Univariate and Bivariate GPD Methods for Predicting Extreme Wind Storm Losses

Erik Brodin Holger Rootzén*

Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden
ebrodin@math.chalmers.se and rootzen@math.chalmers.se

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 Generalised 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.

Keywords: Extreme Value Statistics, Generalized Pareto Distribution, Likelihood Prediction Intervals, Peaks over Threshold, Trend Analysis, Wind storm losses.

*Research supported by the Swedish Foundation for Strategic research through GMMC, the Gothenburg Mathematical Modelling Centre

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 inurance company, the Länsförsäkringar group, was in excess of 2,9 billon 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 learnt 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, often 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 provide prediction intervals for the present setting. In contrast to the traditional approach we in this paper use prediction intervals for the univariate analysis. 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, 2000) 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, 2000) 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 (2006) 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. Swiss Re (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) and 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.

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 (1996). 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 (1996) storm events for these years.

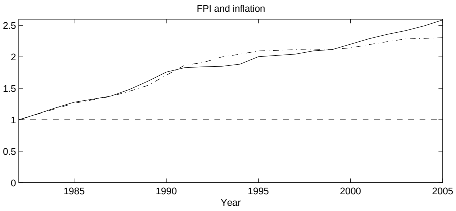

Figure 2.1: Inflation in Sweden 1980-2005. FPI (solid), Consumers inflation (dash-dotted).

Analysis and prediction rely on stationarity. There are four obvious possible causes of nonstationarity: 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 the next section 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 Figure 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.

2.2 Portfolio changes

Figure 2.2 shows the development of the number of LFAB insurance contracts for households, companies and farms from 1980 to 2005. One can clearly see the increase caused by the merger of Länsförsäkringar and Wasa (another Swedish insurance company) in 1998. For households and companies there was an increasing trend also before the merger, while the farm portfolio was rather stable, also after the merger.

Thus, presumably, the proportion of storm losses for household and company damage ought to have increased over the period, while the relative amount of farm damage should have decreased. This can also to some extent be observed in the right plot of Figure 2.3. Trends were more clear in plots similar to the left panel of Figure 2.3 made without exposure correction. The left plot of Figure 2.3 shows that a simple portfolio adjustment

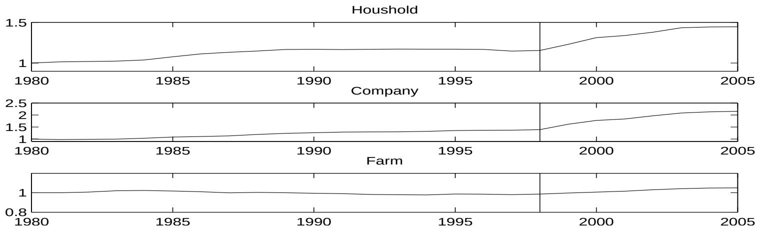

Figure 2.2: The numbers of contracts in the Länsförsäkringar insurance portfolios, divided by the number of contracts 1980. Note the different vertical scales. The vertical lines at year 1998 shows when Länsförsäkringar was merged with Wasa.

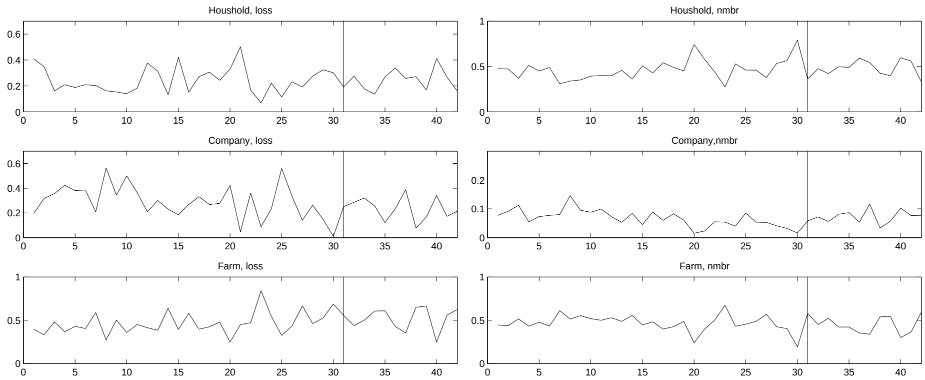

Figure 2.3: Left: Proportions of loss for portfolio adjusted storm events. Right: Proportions of number of claims for storm events, for storms with at least 120 claims. The storm Gudrun is not included. The vertical lines at year 1998 shows when Länsförsäkringar was merged with Wasa.

removes these trends.

