American Statistical Association

2000
Proceedings
of the
Business and Economic Statistics Section

Papers presented at the
Annual Meeting of the American Statistical Association,
Indianapolis, Indiana, August 13-17, 2000
under the sponsorship of
the Business and Economic Statistics Section

American Statistical Association ■ 1429 Duke Street, Alexandria, VA 22314

TABLE OF CONTENTS

Invited Papers by Topic

Contributed Papers by Topic

Contributed Papers-Poster Sessions

ROBUST PARAMETRIC MODELING OF THE PROPORTIONAL REINSURANCE PREMIUM WHEN CLAIMS ARE APPROXIMATELY PARETO-DISTRIBUTED

Vytaras Brazauskas, University of Wisconsin-Milwaukee
UWM, 3200 N. Cramer Ave., EMS Room E457, Milwaukee, WI 53211

KEY WORDS: Pareto, contamination model, efficiency, generalized median, nonparametric estimation, reinsurance, robustness, tail index.

Abstract

A new approach—robust parametric modeling—is introduced for estimation of the proportional reinsurance premium when individual claim sizes are (approximately) distributed according to a single-parameter Pareto model. This methodology is compared with well-established empirical nonparametric approach on the basis of two generally competing criteria, efficiency and robustness, using the asymptotic variance of the estimator as an efficiency criterion, and the breakdown point of the estimator as a robustness criterion.

Various robust and efficient estimators of a Pareto distribution tail index parameter are discussed. In particular, trimmed mean type, and recently introduced (see Brazauskas and Serfling (2000)) generalized median type estimators are considered.

1 Introduction

Extremal (or catastrophic) events in insurance can lead to individual (or grouped) claims which by far exceed the capacity of a single insurance company. A natural response from the insurance world to this problem is the creation of a reinsurance market which can be characterized as the industry that protects insurance companies from insolvency.

A focal statistical problem in reinsurance is to estimate the total claim amount. Based on the net premium, which is defined as the expected value of the total claim amount, the reinsurer calculates the total premium that must be paid by the policy holder (in this case, an insurance company). Therefore a reliable estimate of the distribution function of the total claim amount is necessary. We use a simplified setting for this problem. (For a more general setting see, e.g., [4], p. 507.)

> The individual claim sizes, $X_1, X_2, \ldots$, are independent identically distributed (iid) non-negative random variables with common distribution function F, independent of the number N of claims occurring over a specified time period, for example, a year. The total claim amount of an insurance portfolio is then given by
>
> $$S = \sum_{i=1}^{N} X_i.$$
>
> Typically, the random variable N is assumed to have Poisson distribution. For the purposes of this paper, we can avoid such a restriction. However, we assume that the expected value of N, say λ, is known. It is also assumed that $\mathbf{E}(X_1)$ exists. □

1.1 Proportional Reinsurance

This is a common form of reinsurance for claims of "moderate" size. Here simply a fraction $p \in (0, 1)$

of each claim (hence the pth fraction of the whole portfolio) is covered by the reinsurer. Thus the reinsurer pays for the amount pS whatever the size of the claims.

Further, the parameter of interest, i.e., the expected value of the total claim amount, can be rewritten as follows,

$$R = \mathbf{E}\left\{ p \sum_{i=1}^{N} X_i \right\} = p\,\lambda\,\mathbf{E}(X_1). \qquad (1)$$

Note that the only unknown term which has to be estimated in this expression is $\mathbf{E}(X_1)$. Hence, the problem of estimation of R can be reduced to the estimation of $\mathbf{E}(X_1)$.

1.2 Pareto Models

In the context of reinsurance (insurance), heavy-tailed distribution functions F are a quite realistic choice. A useful and tractable parametric model with relatively high probability in the upper tail is the Pareto distribution $\mathrm{Pa}(\sigma, \alpha)$ having cdf

$$F(x) = 1 - \left(\frac{\sigma}{x}\right)^{\alpha}, \quad x \geq \sigma, \qquad (2)$$

defined for $\sigma > 0$ and $\alpha > 0$. Here we assume that the scale parameter σ is known. In actuarial applications, (2) with σ known is appropriate when losses or claims below a certain level are not relevant, as for example when a deductible applies.

On the other hand, it is unrealistic to expect that individual claim sizes will follow *exactly* $\mathrm{Pa}(\sigma, \alpha)$ model. Therefore, we investigate the behavior of estimators of the proportional reinsurance premium when claims are approximately Pareto-distributed.

