2000 GENERAL INSURANCE CONVENTION
25-28 OCTOBER

PARAMETER UNCERTAINTY FOR EXTREME VALUE DISTRIBUTIONS

Presenter: Ana Mata

Parameter Uncertainty for Extreme Value Distributions

Ana J. Mata

Summary

The main objective of this paper is to study different levels of uncertainty that affect the premiums for high excess of loss layers when pricing risks using extreme value distributions. Using some statistical results and estimated distributions presented in McNeil (1997) for a Danish fire insurance portfolio we give price indicators for excess of loss reinsurance layers. Furthermore, we carry out a statistical analysis in order to make inferences about the premiums calculated with the estimated distributions for the underlying data set.

We incorporate different levels of uncertainty that affect the premiums calculated using estimated distributions: model uncertainty, parameter uncertainty for the loss distribution and parameter uncertainty for the frequency distribution. For each level of uncertainty we simulate the distribution of the premiums for high excess of loss layers and we compare the mean of this distribution with the premium calculated using the original data set.

The paper is organised as follows: in Section 2 we summarise the main theoretical results of Extreme Value Theory when dealing with large losses above high thresholds. Section 3 summarises the main statistical results given in McNeil (1997) for the Danish portfolio, and in Section 3.3 we give price indicators for excess of loss reinsurance layers. Finally in Sections 4 and 5 we present the methodology we use to simulate the distribution of the price indicators described above. We make some comments and conclusions about how each layer of uncertainty affect the premiums given for excess of loss layers.

1 Introduction

Extreme Value Theory studies probabilistic models for the occurrence of rare events in different areas. Recently the use of Extreme Value Theory has become of interest in a wide variety of fields such as modelling of insurance losses, modelling financial indicators and modelling of extreme climatological conditions, to mention but a few.

The literature on Extreme Value Theory is extensive and growing rapidly. If we are interested in the statistical methods to identify extremal events in practice a recent reference is Beirlant *et al* (1996). In that book they present a step-by-step guide on how to use the statistical methods for extreme values in practice, through a detailed analysis of large insurance losses. Another reference which emphasises more the practical applications of extreme value statistics is Falk *et al* (1994), where they review many of the theoretical aspects of Extreme Value Theory and in each chapter they give direct applications of the concepts by means of statistical analysis using data sets in many fields. In the latter book they also introduce the statistical software for extreme value statistical analysis, **xtremes**, and they give a complete introduction to the use of this package through a variety of examples.

A solid theoretical understanding of Extreme Value Theory is provided by Embrechts *et al* (1997). This book presents a detailed review of the mathematical and probabilistic foundations of Extreme Value Theory and its applications. Furthermore, they provide some applications of Extreme Value Theory in insurance and finance with specific examples and a detailed guide on how to use statistical methods that are supported by the probabilistic models developed in Extreme Value Theory.

In a recent paper, McNeil (1997) uses one of the data sets presented in Embrechts *et al* (1997) on large insurance losses for a Danish fire insurance portfolio to give price indicators when pricing high excess of loss layers in reinsurance. He uses extreme value statistical methods to estimate the loss distribution for this insurance portfolio, and then uses the expected value of a single claim to a layer as a price indicator for a reinsurance layer.

In this paper we use the same data set and some of the extreme value distributions fitted by McNeil (1997) to give price indicators for two reinsurance layers. However, unlike most of the references available in the field of statistical analysis for extreme value theory, our objective is not only to use estimated distributions to give price indicators but also to make inference in order to answer the question: **"How uncertain are the premiums given for high layers based on the observed data?"**. In order to give some answers to this question we have to make inferences about the parameters estimated for the extreme value distribution using the data. Our research has been carried out to incorporate different levels of uncertainty that affect the price of excess of loss layers. We study the effect of parameter uncertainty for both the loss distribution and the frequency, and also model uncertainty by choosing two different models for the loss distribution. For each level, we simulate the distribution of the price indicators and make inferences about this distribution with respect to the price calculated with the original parameters estimated from the data.

2 Extreme Value Theory

When it comes to insurance and reinsurance pricing we are interested in having good probabilistic models for the distributions for the numbers and amounts of losses incurred in a contract year. Standard distributions such as the Pareto and the lognormal are commonly used to model insurance losses. In the last decade the reinsurance industry suffered large losses originated by catastrophic events whose magnitude was beyond any predicted loss given by the models available for these types of portfolios. Hence, a real understanding of statistical modelling for extremal events became of great interest among pricing staff and underwriters in many insurance and reinsurance companies. Embrechts *et al* (1997) write: *"Extreme value theory has an important role to play in the pricing of reinsurance contracts, especially in the area of contracts for single events or few events, involving high layers"*.

In the case of pricing high excess of loss layers, the reinsurer's concern lies in those rare events that might cause very large losses to the primary insurer and that therefore are likely to affect the reinsurer. Also it is of interest to have good explanatory models for those large losses in order to calculate premiums that are neither too low nor too high.

