Study 4

Extreme Value Theory in Actuarial Consulting: Windstorm Losses in Central Europe

Dietmar Pfeifer, University of Hamburg and AON Re Jauch & Hübener, Hamburg

Abstract. Scientific consulting especially in reinsurance brokerage has become of growing importance in the recent years, not only in competition with attempts to transfer insurance risk to the capital markets via catastrophe options and other derivative instruments. This study shows that and how extreme value theory can be fruitfully applied to classical and modern problems in reinsurance. As an example, central European windstorm losses which exhibit both temporal as well as spatial aspects are analyzed using some tools of Xtremes.

1. Introduction

A classical field of reinsurance is certainly in the domain of all kind of natural catastrophes like wind and hailstorms, flooding, earthquakes etc. Since the damages occurring here usually are rather costly it is necessary to find good mathematical models which describe the underlying random mechanisms in an appropriate way. One must either calculate a classical PML (Probable Maximum Loss, usually a high quantile of the loss distribution) or perform simulation studies for estimating the performance of typical reinsurance tools such as XL (eXcess–of–Loss) or SL (Stop–Loss) contracts with and without reinstatements, or combinations thereof. See also Chapter 11 in this book for further aspects.

Nice discussions on the use of mathematical methods in the (re)insurance industry, especially extreme value theory, can also be found, e.g., in Embrechts et al. (1997), Hipp (1999) or Woo (1999). Such methods are also useful companions of physical investigations and simulations based on meteorological approaches as pursued by Applied Insurance Research, for example (see Clark (1997)). Indeed, actuarial consulting experience has shown that many of the results of “merely” statistical analyses of catastrophic risks coincide quite well with the sophisticated findings and forecasts of such physical models.

In this study we want to demonstrate the practical applicability of extreme value statistics by analyzing a data set which in its original form has been used for actuarial consulting of a larger group of clients. In order to preserve the necessary discretion the data (stored in aon-re.dat) have been pre–processed, e.g. some economically reasonable detrending was performed, and the original currency was converted to U.S. $.

The data set comprises adjusted claims from windstorm losses in central Europe during the years 1980 to 1997, in three geographically adjacent zones in the direction from west to east. Since in Europe the wind direction is typically like this, the data exhibit a strong spatial dependence structure (see Fig. 1 below) while the temporal dependence over the years is negligible. Both univariate as well as multivariate tools from Xtremes will be applied to this data set in the sequel.

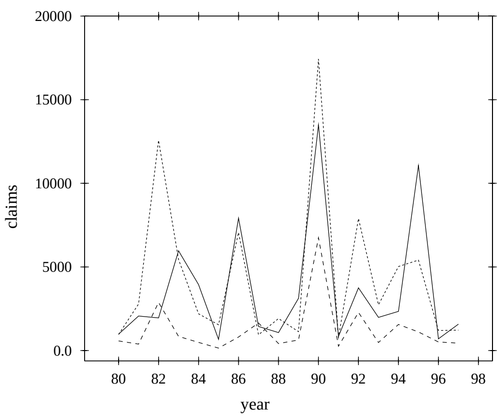

FIG. 1. Time series of windstorm losses (1000 U.S. $); zones 1 (solid), zone 2 (dotted) and zone 3 (dashed).

2. Univariate Data Analysis

In a first step the three zones were each analyzed separately. Due to the comparably small number of 18 observations per zone it was desirable to include all available data in the model fitting procedure. Although it is in general difficult to find a class of distributions which provides a good fit over the whole range of observations (see Beirlant et al. (1996) for a discussion of this point) it seems that the class of Fréchet distributions is often appropriate in modelling windstorm losses (see, e.g., Pfeifer (1997) or Rootzén and Tajvidi (1997)).

The following ML estimates for the parameters α (shape) and σ (scale) were found for the three zones (cf. Table 1).

	zone 1	zone 2	zone 3
$\hat{\alpha}$	1.35418	1.32626	1.27924
$\hat{\sigma}$	1553.05	1791.01	506.45

TABLE 1. Univariate parameter estimates for the Fréchet class.

The estimates for the shape parameter α indicate heavy upper tails of the windstorm losses.

The following illustrations show the corresponding Q–Q plots for the three zones and the empirical vs. the parametric quantile function of zones 1 and 3 (the zone 2 was omitted from the graph for the sake of a better visibility).