1.2.1 Approximate Pareto

We consider the contamination model of form

$$F = (1 - \varepsilon)\,\mathrm{Pa}(\sigma, \alpha) + \varepsilon\, G, \qquad (3)$$

where ε represents the probability that a sample observation comes from the distribution G instead of $\mathrm{Pa}(\sigma, \alpha)$, and G is an arbitrary distribution function labeled as "contamination."

For Monte Carlo simulations, a choice of $G = \mathrm{Pa}(\sigma^*, \alpha)$ offers a flexible option of the contamination model (3). This approach was used in Brazauskas (1999) and it allows us to generate several approximate Pareto models:

- model (3) with *upper* outliers (the case of σ^* many times greater than σ, $\sigma^* \gg \sigma$),
- model (3) with *inliers* (the case of σ^* slightly greater than σ).

(Although it is not of concern in this paper, but the case $\sigma^* \ll \sigma$ results in model (3) with *lower* outliers.)

1.3 Estimation Methodologies

It was suggested by Hipp (1996) that all types of reinsurance premiums along with other important features (e.g., *mean, variance, coefficient of variation, skewness, kurtosis,* the *mean excess function,* and the *loss elimination ratio*) of an underlying distribution F can be represented as functionals of F, i.e., $\mathbf{H}(F)$. This idea serves as a motivation for several estimation methodologies.

1.3.1 Empirical Nonparametric Approach

This approach is motivated by the fact that "in nonparametric estimation problems, a parameter of interest $\mathbf{H}(F)$ of the unknown distribution F is frequently estimated by $\mathbf{H}(\hat{F}_n)$, with $\hat{F}_n$ the *empirical distribution function* based on a sample of n iid observations $X_1, \ldots, X_n$ with distribution F." (For examples of functionals $\mathbf{H}$, see Hipp (1996). For a more comprehensive treatment of premium principles, see Gerber (1979).)

1.3.2 Robust Parametric Approach

In parametric modeling of F (say, F depends on parameters δ_1 and δ_2) one represents functionals $\mathbf{H}(F)$ as explicit functions of the parameters δ_1 and δ_2 and obtains estimates $\widehat{\mathbf{H}(F)}$ by substitution of $\hat{\delta}_1$ for δ_1 and $\hat{\delta}_2$ for δ_2. The major disadvantage of such an approach is that it depends directly upon parametric assumptions, which may be of questionable validity, i.e., it is *nonrobust.* We suggest the following solution: use robust estimators of parameters instead of standard ones. (See Brazauskas and Serfling (2000) for discussion.)

1.4 Efficiency versus Robustness

For a parametric model, the Maximum Likelihood Estimator (MLE) proves to be highly efficient (at least for large sample size n). Typically, however, it is nonrobust: highly sensitive to departures of the actual data from the assumed parametric model. Competing (in this context, robust) estimators are designed to perform well over a specified range of departures from the "ideal" model, necessarily achieve their "robustness" at the expense of some sacrifice of efficiency relative to performance when the "ideal" model is indeed fully accurate. Thus robustness may be viewed as a kind of "insurance" purchased for a "premium" consisting of some loss of efficiency at the assumed model. In selecting an estimator, therefore, we seek a good trade-off between "efficiency" and "robustness," i.e., a high degree of "protection" in return for a given "premium." Let us now introduce precise notions of these criteria.

1.4.1 Efficiency

In terms of its optimum asymptotic variance, the MLE provides a quantitative benchmark for efficiency considerations. Thus we characterize efficiency of competing estimators of the proportional reinsurance premium via

$$\mathrm{ARE}(\hat{R}, \hat{R}_{\mathrm{ML}}) = \frac{\text{asymptotic variance of MLE}}{\text{asymptotic variance of } \hat{R}},$$

where "asymptotic variance" denotes either exact asymptotic variance or the variance parameter in an asymptotic distribution, ARE stands for the *asymptotic relative efficiency*, and $\hat{R}_{\mathrm{ML}}$ is defined in Section 2.1. (Also, our definition of ARE is based on "asymptotic variance." Note that it is a valid approach, because all the estimators under consideration are asymptotically *unbiased.* Otherwise, we should replace term "asymptotic variance" by "asymptotic mean square error.")

1.4.2 Robustness

A popular and effective criterion for robustness of an estimator is its (finite-sample) *breakdown point*, loosely defined as the largest proportion of sample observations which may be corrupted without corrupting the estimator beyond any usefulness. It provides an index valid over a broad and *nonspecific* range of possible sources of contaminating data.