We are interested in layers of the type $m \; xs \; l$, where l is the deductible. Thus, a reinsurer trying to price this layer would be interested in all those losses above a certain threshold u, which satisfies $0 < u \leq l$. If $X_1, X_2 \ldots, X_n$ are the variables representing the losses to the insurer's portfolio, then the reinsurer would carry out the statistical modelling in order to find a good estimate of the tail from the threshold u of the original distribution given by

$$F(x) = P(X \leq x) \qquad \text{for } u \leq x \leq x_F, \tag{1}$$

where x_F is called the right-end point. Note that x_F might take the value infinity. Also the reinsurer might be interested in the conditional distribution of the excesses once they have surpassed the threshold u. The conditional distribution of the excesses is given by

$$\overline{F}_u(x) = P(X - u \leq x \mid X > u) \qquad 0 < x \leq x_F - u, \tag{2}$$

Another function that plays an important role in the pricing process of an excess of loss layer is the so called mean excess function which is the conditional expectation of the excess over a threshold u. The mean excess function is defined as

$$e(u) = E[X - u \mid X > u]. \tag{3}$$

Given the mean excess function the distribution of the excesses can be immediately identified since there is a one-to-one relation between the distribution function and its mean excess function. Embrechts *et al* (1997) give tables and graphs of several standard probability distributions and their mean excess functions. Thus, in a data analysis, the empirical mean excess function from a particular data set will be a very useful tool to identify which family of distributions would be a reasonable model for the observations.

When we look at excesses over a threshold u, Extreme Value Theory points to the generalised Pareto distribution (GPD) which is characterised by the following definition.

Definition 1 *The standard generalised Pareto distribution (GPD) is given by the following distribution function*

$$G_\xi(x) = \begin{cases} 1 - (1+\xi x)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-x) & \text{if } \xi = 0 \end{cases} \tag{4}$$

This distribution has support $x \geq 0$ if $\xi \geq 0$ and $0 \leq x \leq -1/\xi$ if $\xi < 0$. The notation $G_{\xi,\sigma,\mu}$ refers to the generalised Pareto distribution for a location-scale change in the argument of (4), i.e, we substitute the argument x for $(x-\mu)/\sigma$ in formula (4), changing the support of the distribution accordingly.

The following result summarises in a very simple form one of the main theoretical results in Extreme Value Theory for excesses above high thresholds.

Result 1 (The Pickands-Balkema-de Haan Theorem). *Under certain conditions, the generalised Pareto distribution ($G_{\xi,\sigma,u}$) is the limiting distribution for the excesses over the threshold u, as the threshold tends to the right endpoint x_F.*

For more details about the required conditions and the proof of this result see Embrechts *et al* (1997). Basically, this result states that the distribution function defined in (2) can be approximated by the generalised Pareto distribution for a large threshold. Hence, the GPD plays a very important role in the statistical methods used to estimate distributions for the largest observations.

Many of the statistical methods for extreme value analysis are based on the properties of the generalised Pareto distribution. Embrechts *et al* (1997) give a detailed summary and analytical proofs of all the properties of extreme value distributions including the GPD. Some of the properties that will be useful in our analysis are:

1. If X has a GPD with parameters ξ and σ, $(G_{\xi,\sigma})$, the excess over the threshold u has also a GPD $(G_{\xi,\sigma,u})$ and its mean excess function over this threshold is given by

$$e(u) = E[X - u \mid X > u] = \frac{\sigma + \xi u}{1 - \xi}, \qquad \text{for } \xi < 1.$$

   We notice that for the GPD the mean excess function is a linear function of the threshold u. Hence, if the empirical mean excess of loss for a sample has a linear shape then a generalised Pareto might be a good statistical model for the overall distribution. However, if the mean excess plot does not have a linear shape for all the sample but is roughly linear from a certain point, then it suggests that a GPD might be used as a model for the excesses above the point where the plot becomes linear.

2. The generalised Pareto is closed under change of threshold, in other words if X has a GPD $(G_{\xi,\sigma})$, the probability that X exceeds $u + v$ given that it has exceeded u is also a probability in the generalised Pareto family. This property is very useful in reinsurance particularly when the reinsurer has taken two layers of the same risk.

3. If the conditional distribution above the threshold u is a GPD $(G_{\xi,\sigma}(x - u))$, then we can estimate the distribution of the tail of the original distribution defined in (1) as follows

$$\hat{F}(x) = (1 - F_n(u))G_{\xi,\sigma}(x - u) + F_n(u) \qquad x > u,$$

   where $F_n(u)$ is the empirical cumulative distribution function evaluated at the threshold u. The tail of the original distribution also has a GPD with the same parameter ξ, but different scale-location parameters, see McNeil (1997).

For the purpose of pricing reinsurance contracts per event or for few events we are also interested in the number of excesses above a threshold in a certain period of time, for example a year. Loosely speaking, the following result provides the limiting distribution of the number of excesses above a high threshold.

Result 2 *The number of excesses over a high threshold u follows a Poisson Process.*

A detailed proof of the result outlined above can be found in Embrechts *et al* (1997), Chapter 5, where they present an approach to extremes by means of a point process.

3 The data: Danish fire data

The data we use in this section represent large losses for a fire insurance portfolio in Denmark. There are 2157 losses over 1 million Danish Krone in the years 1980-1990. A detailed statistical analysis of this data has been provided by McNeil (1997) and also by Embrechts *et al* (1997, Chapter 6). This data is available in the Extreme Value Statistics library in Splus.

3.1 The individual claim amounts distribution

In this section we summarise some of the results of the statistical analysis for the Danish fire data studied by McNeil (1997). Basically, we review the possible probabilistic models that can be used as explanatory models for the loss distribution of the insurance claims based on the observed losses in the 11 years.

