REINSURANCE RATE-MAKING
WITH PARAMETRIC AND NON-PARAMETRIC MODELS

By
Siqi Chen, Madeleine Min Jing Leong, Yuan Yuan

University of Illinois at Urbana-Champaign
1. Introduction

Reinsurance contract is an insurance policy purchased by primary insurer from one or more reinsurance companies. It is important to insurance companies as they can transfer some high risk business with potential losses to reinsurance companies. This transaction carries a mean of risk management and has become more extensive in the global financial market. There are two reinsurance agreements: proportional and non-proportional coverage.

Proportional coverage means the same proportion of premium and losses incurred by the policyholders are ceded to the reinsurer, and thus, rating consideration may not considered in this agreement. With non-proportional coverage, the reinsurer would receive “reinsurance premium” and pay a predefined portion of the claims incurred by the insurer. Excess of loss (XL) reinsurance, which the insurance coverage is paid by insurer to policyholders up to a maximum retention level and any amount exceeds the level will be recouped by the reinsurer. For instance, the insurer may insure a catastrophic risk with a policy limit up to $30 million, and buy a XL reinsurance with retention level of $15 million. In this case the reinsurer will need to pay $5 million to insurer for recovery if the loss is $20 million.

However, this XL reinsurance is usually modeled using the concept of mean excess loss, without any assumptions of a parametric model. This non-parametric pricing approach may raise problem to reinsurance companies as it usually does not capture the tail behavior of the loss distribution. A visual technique of studying the tail behavior of the claim size is applied in this text. Lastly, a parametric model is introduced for better reinsurance rate-making. This model uses the concept of extreme value theory for finding the parameter in the pricing distribution. Net premiums for both parametric and non-parametric are computed and compared in the latter section.

In this paper, valuation of parametric model and non-parametric model for insurance rate-making is performed. Sample insurance claims data from [4] is used throughout the text to give an idea of the claim data modeling problems faced by actuarial practitioner as well as to manifest the two different pricing methods. A histogram of the sample claims data is shown in next section for the purpose of glancing at the data.

2. Histogram of Sample Claims Data

Histogram is one of the common graph to use for displaying the numeric data and illustrating some behaviors of the data. A histogram of the supported data is shown in Figure 1. Figure 1 shows that the claim data is right-skewed where the mean of the losses is larger than the median. In addition, there are few extremely high values (or outliers) that do not fall near to any other data points. Understanding these extreme loss values is important because these have the greatest impact on the total of losses for reinsurance companies. A statistical method will be applied later for viewing the tail behavior of the claim distribution.
3. Non-Parametric XL Model and Quantile Plotting

As mentioned in the introduction, XL reinsurance pricing method relies on the concept of mean excess loss function. Let $X$ be a random variable that represent the size of a claim made by policyholder and $R$ be the retention or priority level of the reinsurance contract. Reinsurer is required to pay $X - R$ if only $X > R$. Actuaries working in reinsurance companies would pay more attention on the expected value of such claim, which is given by

$$m(R) = \mathbb{E}(X - R | X > R)$$

where the expression is determined by mean excess function and is existed if $\mathbb{E}(x) < \infty$.

Let $x_1, x_2, ..., x_n$ be the claim data that are from a random sample of $X_1, X_2, ..., X_n$, the mean excess function $m$ is often estimated at $R = X_{n-k,n}$, the $(k+1)$-th largest observation for $k = 0, 1, 2, ..., n-1$ by the empirical excess of the $k$ points that are higher than $X_{n-k,n}$:

$$\hat{m}_{k,n} = \frac{1}{k} \sum_{j=1}^{k} X_{n-j+1,n} - X_{n-k,n}.$$
The estimates of the mean excesses are illustrated in Figure 2 and they play crucial roles in the rating of XL insurance with a retention level $R$ because the corresponding fair net premium (without any policy expenses and other administrative costs) is determined to be [4]:

$$\Pi(R) = \mathbb{E}[(X - R)_+]$$

To estimate the net premium, $\Pi(R)$, the following equation of the empirical (non-parametric) estimator with a range of retention level from 0 to 7,875,000 (step = 125,000) is given by

$$\hat{\Pi}(R) = \frac{1}{n} \sum_{i=1}^{n} (X_{i+1,n} - R)_+$$

The non-parametric estimator of the net premium as a function of $R$ is shown in Figure 3.
Non-Parametric Estimator of Premium

Figure 3: Non-parametric estimator of the net premium as a function of the retention level $R$

Although the calculation was done with for a wide range of retention levels, the results for high retention level would be more valuable to reinsurer as they are meant to transfer sizable risks (e.g. catastrophic losses). In other words, in order to value the net premium precisely, finding distributions that have similar tail behavior should be emphasized as well as taking account on parametric assumptions.

In order to assist in the rating of XL reinsurance on the statistical side, quantile plotting is addressed to view the tail weight of the claim distribution. Random variables that tend to assign higher probabilities to larger observations are said to be heavier tailed. An definition of heavy-tailed distribution can be described as a heavy-tailed distribution has a tail which is heavier than any exponential tail:

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

in rectangular coordinate system, an exponential quantile plot of points have the coordinates of

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

where the empirical distribution, $X_{n-i+1,n}$ is the estimated quantile of $Q(1 - \frac{i}{n+1})$, for several values of $1 - \frac{i}{n+1} \in (0,1)$[1]. The exponential QQ-plot of the claim size data is constructed below in Figure 4.
From the exponential QQ-plot, the horizontal axis shows the quantile of a standard exponential distribution and the vertical axis represents the empirical quantiles of the claim data. The empirical points in the plot bend upwards and exhibit a convex pattern indicating that the claim size distribution has a heavier tail than expected from a standard exponential distribution\cite{2}. Since most of the reinsurance contracts have the purpose of transferring huge risks of losses, pricing actuaries should use a model that can capture the tail section of the claim distribution. Consequently, an exponential distribution may not be the accurate model of the reinsurance contract since the largest three observations do not fit so well in the exponential QQ-plot.