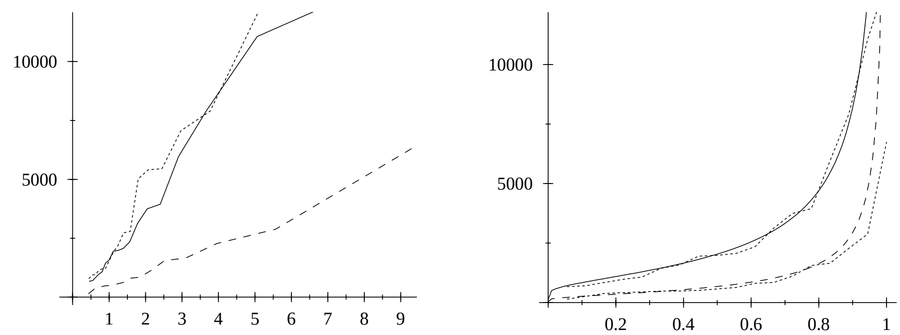

FIG. 2. (left.) Q–Q plots for windstorm losses: zone 1 (solid), zone 2 (dotted), zone 3 (dashed). (right.) Parametric quantile functions: zone 1 (solid), zone 3 (dashed); pertaining empirical quantile functions (dotted).

It is remarkable that in particular in the lower and central part of the distributions the fit is quite satisfactory.

This can also be seen from the following illustrations. In Fig. 3 (left) we compare the parametric density of windstorm losses from zone 1 with a kernel density estimate (uniform kernel, global bandwidth $b = 1470$, local bandwidth $= 760$ at minimum data point $x_{1:18} = 669$). On the left–hand side one finds the corresponding illustrations for zone 3 (uniform kernel, global bandwidth $b = 550$, local bandwidth $= 107$ at minimum data point $x_{1:18} = 146$).

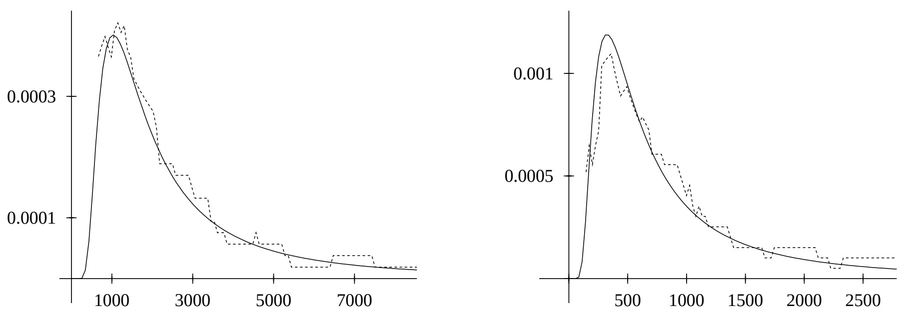

FIG. 3. Estimated parametric density vs. kernel density estimate, zone 1 (left) and zone 3 (right).

On the basis of the parameter estimates $\hat{\alpha}$ and $\hat{\sigma}$, PML estimates for a return period of T years can be represented as

$$\mathrm{PML}(T) = \hat{\sigma}\left(-\ln\left(1 - 1/T\right)\right)^{-1/\hat{\alpha}} \;\approx\; \hat{\sigma} T^{1/\hat{\alpha}}$$

for large values of T, the approximation being quite good for $T \geq 20$ already as it is shown in the subsequent Table 2.

TABLE 2. Tabulated PMLs for different return periods

	zone 1		zone 2		zone 3	
T	PML(T)	approx.	PML(T)	approx.	PML(T)	approx.
20	13923	14188	16815	17142	5162	5267
50	27705	27912	33948	34207	10696	10780
100	46396	46568	57471	57689	18461	18533
200	77551	77694	97107	97291	31800	31862
500	152732	152845	193996	194142	65165	65216
1000	254910	255004	327291	327414	112075	112118

The maximum observed losses in zones 1 to 3 within 18 years were 13496, 17430 and 6749 resp., all occurring in the same year 1990 (cf. Fig. 1). This is in coincidence with the 20 year PMLs shown above. It is also remarkable that the estimates for the shape parameter α are all around the value of 1.3 which shows a great constancy of the "dangerousness" of the loss distribution even over a larger geographical region. On the other hand, an investigation using the pot approach does not provide useful information about α due to the small number of data.

Of course, results as above must be considered with great care, especially if PML estimates for a large return period are to be considered and the data basis is small. They should always be accompanied by further statistical investigations, not only on the basis of extreme value theory. It must however be kept in mind, that in reinsurance practice, return periods between 200 and 500 years are a common basis for XL and SL–contracts nowadays. We refer to the corresponding discussion in Hipp (1999) and Berz (1999).

3. Multivariate Analysis

Due to the geographical dependencies of windstorm losses the present data set is a good candidate for a multivariate extreme value analysis, in particular w.r.t. the Marshall–Olkin, Gumbel–McFadden und Hüsler–Reiss models. In this section, we shall concentrate on these aspects in detail, using zones 1 and 2 only for the graphical analysis. Other comparisons can easily be performed in a similar manner.