Usually in reinsurance there would be a large number of small losses that do not result in liabilities for the reinsurer and therefore the ceding company does not report them. Nevertheless, there is a small number of large losses that are unlikely to occur, but once they occur they might cause large losses to the reinsurer. For the Danish data, all the losses are above 1 million Danish Krone (DKK), and they represent the individual losses to the insurance portfolio $Y_1, Y_2, \ldots$. The losses are in units of 1 million DKK.

In any statistical analysis of loss distributions there are several possibilities for probabilistic models for a particular data set. The first option would be to find a parametric model to estimate the overall distribution. In this case the resulting distribution would approximate the following distribution

$$F_{Y^\delta}(y) = P(Y \leq y \mid Y > \delta) \qquad y > \delta, \tag{5}$$

where $\delta = 1$ for the Danish data. In this case the standard graphical tools would provide good guidance to the shape of the distribution and then the parameters can be estimated by maximum likelihood methods. McNeil (1997) fitted some standard distributions to the overall data set typically used to model insurance losses, which are the lognormal and the Pareto, and he also included the generalised Pareto distribution.

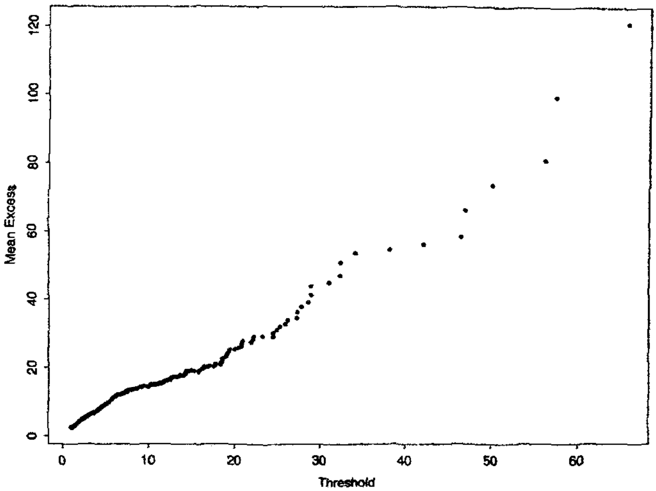

Figure 1: Mean excess plot for Danish data

In his discussion McNeil (1997) writes:

"The lognormal is a reasonable fit, although its tail is just a little too thin to capture the behaviour of the very highest observed losses. The Pareto, on the other hand, seems to overestimate the probabilities of large losses...The GPD is somewhere between the lognormal and the Pareto in the tail area and actually seems to be quite a good explanatory model for the highest losses..."

However, if a reinsurer is using historical information to find statistical models for pricing high excess of loss layers, then he would be more interested in good models for the largest losses. In fact, if the reinsurer wants to price a layer with deductible l he would be interested in all the losses above a certain threshold u, such that $\delta \ll u \leq l$ as we discussed in Section 2. In this case the statistical methods to find good models for the tail area are based on the Extreme Value Theory results summarised in Section 2, which state that for a reasonably high threshold the losses follow, in the limit, a generalised Pareto distribution. Figure 1 shows the empirical mean excess function for the Danish data. Based on the properties of the GPD described in Section 2, McNeil (1997) chooses two different thresholds u= 10 and 20 million DKK, from which the mean excess function plot becomes roughly linear, see Figure 1. After choosing the threshold, McNeil estimates by maximum likelihood the parameters ξ and σ for the generalised Pareto distribution given in (4), with location parameter u and scale parameter σ. This estimation provides an estimate of the conditional probability distribution function over the threshold u as defined in formula (2). Then using this conditional distribution we can approximate the tail of the original distribution as defined in formula (1) by

$$\hat{F}_{Y^\delta}(y) = (1 - F_n(u))G_{\hat{\xi},\hat{\sigma}}(y-u) + F_n(u) \quad y > u, \tag{6}$$

where $F_n(u)$ is the empirical estimate of $P(Y^\delta \leq u)$.

In our case study we decided to chose two consecutive layers, 80 xs 20 and 100 xs 100. Notice that the first deductible is 20 which coincides with one of the thresholds used to fit the generalised Pareto distribution. Since we are interested only in those claims that are above 20 million DKK we will use the conditional probability of a single claim being greater than 20 million DKK.

For a threshold of 10 million DKK, there are only 109 data points above this threshold. In this case the maximum likelihood estimates for the parameters of the generalised Pareto distribution are $\hat{\xi} = 0.497$ and $\hat{\sigma} = 6.98$, and $F_n(10) = 0.95$. Therefore the following estimate for the distribution can be used to approximate the distribution in (5)

$$\hat{F}_{Y^\delta}(y) = 0.05\left(1 - \left(1 + \frac{\hat{\xi}(y-10)}{\hat{\sigma}}\right)^{-1/\xi}\right) + 0.95 \qquad y > 10. \tag{7}$$

Then we estimate the conditional distribution of being greater than 20 million DKK as follows

$$\hat{F}_{Y^\delta > 20}(y) = P(Y^\delta \leq y \mid Y^\delta > 20) = \frac{\hat{F}_{Y^\delta}(y) - \hat{F}_{Y^\delta}(20)}{1 - \hat{F}_{Y^\delta}(20)} \quad y > 20, \tag{8}$$

where $\hat{F}_{Y^\delta}(y)$ is as defined in (7).