Since exponential distribution is not accurate for the pricing model due to the skew errors in the QQ-plot, an assumption of using another distribution is made. In general, Pareto distribution, as a heavy-tailed distribution provides a good fit of a large claim size data. This statement can be
proved by looking at Pareto QQ-plot (Figure 5). The Pareto QQ-plot is more linear compared to the exponential QQ-plot which indicating a reasonable fit of the Pareto model to the tail of the claim sizes. Thus, further discussion on using a Pareto-type model with parametric assumption for reinsurance pricing is needed.

![Pareto Q-Q Plot for Sample Data](image)

Figure 5: Pareto quantile plot for sample claims data

4. **Parametric XL Model: Pareto-type Distribution**

A scaled Pareto distribution for the sample data is considered and it has a survival probability function of

$$\bar{F}(x) = \frac{x^{-\alpha}}{C}, x > C$$

for some large $C > 0$ and $x$ is $C$ times a strict Pareto random variable. The mean excess function of the random variable is found by

$$e(t) = \frac{t}{\alpha - 1}, t > 1$$

where $\alpha$ is the shape parameter of the strict Pareto distribution. $\alpha$ is using to be estimated using
the idea of extreme value theory. The estimate of $\alpha$, which has a reciprocal relationship with a
extreme value index (EVI): $\gamma = \frac{1}{\alpha} > 0$. A well-known estimator of the EVI is the Hill Estimator
[3]. Combining the result of log-transformed data is an exponential distribution with the the mean
excess function of the Pareto random variable, the mean excess value of the log-transformed data
(Hill Estimator) [1]

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

where $\hat{\gamma}_{k,n} = \frac{1}{\hat{\alpha}_{k,n}}$.

However, the Hill estimator is based on $k$-th largest observation, which means for each choice of $k$
there would be a different estimator.

**Plot of Hill Estimator as a Function of $k$**

![Plot of Hill Estimator as a Function of $k$](image)

Figure 6: Plot of Hill Estimator as a function of $k$

The ways of finding optimal $k$ using various statistical properties are developed by many statistical
practitioners. One method of finding optimal $k$ is shown by Danielsson et al. (1997) which $k$
is determined by a two-step subsample bootstrap method through minimizing asymptotic mean-
squared error (AMSE)[5]. For a given size of $n_1 < n$, a resample of a $n_1$ size is drawn and the AMSE
for each level of $k_1$ is calculated. Then, the level $k_{10}^*(n_1)$ that minimizes AMSE for this bootstrap
$n_1$ is found. The procedure for a smaller sample size of $n_2 < n_1$ is repeated, where $n_2 = \frac{n_1^2}{n}$. This
procedure can find the level of \( k_{2,0}(n_2) \) that minimize AMSE for this bootstrap \( n_2 \). Lastly, the best choice of value \( k \) can be calculated using the equation given by

\[
\hat{k}_0(n) = \frac{\left( k_{1,0}(n_1) \right)^2}{k_{2,0}(n_2)} \left( \frac{(ln k_{1,0}(n_1))^2}{(2 ln n_1 - ln k_{1,0}(n_1))^2} \right)^{(ln n_1-ln k_{1,0}(n_1))/ln n_1}
\]

With the optimal choice of \( k = 95 \), the Hill estimator, \( \hat{\gamma} \) is computed to be 0.2711.

Knowing the estimated EVI, the calculation of net premium under parametric assumption is clearly shown in [4] and is written as

\[
\hat{\Pi}(R) = \frac{1}{1/\hat{\gamma}_{k,n} - 1} R \frac{k}{n} \frac{R}{X_{n-k,k}}^{-1/\hat{\gamma}_{k,n}}
\]

The relationship of the net premium estimator with retention levels of \( R \) where \( R \in (0, 7875000) \) (incremental step = 125000) is illustrated in Figure 7.

![Parametric Estimator of Premium](image)

Figure 7: Parametric estimator of the net premium as a function of the retention level \( R \)

5. Comparing Parametric and Non-parametric for Different Retention Levels

Comparison between parametric and non-parametric estimators for \( 0 < R < 7875000 \) and \( R > 3000000 \) has been made. From Figure 8, the parametric estimator does not capture the behavior
of left tail part because of the emphasis on the right tail behavior, where the larger observations are located. At a high retention level (around 3 million to 4.5 million), two estimators provide close results. However, beyond 4.5 million retention level, a deviation from two estimators can be visualized. Due to the lack of sample data at high retention level, one would get a low quality result of the net premium estimation using non-parametric approach. On the other hand, the parametric estimator predicts a higher amount of premium because the parametric distribution tends to assign more weight on the tail. Using the technique of modeling larger losses data with a heavy-tailed distribution enables reinsurance companies to make a better forecast on the net premium.

![Comparison of Estimators of Premium](image1)

![Comparison of Estimators of Premium (R>3000000)](image2)

Figure 8: Comparison of the net premium estimators

### 6. Conclusion

An overview of loss modeling process and net premium calculation is done using two methods: non-parametric and parametric. An inspection of the tail behavior for the claims data should be constantly done in order to see if the pricing model is accurate to apply. This paper has provided a suggestion on using parametric model for reinsurance pricing when having insufficient large claims data. However, it may not be a final solution for reinsurers due to different sizes of claims with different risks involved in it.
References