In the reinsurance context, protection against upper contamination is more important than protection against inliers or lower outliers. Thus we define

> *Upper Breakdown Point* (UBP): the largest proportion of *upper* sample observations which may be taken to an upper limit without taking the estimator to an uninformative limit not depending on the parameter being estimated.

Also, it is important to decide upon the level of contamination (i.e., probability ε). For this we are guided by the following principle.

> *The "$\varepsilon(n) \to \varepsilon_\infty$ principle."* The proportion $\varepsilon(n)$ of "contamination" that we wish to protect against should be allowed to decrease to some small limit ε_∞ (e.g., $\varepsilon_\infty = 0.05$) as $n \to \infty$. Otherwise, having an abundance of data when n is large, we should estimate the entire mixture model, not the "ideal" model. Also, in the case of many outliers in a large data set, the problem of identification of outliers reduces to a classification problem.

See discussion by Jaeckel (1971), for example.

On the basis of these criteria various estimators of R are compared. In particular, empirical nonparametric, trimmed mean type, and generalized median type estimators are considered. In Section 2 we briefly introduce the estimators in the study and evaluate their UBP and ARE. Comparisons and conclusions are presented in Section 3.

2 The Estimators

Consider a sample of claims $X_1, \dots, X_n$ from the $\text{Pa}(\sigma, \alpha)$ model as described by (2) and denote the ordered sample values by

$$X_{n1} \le X_{n2} \le \cdots \le X_{nn}$$

and the sample mean by $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$.

2.1 Maximum Likelihood Estimator

As is well-known (e.g., Arnold (1983)), for model (2) the MLE of α is given by

$$\hat{\alpha}_{\text{ML}} = \frac{1}{n^{-1} \sum_{i=1}^{n} \log X_i - \log \sigma}.$$

Additionally, the expected value of Pareto distribution with cdf (2) is

$$\frac{\sigma \, \alpha}{\alpha - 1}. \tag{4}$$

Therefore, using (1) and (4) we find that the corresponding MLE of R is given by

$$\hat{R}_{\text{ML}} = p\lambda \, \frac{\sigma \, \hat{\alpha}_{\text{ML}}}{\hat{\alpha}_{\text{ML}} - 1}. \tag{5}$$

Since the estimator of α involves a *mean* of the $\log X_i$'s, it has UBP = 0. Therefore, $\hat{R}_{\text{ML}}$ also has UBP = 0 and it is classified as *nonrobust*.

Nevertheless, the MLE provides a benchmark for *efficiency* considerations. Namely, $\hat{\alpha}_{\text{ML}}$ *is asymptotically normal with mean* α *and variance* α^2/n, which we denote

$$\hat{\alpha}_{\text{ML}} \ \text{ is AN} \left(\alpha \, , \, \frac{\alpha^2}{n} \right).$$

Further, we have to find asymptotic distribution of $\hat{R}_{\text{ML}}$ which is the transformation of $\hat{\alpha}_{\text{ML}}$ described by (5). A straightforward application of asymptotic distribution theory of transformations leads to the fact that

$$\hat{R}_{\text{ML}} \ \text{ is AN} \left(p\lambda \, \frac{\alpha \, \sigma}{\alpha - 1} \, , \, (p\lambda)^2 \, \frac{\alpha^2 \sigma^2}{(\alpha - 1)^4} \, \frac{1}{n} \right).$$

See Serfling (1980), Chapter 3, for example.

2.2 Nonparametric Estimator

Replace F by $\hat{F}_n$ in $\mathbf{E}(X_1)$. This leads to

$$\widehat{\mathbf{E}(X_1)} = \int x \, d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}.$$

Hence, the empirical nonparametric estimator of the proportional reinsurance premium R is given by

$$\hat{R}_{\text{N}} = p\lambda \, \overline{X}.$$

Clearly, $\overline{X}$ has UBP = 0. Consequently, $\hat{R}_{\text{N}}$ also has UBP = 0 and thus is *nonrobust*.

For efficiency considerations we have that

$$\hat{R}_{\text{N}} \ \text{ is AN} \left(p\lambda \, \frac{\alpha \, \sigma}{\alpha - 1} \, , \, (p\lambda)^2 \, \frac{\alpha \, \sigma^2}{(\alpha - 1)^2 (\alpha - 2)} \, \frac{1}{n} \right)$$

and therefore

$$\text{ARE}(\hat{R}_{\text{N}}, \hat{R}_{\text{ML}}) = \frac{\alpha(\alpha - 2)}{(\alpha - 1)^2},$$

which are valid provided $\alpha > 2$.