For a threshold of 20 million DKK, the fitted generalised Pareto distribution gives the conditional distribution for all those losses that are greater than 20 million DKK. Therefore we use this distribution directly. In this case there are only 36 data points and the maximum likelihood estimates for the generalised Pareto distribution are $\hat{\xi} = 0.684$ and $\hat{\sigma} = 9.63$, which gives the following distribution function

$$\hat{F}_{Y^\delta > 20}(y) = P(Y^\delta \leq y \mid Y^\delta > 20) = 1 - \left(1 + \frac{\hat{\xi}(y-20)}{\hat{\sigma}}\right)^{-1/\xi} \quad y > 20. \tag{9}$$

We will use for our analysis both conditional distributions (8) and (9) to show how the results are model dependent and how they are sensitive to the choice of threshold, which has also been discussed in McNeil (1997).

3.2 The distribution of the number of claims

As we discussed above, in the case of reinsurance contracts with limited number of losses or events we also need good estimates for the frequency distribution in order to estimate the total losses for the reinsurer.

For the layers we have chosen we are looking at losses above 20 million DKK, hence we only require to fit a distribution for the number of claims in a year that lie above this amount. From the data we obtain 36 observations that are greater than 20 million DKK and the freauency is given in Table 1.

Years (1980-1990)	80	81	82	83	84	85	86	87	88	89	90
No. Claims > 20 (N_i)	3	4	5	0	0	3	1	4	8	5	3

Table 1: Frequency of claims above 20 million DKK

If we denote N the total number of claims above 20 million DKK, from Table 1 we calculate the sample mean and variance of N. It can be seen that

$$E[N] = \frac{36}{11} = 3.27 \qquad \text{and} \qquad Var(N) = \frac{\sum_{i=1}^{11}(N_i - E[N])^2}{10} = 5.1074.$$

In Section 2, Result 2 states that for high thresholds the number of excesses follows a Poisson Process. Hence, we fit the Poisson distribution as the probabilistic model for the distribution of the number of losses above 20 million DKK. The probability function is given by

$$P(N = n) = \frac{e^{-\lambda}\lambda^n}{n!} \qquad n = 0, 1, 2, \ldots,$$

and in this case the maximum likelihood estimate is $\hat{\lambda} = 3.27$. Table 2 gives the observed frequency, the estimate probabilities and the expected frequency under the model proposed. We would like to test whether the proposed model provides a good explanatory model for the observed data.

If we use the Pearson χ^2 test with the results in Table 2 we must calculate the statistic given by

$$X^2 = \sum_{k=0}^{5} \frac{(Obs.freq. - Exp.freq.)^2}{Exp.freq.} \sim \chi^2_{k-p-1}, \tag{10}$$

to test the null hypothesis that the observations follow a Poisson distribution. In equation (10) k is the number of classes in which the observations have been grouped, and p is the number of parameters estimated from the sample. In this

N	0	1	2	3	4	≥ 5
Obs. freq.	2	1	0	3	2	3
Prob (p_i)	0.038	0.124	0.203	0.221	0.181	0.232
Exp. freq. (np_i)	0.418	1.364	2.233	2.431	1.991	2.552

Table 2: Expected frequency for Poisson Distribution

case the statistic is $X^2 = 8.528$ that we compare with $\chi^2_{4,0.95} = 9.488$. This result indicates that for a 95% significance level we cannot reject the null hypothesis that the observations follow a Poisson distribution.

However, the χ^2 test is applicable only under certain conditions: we require at least 25 observations, in this case we only have 11 years of observations. Also, this empirical test is supported by the Central Limit Theorem and its justification is based on the assumption that the expected frequency for each class should be greater than or equal to 5, which is not satisfied in our case. If we try to meet this requirement we would have to regroup the observations, such that $np_i \geq 5$, therefore we would have only two classes ($0 \leq N \leq 2$) and ($N \geq 3$). This would leave us with insufficient degrees of freedom to estimate the parameters and apply the test. Even though we have not rejected the null hypothesis the validity of the test is not justified.

3.3 Price indicators

In this section it is our objective to give price indicators when pricing excess of loss reinsurance. We choose as price indicator for the reinsurance layers the expected value of the aggregate claim amounts.

The layers we are going to price are 80 xs 20 and 100 xs 100. Since we use the conditional probability distribution, we define the conditional excesses above the deductible of 20 million DKK to be $Y_1^{(20)}, Y_2^{(20)}, \ldots,$ where $Y_i^{(20)} = (Y_i^\delta - 20 \mid Y_i^\delta > 20)$. Therefore the random variables representing the individual claim amounts to the reinsurer for each layer are

$$Z_{i1} = \min(Y_i^{(20)}, 80) \quad \text{and} \quad Z_{i2} = \min(\max(Y_i^{(20)} - 80, 0), 50).$$

The expected value of a single claim to the first layer can be calculated as follows

$$E[Z_{i1}] = \int_{20}^{100} (y - 20) d\hat{F}_{Y^\delta > 20}(y) + 80 \left(1 - \hat{F}_{Y^\delta > 20}(100)\right),$$

Layer	$u = 10$	$u = 20$
80 xs 20	18.3634	17.8030
100 xs 100	2.6658	3.6030
180 xs 20	21.0292	21.4060

Table 3: Expected value of a single claim for each layer

where $\hat{F}_{Y^\delta>20}(y)$ can be either the conditional distribution in (8) or (9). The same type of integral can be used for the expected value of a single claim to the layer 100 xs 100. Table 3 shows the expected value of a single claim to each layer for both choices of thresholds.