This portfolio adjustment was made by multiplying the storm loss for a storm in year i by (number of insurance contracts of the exposure type in year 2005)/(number of insurance contracts of the exposure type in year i). In the sequel all loss amounts used in the analysis are adjusted for portfolio changes in this way.

This is a rather crude way of correcting for changes in exposure. In fact, how to make the best correction for heavy tailed data, such as the present windstorm loss data, seems like an interesting research question. A further discussion of effects of changes in portfolio size is made in the final bivariate analysis.

2.3 The storm loss data sets

From the Länsförsäkringar wind storm data base we constructed three storm loss data sets. The first one was used in a preliminary analysis to determine the sizes of the thresholds to be used in the final analyses. The second storm loss data set was used for the univariate analysis, and the third one for the bivariate analysis. The procedures we used to construct the data sets were inspired by both meteorological and reinsurance contract considerations.

The preliminary storm loss data set: We identified in total 104 storm events with losses larger than 1.5 million Swedish crowns (MSEK) (in 2005 prices, and corrected for portfolio changes) between 1982 and 2005 in the following way. First we computed the total loss in a moving three day window which was moved over the period 1987 - 2005. The three day periods when the aggregate loss exceeded 1.5 million where then considered as "potential storm" events. If a potential storm was isolated (taken to mean that there was at least two days between it and the next potential storm) this three day period was accepted as a storm event and put into the storm event data set. In 5 cases there were two consecutive overlapping potential storms (i.e. four days with large losses). From each of these five cases we selected the three day period which contained the largest aggregate loss and in this way obtained five storm events which were entered into the storm loss data set. In two cases there was five consecutive days which started potential storms (so that in all seven days had quite large damage). For these we considered the second three day period and the last three day period as separate storm events and entered these $2 \times 2 = 4$ storm events into the storm loss data set. For both cases these two storm events roughly corresponded to the largest losses of the five consecutive potential storms. These storms occurred in 1989 and 2000 and are storms number 12 and 13, and number 65 and 66, respectively, in the storm event data set. Finally, in one case during 2003 there where six consecutive days which started potential storms. For this case we included the third three day period and the last three day period into the storm loss data set. These two are storms number 77 and 78. For 1982 - 1986 we used the storm events identified by the same procedure in Rootzén and Tajvidi (1997). All storm events in this period were isolated

From Figure 2.4 it is seen that distribution of the losses in these storm events are quite heavy tailed. E.g., the Gudrun event led to 57% of the total loss in the entire period, 48 % of the total number of claims and was 4 times larger then the second largest loss. Similarly the second largest loss was 37% of the sum of the remaining losses.

For such heavy tailed distributions means and variances are of little interest. Nevertheless, as an aside, the average loss was MSEK 48.7 million and the standard deviation MSEK 295.2. The total loss in the 104 storm events was MSEK 5064. Only 9 of the 104 losses were larger than the average.

The univariate storm loss data set: One ingredient of the PoT method is to select a threshold, and then only use values which exceed this threshold in the further analysis. Using the preliminary storm loss data set we selected the threshold 2 MSEK for the univariate analysis (see discussion below). We redid the procedure described above to define wind storm events with the new threshold 2 MSEK. The resulting final univariate storm loss data set contains 80 storm events. There were 3 cases with two consecutive overlapping potential storms. For each of these we selected the 3-day period with the largest loss and entered it into the univariate storm loss data set. In one case, in 2003,

