Insurance: Mathematics and Economics 44 (2009) 345–356

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Univariate and bivariate GPD methods for predicting extreme wind storm losses

Erik Brodin *, Holger Rootzén

Department of Mathematical Sciences, Chalmers University of Technology, SE-412 96 Göteborg, Sweden
Göteborg University, SE-412 96 Göteborg, Sweden

ARTICLE INFO

Article history:
Received September 2008
Accepted 5 November 2008

Keywords:
Extreme value statistics
Generalized Pareto distribution
Likelihood prediction intervals
Peaks over threshold
Trend analysis
Wind storm losses

ABSTRACT

Wind storm and hurricane risks are attracting increased attention as a result of recent catastrophic events. The aim of this paper is to select, tailor, and develop extreme value methods for use in wind storm insurance. The methods are applied to the 1982–2005 losses for the largest Swedish insurance company, the Länsförsäkringar group. Both a univariate and a new bivariate Generalized Pareto Distribution (GPD) gave models which fitted the data well. The bivariate model led to lower estimates of risk, except for extreme cases, but taking statistical uncertainty into account the two models lead to qualitatively similar results. We believe that the bivariate model provided the most realistic picture of the real uncertainties. It additionally made it possible to explore the effects of changes in the insurance portfolio, and showed that loss distributions are rather insensitive to portfolio changes. We found a small trend in the sizes of small individual claims, but no other trends. Finally, we believe that companies should develop systematic ways of thinking about "not yet seen" disasters.

1. Introduction

In January 2005 the wind storm Gudrun (Erwin in the German weather service terminology) struck the southern part of Sweden and caused widespread damage to infrastructure and forests. The loss to the largest Swedish insurance company, the Länsförsäkringar group, was in excess of 2.9 billion SEK. The damage can by no means be compared with the devastation caused by hurricane Katarina, but still Gudrun was one of the worst wind storms to hit Sweden for centuries, and the economic losses were large. Windstorms continue to be a serious threat to Sweden and Europe in general — at the time of writing a first draft of this paper, Germany was paralyzed by wind storm Kyrill.

The aim of this paper is methodological, (i) to take benefit of the rapid development of Extreme Value Statistics (EVS), in particular the new multivariate Generalized Pareto Distributions (GPD), to choose best practice methods for analysis of wind storm losses, (ii) to experiment with methods for assessing the impact of changes in the insurance portfolio, and (iii) to make a first attempt at approaching new methodological problems for EVS which are raised by wind storm loss data. Prediction of the sizes of future very large losses, often presented in terms of Probable Maximum Loss, PML, are at the center of attention, and we try to provide a basis for evaluation of reinsurance strategies and for calculation of regulatory demands. We illustrate and test the methods by analyzing a big proprietary data set, the Länsförsäkringar windstorm loss data for 1982–2005, which Länsförsäkringar kindly has given us access to.

As a short summary of results, we concluded that a bivariate model gave the most realistic predictions for this data set, and that it seemed to give a reasonable picture of the effects of portfolio changes. Taking statistical uncertainty into account, the standard univariate model lead to qualitatively similar results as the bivariate one. We also discuss issues of estimation, computation, and model control for the bivariate model.

A further question was what can be learned from Gudrun. In particular, does Gudrun drastically alter earlier perceptions of wind storm risk? Was Gudrun a complete surprise, or was she in line with what could be expected? Briefly, we found that Gudrun was larger than what was expected because of the not previously experienced very big forest losses, but that the size of the loss still was not a complete surprise.

The end goal of risk assessment is good estimates of probability distributions, typically expressed as the PML for future losses. Traditionally this has been approached through point estimates of quantiles. The statistical uncertainty of these could then be assessed by confidence intervals. However, in the final evaluation, how should one weigh together the risk level associated with the quantile with the significance level of the confidence interval? A way to solve this problem is to use prediction intervals, since such intervals takes both the uncertainty of the world and statistical uncertainty into account. A very useful development has been Hall et al. (1999, 2002) which provides prediction intervals for the present setting. In contrast to the traditional approach we in this paper use prediction intervals for the univariate analysis.

* Corresponding author.
E-mail addresses: ebrodin@math.chalmers.se (E. Brodin), rootzen@math.chalmers.se (H. Rootzén).

0167-6687/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2008.11.002

Unfortunately such intervals are not yet available for the bivariate analysis.

EVS presumes asymptotic Generalized Pareto behavior of tails. This in fact very often holds. However, there is one important exception: situations when there are two (or more) different causes of large losses, and one of these is much more serious but also rarer. There are then three possible cases. The first one is that there is much data. The rare but large losses will then dominate the empirical extreme tails of the overall loss distribution, and standard univariate EVS will work as intended. The second one is when there is little data so that one doesn't have any experience of the rare but serious cause. In this case statistics can show that risks are at least of a certain size, but can't help in giving upper bounds for the risk. For this case a systematic and serious effort to identify and evaluate risks which aren't represented in the data is important. We stress that companies should develop systematic qualitative ways of thinking about such "not yet seen" kinds of disasters — the unexpected large forest loss from Gudrun provides an example of the importance of this.

The third, intermediate, case is that there has been a few occurrences of the serious eventuality, but that the less serious one dominates the empirical tail distribution of overall loss. A bivariate analysis may be appropriate for such cases.