We notice from the results in Table 3 that the results are very variable with the choice of threshold, particularly for higher layers. In the case of the first layer, for a threshold of 10, the expected value is higher than for a threshold of 20, while for the second layer we have the opposite result. We notice as well that the relative difference is greater for the second layer than for the first. Even though in both cases we are using conditional distributions, when we use the threshold of 10 we have more data points to calculate the maximum likelihood estimates for the GPD than for a threshold of 20. Moreover the variability for the second layer is higher because the data in that interval is very scarce and the parameters are subject to higher standard errors. When we consider the combined layer with the estimated distributions the premium is equivalent to the sum of the premiums for both layers separately. However, in Section 5 we will notice that the distribution of the premiums for the combined layer follows the same pattern as the premiums for the first layer.

We define N as the number of losses above 20 million DKK. Hence the aggregate claim amounts at the end of the year are

$$S_1 = \sum_{i=1}^{N} Z_{i1} \qquad \text{and} \qquad S_2 = \sum_{i=1}^{N} Z_{i2}.$$

For each layer we give as price indicator the expected value of the aggregate claim amounts, i.e. $P_1 = E[N]E[Z_{i1}]$ and $P_2 = E[N]E[Z_{i2}]$, and from Section 3.2 we have $E[N] = 3.27$. For each choice of threshold, $u = 10$ and $u = 10$, Table 4 shows the price indicator for each layer. Note that the results in Table 4 are the same results as in Table 3 multiplied by 3.27, and therefore we observe the same patterns as we discussed above.

Layer	$u = 10$	$u = 20$
80 xs 20	60.0483	58.2158
100 xs 100	8.7172	11.7818
180 xs 20	68.7655	70.0076

Table 4: Expected value of aggregate claim amounts

4 Parameter uncertainty for extreme value distributions

In previous sections we have discussed the possible distributions we can use to approximate the loss distribution and the frequency for the Danish fire data. Since we are interested in having good estimates for the largest losses we decided to use the estimated distributions for the tail. Using these distributions we have calculated price indicators for excess of loss layers.

Now we have reached the point where we have to ask:

How certain/uncertain are these premiums?

Of course, when pricing insurance contracts uncertainty is always present since the insurer wants to be sure he charges adequate premiums for any line of business. However, when it comes to extremal events or observations, the insurer/reinsurer needs to make inferences about possible occurrences outside the observations where there is very little information, therefore the uncertainty of any estimator is even higher.

In his analysis of price indicators for the Danish portfolio, McNeil (1997) writes:

"We should be aware of various layers of uncertainty which are present in any data analysis, but which are perhaps magnified in an extreme value analysis.
On one level, there is parameter uncertainty. Even when we have abundant, good-quality data to work with and a good model, our parameter estimates are subject to standard errors...
Model uncertainty is also present ... If we set the threshold too high we have few data and we introduce more parameter uncertainty. If we set the threshold too low we lose our theoretical justification for the model ..."

Apart from these layers of uncertainty discussed in McNeil (1997), in our case we have an extra layer of uncertainty which is the frequency distribution. We have very few years of observation and therefore the χ^2 test does not provide a satisfactory answer to which probabilistic model will be an appropriate choice for the frequency of the losses.

It is our objective in the following section to calculate distributions and standard errors of the price indicators given in Section 3.3. We study the variability of these price indicators at various levels:

1. Fix the parameter for the frequency distribution and vary the parameters for the loss distribution. For each combination calculate the premiums with the corresponding aggregate distribution.
2. Vary the parameters for both the loss and the frequency distribution and for each combination calculate the corresponding premium.

In each of these steps we are taking into account the two choices of threshold we discussed in Section 3.3, and in each case we calculate the expected value of the aggregate claim amount. In the following sections we give the methodology we have used.

4.1 Uncertainty in the loss distribution

The loss distribution we are using is the GPD as defined in formula (4), with location-scale parameters; thus we substitute x by $(x-u)/\sigma$ in the distribution function. As we have discussed in the previous sections, the threshold u is chosen using the mean excess function graph, and after the threshold is fixed the parameters ξ and σ are estimated by maximum likelihood. Even though for both choices of thresholds we used the conditional probability of the excesses being greater than 20 million DKK as defined in formulae (8) and (9), in each case the parameters have been estimated with different numbers of sample observations.

If we were interested in making inferences about the estimated parameters, the theory of maximum likelihood states that under certain conditions of regularity the maximum likelihood estimates follow asymptotically a multivariate normal distribution. In the case of Extreme Value Theory, the parameter ξ determines whether the distribution is heavy tailed or not. McNeil (1997) indicates that for $\xi > -0.5$ the generalised Pareto distribution satisfies the regularity conditions of maximum likelihood estimation. Therefore, the maximum likelihood estimates for ξ and σ from a sample of n excesses follow in the limit a bivariate normal distribution, as specified in the following relation:

$$n^{1/2}\begin{pmatrix}\hat{\xi}_n\\ \hat{\sigma}_n\end{pmatrix} \xrightarrow{d} N\left[\begin{pmatrix}\xi\\ \sigma\end{pmatrix}, \begin{pmatrix}(1+\xi)^2 & \sigma(1+\xi)\\ \sigma(1+\xi) & 2\sigma^2(1+\xi)\end{pmatrix}\right] \quad (11)$$

In our case we are not interested in making inferences about the parameters, but about the price indicators calculated in Section 3.3. However, the premiums were calculated using the estimated parameters. Hence at first, the asymptotic normality seems to be the answer to our problem. We could simulate M pairs of parameters (ξ, σ) from the bivariate normal distribution, with mean vector and covariance matrix calculated using the maximum likelihood estimates. Then for each pair (ξ_i, σ_i) and each threshold we can use the GPD $(G_{\xi_i,\sigma_i}(x-u))$ for the loss distribution, and finally with this loss distribution we can calculate the corresponding premium. This method would produce M values of the price indicators that will give us a clear idea of the distribution of the premiums.