2.3 Trimmed Mean Estimators

For specified β_1 and β_2 satisfying $0 \le \beta_1 \, , \beta_2 < 1/2$, a trimmed mean is formed by discarding the proportion β_1 lowermost observations and the proportion β_2 uppermost observations and averaging the remaining ones in some sense. In particular,

$$\hat{\alpha}_{\text{T}} = \left(\sum_{i=1}^{n} c_{ni} \, (\log X_{ni} - \log \sigma) \right)^{-1},$$

with

$$c_{ni} = \begin{cases} 0, & 1 \le i \le [n\beta_1]; \\ 1/d, & [n\beta_1] + 1 \le i \le n - [n\beta_2]; \\ 0, & n - [n\beta_2] + 1 \le i \le n, \end{cases}$$

where $[\cdot]$ denotes "greatest integer part," and

$$d = d(\beta_1, \beta_2, n) = \sum_{j=[n\beta_1]+1}^{n-[n\beta_2]} \; \sum_{i=0}^{j-1} (n - i)^{-1}.$$

These estimators correspond to the trimmed mean estimators introduced and studied by Kimber (1983a,b) for the equivalent problem of estimation of $\theta = \alpha^{-1}$ in the two-parameter exponential model

$E(\mu, \theta)$ with μ known. The above c_{ni}'s are a choice making $\hat{\theta}_T = \hat{\alpha}_T^{-1}$ mean-unbiased for θ.

It follows from (1) and (4) that the trimmed mean estimators of R are given by

$$\hat{R}_T = p\lambda \frac{\sigma \hat{\alpha}_T}{\hat{\alpha}_T - 1}.$$

We see from the definition of trimmed mean estimators that $\hat{\alpha}_T$ is unaffected by proportion β_2 of uppermost observations. This implies that $\hat{\alpha}_T$ (consequently, $\hat{R}_T$) has UBP $= \beta_2$. Therefore, the trimmed mean estimators $\hat{R}_T$ are classified as *robust*.

For efficiency considerations we have that

$$\hat{R}_T \text{ is AN}\left(p\lambda \frac{\alpha\sigma}{\alpha - 1}, (p\lambda)^2 \frac{\alpha^2\sigma^2}{(\alpha-1)^4} \frac{D_{\beta_1,\beta_2}}{n}\right),$$

with D_{β_1,β_2} computable following general methods for L-statistics in Serfling (1980), Chapter 8. It follows immediately that

$$\text{ARE}(\hat{R}_T, \hat{R}_{ML}) = \frac{1}{D_{\beta_1,\beta_2}}.$$

2.4 Generalized Median Estimators

Generalized median (GM) statistics are defined by taking the median of the $\binom{n}{k}$ evaluations of a given kernel $h(x_1, \ldots, x_k)$ over all k-sets of the data. In Brazauskas and Serfling (2000), such an estimator was considered for the parameter α in the case of σ known:

$$\hat{\alpha}_{GM} = \text{Median}\{h(X_{i_1}, \ldots, X_{i_k}; \sigma)\},$$

with a particular choice of kernel h:

$$h(x_1, \ldots, x_k; \sigma) = \frac{1}{C_k} \frac{1}{k^{-1}\sum_{j=1}^{k} \log x_j - \log \sigma},$$

where C_k is a multiplicative median-unbiasing factor, i.e., chosen so that in each case the distribution of $h(X_{i_1}, \ldots, X_{i_k}; \sigma)$ has *median* α. Values of C_k are provided in the following table. (For $k > 8$, C_k is given by a very accurate approximation, $C_k \approx k/(k - 1/3)$.)

Table 1. Values of C_k, for $k = 2:8$.

			k			
2	3	4	5	6	7	8
1.19	1.12	1.09	1.07	1.06	1.05	1.04

Similarly to previous derivations, it follows from (1) and (4) that the GM estimators of the proportional reinsurance premium R are given by

$$\hat{R}_{GM} = p\lambda \frac{\sigma \hat{\alpha}_{GM}}{\hat{\alpha}_{GM} - 1}.$$

A detailed study of robustness and asymptotic distribution theory of the generalized median estimators is available in Brazauskas (1999). It was found that these estimators are *robust*, thus endowing estimators $\hat{R}_{GM}$ with good robustness properties. The UBP of $\hat{\alpha}_{GM}$ (consequently, $\hat{R}_{GM}$) is given by the following formula:

$$\text{UBP} = 1 - 2^{-1/k}.$$

For efficiency considerations we have that

$$\hat{R}_{GM} \text{ is AN}\left(p\lambda \frac{\alpha\sigma}{\alpha - 1}, (p\lambda)^2 \frac{\alpha^2\sigma^2}{(\alpha-1)^4} \frac{\gamma_k}{n}\right),$$

which implies that

$$\text{ARE}(\hat{R}_{GM}, \hat{R}_{ML}) = \frac{1}{\gamma_k}.$$

3 Comparisons

In the table below, several of the estimators of the proportional reinsurance premium R considered above are compared from the standpoint of *efficiency versus robustness*.