For the present wind storm insurance problem there is indeed two such different loss mechanisms: for most wind storm events, damage to buildings, etc., dominate completely, but for the most serious event, Gudrun, the damage to forest was 2.6 times larger than the building damage. We throughout compare with the Rootzén and Tajvidi (1997, 2001) analysis of the 1982–1993 Länsförsäkringar data. There was little forest losses in this data set, corresponding to case two above. However, for the 1982–2005 data studied in this paper we are in the third, intermediate, situation.

There is a considerable literature on wind storms. The two Rootzén and Tajvidi (1997, 2001) papers cited above contain EVS analyses of the Länsförsäkringar wind storm losses for 1982–1993 and were a starting point for this paper. The first one discussed estimation of loss quantiles for various risk levels and time periods and also concluded that there were no significant trends in the cumulated loss sizes, although there was an increasing trend in the size of small claims. The second one argued that the link between Swedish meteorological data and loss sizes is too weak to make it practical to use it to predict losses.

Among other papers of special interest to us are the detailed analyses Valinger et al. (2006) and SMHI (2005) of Gudrun, and the report Holmberg (2005) which lists all severe wind storms in Sweden during the last 210 years. From the latter report one can learn that there were storms with wind speeds comparable to Gudrun's in December 1902 and in September/November 1969. However, the economic losses were smaller because of the higher cost of modern infrastructure, and also because the ground was not frozen at the time of the Gudrun storm, which contributed importantly to the amount of damage to forest — 75 million m^3 forest was lost in Gudrun, but only 35 million m^3 in the 1969 storm. For 1902 the amount of damage is not known. Lies (2000) argues forcefully that prices for wind storm reinsurance have been too low. Theoretical studies of wind storm and more general catastrophes insurance include Cossette et al. (2003), Jaffee and Russell (1996), Lescourret and Robert (2006) and references therein.

In Section 2 we identify storm events and make inflation and portfolio change adjustments to obtain a final wind storm loss event data base for 1982 to 2005. Section 3 makes a brief discussion of the univariate EVS methods used in this paper, and Section 4 contains the results of the univariate analysis, and in particular prediction intervals for PML and a trend analysis. In Section 5 we introduce the bivariate approach. The results of the bivariate analysis are presented in Section 6. We discuss an alternative presentation of results from heavy tailed risk analysis in Section 7. Section 8 contains the conclusions of this paper.

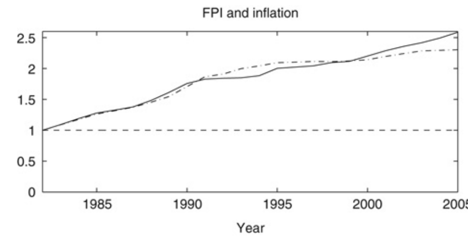

Fig. 2.1. Inflation in Sweden 1980–2005. FPI (solid), Consumers inflation (dash-dotted).

2. The windstorm loss data

The Länsförsäkringar 1982–2005 wind storm data base contains all individual storm related claims for household, company and farm insurance made to Länsförsäkringar, and includes a wealth of information on each claim. Here we have used the date when the damage occurred, the amount paid out by Länsförsäkringar, split up into building and forest claims, and the classification into household, company and farm insurance. In addition to claims for damage to buildings and to forest, there is a residual claim category. This residual category has about 1% of the total amount paid out and is excluded in our analysis.

The data base also contains a classification of the claims into wind storm events. This classification depends both on the loss (= total sum paid out) to Länsförsäkringar in a moving three day window and on the sizes of the sum of payments made by the individual regional companies which together make up the Länsförsäkringar group. In the present paper we use a different definition of storm events, see Section 2.3. However, for the larger events the Länsförsäkringar storms are very similar to our wind storm events.

In addition we have had use of a storm event data set from Rootzén and Tajvidi (2006). This is constructed in the same way as here but from an earlier version of the Länsförsäkringar database which covered the years 1982–1993. For 1987–1993 the storm events in this data set are virtually identical with the ones obtained from the present version of the data base. However, for 1982 to 1986 there are some differences. We believe that these may be caused by transcription and storage errors and that the earlier version is more accurate for 1982–1986, and hence have used the Rootzén and Tajvidi (2006) storm events for these years.

Analysis and prediction rely on stationarity. There are four obvious possible causes of non-stationarity: (a) inflation, (b) changes in the size and composition of the Länsförskringar insurance portfolio, (c) changes in building standards and changes in the propensity to build in more exposed places, and (d) changes in the wind storm climate. We discuss inflation adjustment in Section 2.1 and portfolio changes in Section 2.2 and Section 6. Further, (c) would result in trends in the amounts paid out in storm events, and (d) as trends in the yearly numbers of storms and/or in the severity of the storms. The existence (or not) of trends in our data is studied in Section 4.

2.1. Inflation adjustment

We have used the Swedish FPI (Faktorprisindex för byggnader), which can be downloaded from www.scb.se, to recompute all amounts into 2005 prices. The FPI index reflects the cost of building, including salaries. It is rather similar to the consumer inflation index, but there are some differences, see Fig. 2.1.

Parts of the claims are for forest damage. There doesn't seem to exist any suitable index for forest prices, so we have used the FPI also for this part. The forest loss caused by storm Gudrun completely dominated the forest damage in the other storms. Gudrun occurred in 2005 and hence there wasn't any need to adjust it for inflation.