The problem with this approach is that when we simulate pairs (ξ_i, σ_i) from the bivariate normal we obtained positive and negative values for both parameters ξ and σ. For ξ negative, we have to truncate the GPD at the right-end point $-\sigma/\xi$ as defined in Section 2 in the definition of the generalised Pareto distribution. If the right-end point is smaller than the deductible for the corresponding layer, then the corresponding premium associated with the GPD for this pair of parameters is set as zero. In the case of negative values for the parameter σ the GPD is not defined. Thus there is no interpretation of the corresponding premium for a pair of parameters where $\sigma < 0$. Hence, the asymptotic normality does not provide a good method to simulate the distribution of our price indicators.

Therefore we decided to use the bootstrap method to calculate standard errors for any statistic estimated from an observed sample. We give step by step the methodology we used to simulate parameters with the bootstrap method. For more information about this method, see, for example, Efron and Tibshirani (1993).

4.1.1 The bootstrap method

We call $\mathbf{x} = (x_1, \ldots, x_n)$ the observed sample from which the parameters $\hat{\xi}$ and $\hat{\sigma}$ and the premium indicator $P(\hat{\xi}, \hat{\sigma}, u)$ have been estimated. For threshold $u = 10$, $n = 109$ and for $u = 20$, $n = 36$.

We used the bootstrap method to simulate pairs (ξ_i, σ_i) for $i = 1, \ldots, B$ for each chosen threshold as follows

1. Assign the same probability to each observation of the original sample $\mathbf{x}$. In other words, we generate an empirical uniform distribution for the observations, with $p(x_i) = 1/n$.

2. Generate the bootstrap sample: We produce a random sample $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ by sampling with replacement from the empirical distribution defined in step 1. Notice that the bootstrap sample is of the same

length of the original vector, but some observations might be repeated and some might not appear. We repeat this procedure to produce B bootstrap samples.

3. For each sample we estimate by maximum likelihood the parameters for the generalised Pareto using the Extreme Value library available in Splus. Thus we have (ξ_i, σ_i) for $i = 1, \ldots, B$.

4. For each pair we use the corresponding GPD $(G_{\xi_i,\sigma_i}(x-u))$ to calculate the conditional distributions defined in formulae (8) and (9). Then for each pair of parameters and each combination of compound distribution we calculate the corresponding premium with the methodology given in Section 3.3. Therefore we have $P_i(\xi_i, \sigma_i, u)$ for $i = 1, \ldots, B$.

5. The bootstrap standard error for any estimated statistic, in particular for the premiums, is given by

$$\widehat{se}_B(P) = \left(\frac{\sum_{i=1}^{B} (P_i(\xi_i, \sigma_i, u) - \overline{P}(\xi_i, \sigma_i, u))^2}{B-1} \right)^{1/2}, \tag{12}$$

where $\overline{P}(\xi_i, \sigma_i, u) = \left(\sum_{i=1}^{B} P_i(\xi_i, \sigma_i, u) \right) / B$.

In Section 5 we show our numerical results for the standard errors and the premium distributions for each layer.

4.2 Uncertainty in the frequency distribution

Apart from the uncertainty generated by the parameters of the loss distributions we also have uncertainty in the frequency. First we have model uncertainty. Since there are few years of observation, the χ^2 test does not provide a valid method to test any of the models proposed. We estimated from the data the parameters for the Poisson distribution, therefore when we combine this distribution with the loss distribution, the distribution of the total aggregate claim amount is subject to higher uncertainty.

We study the effect of the uncertainty in the frequency distribution and in this case we applied the following methodology:

1. Generate B sets of 11 random numbers from a Poisson distribution with parameter $\lambda = 3.27$.

2. For each set we calculate the sample mean.

3. We use the same set of simulated parameters for the loss distribution as in Section 4.1 and for each set of parameters we associate a Poisson distribution with parameter λ_i as calculated in step 2.

4. For each threshold we have three parameters associated with the compound Poisson distribution, $(\xi_i, \sigma_i, \lambda_i)$. For each vector of parameters we have a premium associated for each layer, that is $P_i(\xi_i, \sigma_i, \lambda_i, u)$.

5. Then we calculate the standard errors of these premiums as defined in step 5 of the bootstrap method in the previous section.

We compare the standard errors for the premiums produced with this extra level of uncertainty with the standard errors obtained when we fixed the Poisson parameter.

5 Premium distributions

In this section we present the numerical results of our research using the methodology described in Section 4. With the bootstrap method to calculate the standard error, usually between 50-200 bootstrap samples $\mathbf{x}^*$ are sufficient, see Efron and Tibshirani (1993). For our calculations we used 500 bootstrap samples to produce more accurate calculations since we are using extreme value distributions. When we simulated 500 pairs of parameters for the GPD the bootstrap standard error for the parameters were $\widehat{se}_B(\xi) = 0.15$ for $u = 10$ and $\widehat{se}_B(\xi) = 0.28$ for $u = 20$. These are the same as the standard errors given in McNeil (1997) for the shape parameter ξ.