The illustration in Fig. 4 shows a superposition of a scatterplot of the data, and contourplots of a kernel density and a bivariate extreme value density. The kernel density is based on bandwidths 2100 and 1900 and a direction $\varphi = 1.1$ which corresponds to the observed direction of the data (obtained with the Visualize menu for multivariate data sets in Xtremes). The estimated extreme value density will be explained below.

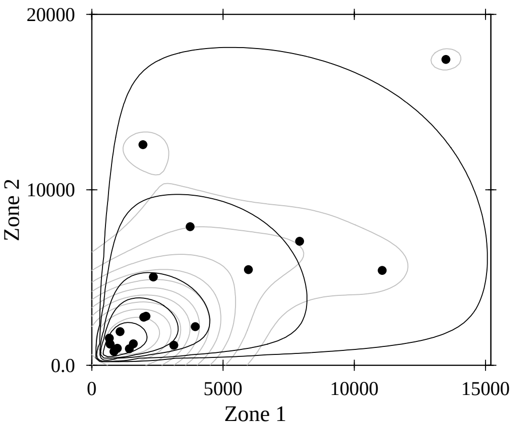

FIG. 4. Scatterplot of zone 2 plotted against zone 1 superposed with the contour plots of a bivariate kernel density (dotted) and a bivariate extreme value density (solid).

The contour plot of the kernel density represents the data in a reasonable manner with the exception of the isolated data point (representing the losses in the year 1990) in the north–east corner of the graph. One may argue that especially larger losses are highly dependent. This is mainly due to the fact that severe storms in central Europe usually affect several states simultaneously (such as in 1990 with storms Daria and Vivien, see Berz (1999)).

The following Table 3 contains the estimated dependence parameters for the models mentioned above (see also Chapter 9 in this book), obtained with the Estimate menu for the MAX–domain in Xtremes. The univariate margins are the Fréchet distributions pertaining to the estimated parameters in Table 1. In all three models, the estimated parameters reveal the strong dependence of loss data between adjacent geographical zones, as expected.

The parametric contourplot in Fig. 4 belongs to the estimated Gumbel–McFadden density with dependence parameter $\lambda = 1.903$.

TABLE 3. Matrices of estimated dependence parameters λ.

Marshall–Olkin model   Gumbel–McFadden model   Hüsler–Reiss model

	zone 1	zone 2
zone 2	0.585	
zone 3	0.578	0.427

	zone 1	zone 2
zone 2	1.858	
zone 3	1.517	1.782

	zone 1	zone 2
zone 2	1.903	
zone 3	2.856	1.462

4. Conclusion

Extreme value theory repeatedly turns out to be one of the most powerful and meanwhile widely accepted statistical tools in the (re)insurance industry, at least w.r.t. univariate statistics. However, also multivariate statistical procedures will gain more importance in the near future, e.g. in connection with ratings of insurance bonds or hedging strategies in the framework of the so–called alternative risk transfer (see Hipp (1999)). For this purpose, suitable copula models for insurance losses and risks play an important role. Regrettably, it is still an open question how such models can efficiently be identified, in particular when the data sets are small (see also Dall'Aglio et al. (1991) for a careful discussion of this topic).

References

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

Berz, G. (1999). Naturkatastrophen an der Wende zum nächsten Jahrhundert—Trends, Schadenpotentiale und Handlungsoptionen der Versicherungswirtschaft. Zeitschrift für die gesamte Versicherungswissenschaft 2/3, 427–442.

Clark, K.M. (1997). Current potential impact of hurricane variability on the insurance industry. In: H.F. Diaz, R.S. Pulwarty (Eds.): Hurricanes—Climate and Socioeconomic Impacts, Springer, N.Y., 273–283.

Dall'Aglio, G., Kotz, S. and Salinetti, G. (Eds.) (1991). Advances in Probability Distributions with Given Marginals. Kluwer, Dordrecht.

Embrechts, P., Klüppelberg, C. and Mikosch, Th. (1997). Modelling Extremal Events for Insurance and Finance. Springer, N.Y.

Hipp, Ch. (1999). Risikomanagement von Naturkatastrophen: helfen mathematische Methoden? Zeitschrift für die gesamte Versicherungswissenschaft 2/3, 443–456.

Pfeifer, D. (1997). A statistical model to analyse natural catastrophe claims by means of record values. Proceedings of the XXVIIIth International ASTIN Colloquium, Cairns, August 10–12, 1997, The Institute of Actuaries of Australia, Sydney, 45–57.

Rootzén, H. and Tajvidi, N. (1997). Extreme value statistics and wind storm losses: a case study. Scand. Act. J., 70–94.

Woo, G. (1999). The Mathematics of Natural Catastrophes. Imperial College Press, London.