Estimator	ARE	UBP
MLE	1	0
$\hat{R}_N$ $(2 < \alpha \leq 2.5)$	≤ 0.56	0
$\hat{R}_T$, $\beta_1 = \beta_2 = .25$	0.67	0.25
$\hat{R}_T$, $\beta_1 = \beta_2 = .20$	0.72	0.20
$\hat{R}_T$, $\beta_1 = \beta_2 = .15$	0.78	0.15
$\hat{R}_T$, $\beta_1 = \beta_2 = .10$	0.85	0.10
$\hat{R}_T$, $\beta_1 = \beta_2 = .05$	0.92	0.05
$\hat{R}_{GM}$, $k = 2$	0.78	0.29
$\hat{R}_{GM}$, $k = 3$	0.88	0.21
$\hat{R}_{GM}$, $k = 4$	0.92	0.16
$\hat{R}_{GM}$, $k = 5$	0.94	0.13
$\hat{R}_{GM}$, $k = 10$	0.98	0.07

The following conclusions are quite evident:

- The "empirical nonparametric" estimator is neither robust nor efficient and thus is not competitive.

- The "trimmed mean" type estimators offer a good trade-off between ARE and UBP, but they are improved upon by the "generalized median" type estimators. For example, $\hat{R}_T$ for $\beta_1 = \beta_2 = .20$, with ARE = 0.72 and UBP = 0.20, is dominated by $\hat{R}_{GM}$ for $k = 2$, with ARE = 0.78 and UBP = 0.29, and by $\hat{R}_{GM}$ for $k = 3$, with ARE = 0.88 and UBP = 0.21. Likewise, $\hat{R}_T$ for $\beta_1 = \beta_2 = .05$, with ARE = 0.92 and UBP = 0.05, is dominated by $\hat{R}_{GM}$ for $k = 4$, with ARE = 0.92 and UBP = 0.16, and by $\hat{R}_{GM}$ for $k = 5$, with ARE = 0.94 and UBP = 0.13, and also by $\hat{R}_{GM}$ for $k = 10$, with ARE = 0.98 and UBP = 0.07.

The superiority of the generalized median estimators can be explained by the following principle.

> *Smoothing* followed by *medianing* yields a very favorable combination of *efficiency* and *robustness.*

That is, the two-step procedure

1. "Smooth" the data by taking a function of several observations at a time over all subsets of the data and replace the data with these function evaluations, and

2. Take the median of these evaluations,

leads to a statistic which possesses a favorable combination of efficiency and robustness.

4 Acknowledgments

The author is very appreciative of valuable comments provided by Professor E. W. Frees. Also, support by grants from the Casualty Actuarial Society and the Society of Actuaries, with administrative support from the Actuarial Education and Research Fund, is gratefully acknowledged.

References

[1] Arnold, B. C. (1983). *Pareto Distributions.* International Cooperative Publishing House. Fairland, Maryland.

[2] Brazauskas, V. (1999). *Robust and Nonparametric Statistical Methods for Pareto Distribution Tail Index Estimation, with Actuarial Science Applications.* Ph.D. dissertation, University of Texas, Dallas.

[3] Brazauskas, V., and Serfling, R. (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal.* (to appear)

[4] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[5] Gerber, H. U. (1979). *An Introduction to Mathematical Risk Theory.* Huebner Foundation for Insurance Education, University of Pennsylvania.

[6] Hipp, C. (1996). The delta-method for actuarial statistics. *Scandinavian Actuarial Journal*, 1, 79–94.

[7] Jaeckel, L. A. (1971). Robust estimates of location: symmetry and asymmetric contamination, *Annals of Mathematical Statistics*, **42**, 1020–1034.

[8] Kimber, A. C. (1983a). Trimming in gamma samples. *Applied Statistics*, **32**, 7–14.

[9] Kimber, A. C. (1983b). Comparison of some robust estimators of scale in gamma samples with known shape. *Journal of Statistical Computation and Simulation*, **18**, 273–286.

[10] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.