First we consider the aggregate claim amount with fixed parameter for the frequency distribution. The numerical results we show in our tables are the sample mean from the 500 simulated premiums and the number in brackets represents its standard error. Table 5 shows the mean of the distribution of the expected value of the aggregate claim amount with fixed parameter for the Poisson distribution and its standard error.

We notice from the results in Table 5 that for the first layer and the combined layer the mean of the expected value of the aggregate claim amount is lower than the expected value calculated with the original data set, see Table 4. For the second layer, for both choices of thresholds, there is very little difference between the mean of the expected value of the aggregate claim amount and the expected value calculated in Table 4. However, we notice that the standard errors are always higher for the threshold $u = 20$, particularly for the second layer where the increase of the standard error is around 20%. For the first layer and the combined layer the standard error increases around 12% for $u = 20$. We discussed in Section 3 that the number of observations above 20 million DKK is only 36, therefore the maximum likelihood estimates are subject to higher standard errors, and so are the premiums calculated with these parameters. Moreover, for the second layer the variability is even higher since only 3 observations fall in this layer. Thus, for high layers the extreme value models try to capture the behaviour of the tail with very little information, which means more uncertainty.

Layer	$u = 10$	$u = 20$
80 xs 20	58.6695 (10.5087)	57.8581 (11.7270)
100 xs 100	8.8857 (5.8613)	11.6753 (7.1542)
180 xs 20	67.5551 (16.2073)	69.5334 (18.0166)

Table 5: Mean and standard error for the price indicators, fixed $\lambda = 3.27$

Figure 2 shows the distribution of the price indicators given in Section 3.3 for each layer and each threshold. We notice that the distribution of the expected value of the aggregate claim amount for the first layer and the combined are symmetric around their own mean, while for the second layer the distribution is skewed to the right and there is more variability. The second layer is subject to very low premiums since there is a very small probability of a claim being greater than 100, but on the other hand there is a high uncertainty in this layer due to the few observation; therefore there are a few large values of the premiums that make the overall distribution skewed. The distribution of the mean amounts are not very different when we change threshold; in all the cases the scale for the premiums is wider for $u = 20$.

After studying the effect of the uncertainty in the loss distribution on the price indicators we incorporate the uncertainty in the frequency distribution. Using the methodology described in Section 4.2 we simulate different parameters for the Poisson distribution. Table 6 shows the mean of the distribution of the expected value of the aggregate claim amounts varying both the parameters for the loss distribution and for the frequency distribution, and the corresponding standard errors.

When we compare the results in Table 6 with the results in Table 5 we notice that in general the means of the premiums when we vary the Poisson parameter are lower than the means with a fixed Poisson parameter, and also lower than the premiums calculated with the original data, see Table 4. However, the standard errors in Table 6 are higher than the standard errors in Table 5, particularly for the first and combined layer.

The standard error for the first layer increases around 36% for $u = 10$ and 27%

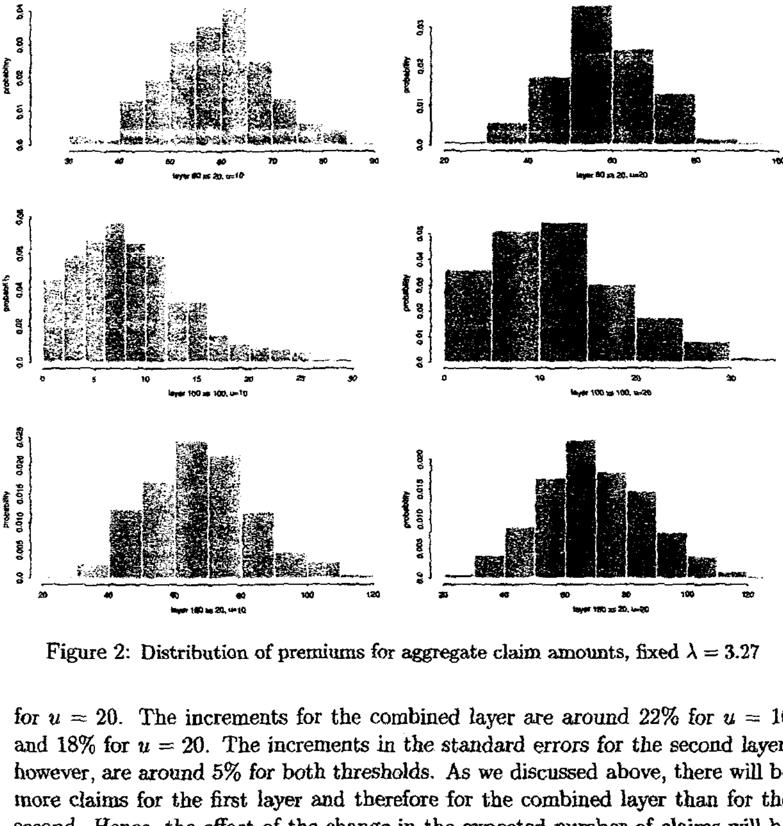

Figure 2: Distribution of premiums for aggregate claim amounts, fixed $\lambda = 3.27$

for $u = 20$. The increments for the combined layer are around 22% for $u = 10$ and 18% for $u = 20$. The increments in the standard errors for the second layer, however, are around 5% for both thresholds. As we discussed above, there will be more claims for the first layer and therefore for the combined layer than for the second. Hence, the effect of the change in the expected number of claims will be higher for the first layer and the combined. The probability of a claim being greater than 100 million DKK is less than 10^{-3} for a high proportion of the simulated parameters, therefore the variation of expected value of the number of claims for this layer is not very significant when we compare it with the original Poisson parameter.

Figure 3 shows the distribution of the premiums *for simulated values of* λ. Comparing Figures 2 and 3 we do not observe major differences in the shape of the distribution of the premiums by including an extra layer of uncertainty. Nevertheless, we observe that in all the cases the scale is wider when we vary the Poisson parameters. In other words, we observe more values in both extremes of the histograms, particularly for the first layer and the combined layer.

Layer	$u = 10$	$u = 20$
80 xs 20	58.2018	57.4069
	(14.3462)	(14.9155)
100 xs 100	8.7979	11.5552
	(6.1195)	(7.4599)
180 xs 20	66.9997	69.0621
	(19.8031)	(21.1883)

Table 6: Mean and standard error for the price indicators, simulated λ

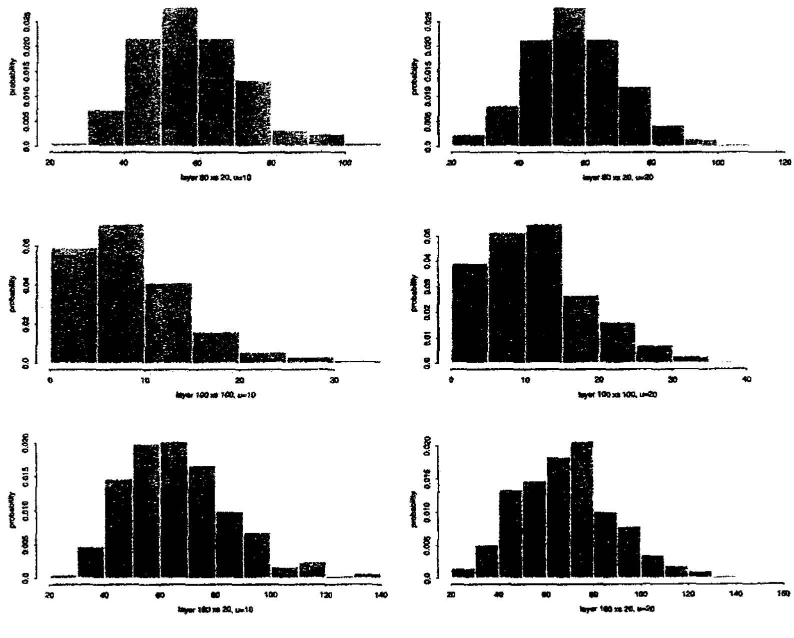

Figure 3: Distribution of premiums for aggregate claim amounts, simulated λ

6 Conclusions

When using historical data to fit loss distributions uncertainty is always present. In insurance and reinsurance it is not rare to have insufficient data to make inferences about the loss distribution. In the case of fitting the tail of the loss distribution the uncertainty is even higher since there would be very few observations that fall in the tail of the distribution. When we use the generalised Pareto distribution to fit the tail of the loss distribution above high thresholds the parameters are estimated with very few data points and therefore any statistic estimated using these parameters would be subject to high standard errors, in particular the premiums.

With the results shown in this paper it was our objective to provide some insight on how the premiums for high excess of loss layers are affected by different layers of uncertainty when using extreme value distributions. From the numerical results we observed that for high layers where there are very few data points the premiums are very variable and they are very sensitive to the choice of threshold. We noticed that for higher thresholds the premiums are subject to higher standard errors, particularly for layers with high deductible. The uncertainty added by the frequency distribution affects more lower layers than higher layers, this is due to the fact that there would be more claims affecting lower layers.

In the data set we are considering it seems reasonable to use extreme value statistical methods as has been discussed in McNeil (1997), but for any other data set other distributions might provide a better model depending on the line of business and the type of claims.

By incorporating different levels of uncertainty that might affect the premiums for a certain risk the pricing actuary is given tools to make decision about what premium should be reasonable that is neither too low nor too high. Other external sources of information should also be taken into account when princing excess of loss reinsurance such as the market competition and insurance cycles.

Acknowledgements

The author would like to thank Prof. Howard Waters for his continuous encouragement and useful comments and ideas throughout the preparation of this paper. The author is also grateful to The British Council and FUNDAYACUCHO, Venezuela, for the financial support that made possible the development of this research, and the financial support of KPMG that made possible the presentation of this workshop at the GIRO Convention 2000.

References

[1] BEIRLANT, J., TEUGELS, J. AND VYNCKIER, P. (1996) *Practical analysis of extreme values.* Leuven University Press, Leuven.

[2] EFRON, B. AND TIBSHIRANI, R.J. (1993) *An Introduction to the bootstrap.* Chapman and Hall, London.

[3] EMBRECHTS. P, KLÜPPELBERG, C. AND MIKOSCH, T. (1997) *Modelling Extremal Events.* Springer, Berlin.

[4] FALK, M., HÜSLER, J. AND REISS, R. (1994) *Laws of large numbers: extremes and rare events.* Birkhäuser, Basel.

[5] MCNEIL, A.J. (1997) Estimating the tail of loss severity distributions using extreme value theory. *ASTIN Bulletin* **27**: 117-137.