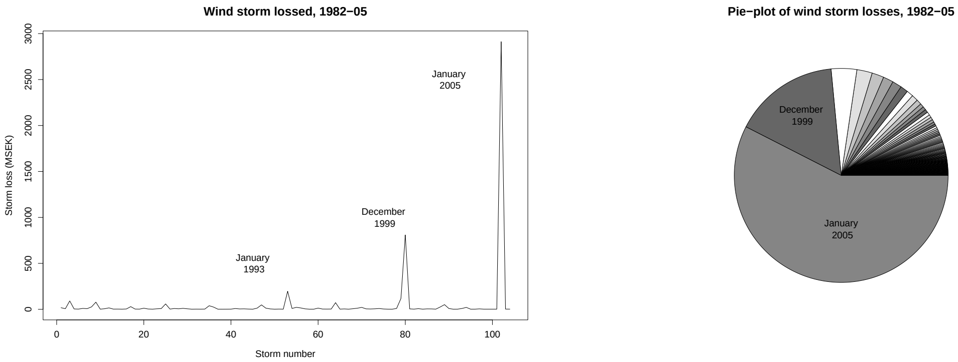

Figure 2.4: Storm events with losses exceeding MSEK 1.5 in 1982-2005. Left: Storm events ordered in time. Right: Pie-plot of storm events

there were five consecutive days which started potential storms. For this case we included the second three day period and the last three day period as separate wind storm events into the univariate storm loss data set. These two storm events roughly corresponded to the largest losses of the five consecutive potential storms.

The bivariate storm loss data set: As a background, in 14 of the storm events in the univariate storm loss data set there were no forest losses, and of the remaining ones only two had building losses less than 2 MSEK. Only three wind storms had larger forest losses than building losses. For Gudrun the forest loss was 2,115 MSEK out of a total loss of 2912. The second largest forest loss was 86 MSEK.

As a preliminary analysis to determine thresholds for the bivariate analysis we added all three day periods where the forest losses were larger than 0.06 SEK to the univariate storm loss data set – provided they were not already included. All the added storms were isolated. Based on this new preliminary bivariate data we, in a similar way as for the univariate analysis decided to use the thresholds $u_b = 2$ MSEK for building loss and $u_f = 0.14$ MSEK for forest loss. Using these thresholds, a second preliminary bivariate storm loss data set came to consist of all the storm events in the univariate storm loss data set plus 19 new storm events where forest loss exceeded 0.14 MSEK. However 6 of the 19 new events occurred in 2005, 2 before and 4 after Gudrun. This is a rather high number and we suspect that these in fact were due to damage caused by Gudrun but misreported. Since doing so is risk conservative we hence decided to removed these 6 storms, to obtain a final bivariate storm loss data set.

In this data set there then was 35 storm events where both thresholds were exceeded, 15 events where the forest threshold was exceeded but not the building threshold, and 44 events where the forest loss was below its threshold while the building loss exceeded 2 MSEK.

3 Univariate analysis

Extreme Value Statistics is a by now well established and well documented approach. We refer to Rootzén and Tajvidi (1997) for motivation and description in the windstorm insurance context, and for general accounts e.g. to the recent books Embrechts et al. (1997), Coles (2001), and Beirlant et al (2005). However, for ease of reference, and to introduce notation needed later, we briefly describe the main EVS model for exceedances over high levels, the Peaks over Thresholds (PoT) model.

In this model, the exceedances over a threshold u are assumed to be mutually independent and have a generalized Pareto (GP) distribution with distribution function (d.f.):

$$F(x) = 1 - (1 + \gamma \frac{x}{\sigma})_{+}^{-1/\gamma} \tag{3.1}$$

defined on $\{y : y > 0 \text{ and } (1 + \gamma y/\sigma) > 0\}$. Here $\sigma > 0$ is a scale parameter and γ is a shape parameter. In the present heavytailed situation only the case $\gamma > 0$ is of interest. The exceedances are assumed to occur as a stationary Poisson process (with intensity denoted by λ) which is independent of the sizes of the exceedances. As a minor extension, mentioned in Rootzén and Tajvidi (1997), the methods can be used without change if the Poisson intensity varies over unit blocks of time, in the sequel taken to be years, but is the same from year to year, as long as only integer multiples of years are considered.

Straightforward computation shows that that excess over a higher level $u + v$, $v > 0$ also has a Generalized Pareto df

$$1 - (1 + \gamma \frac{x}{(\sigma + v\gamma)})^{-1/\gamma}, \tag{3.2}$$

and that the median of the distribution of excesses of $u + v$ is

$$m(u + v) = \frac{\sigma}{\gamma}(2^{\gamma} - 1) + v(2^{\gamma} - 1). \tag{3.3}$$

It follows from (3.2) that if σ_u stands for the scale parameter for the excesses over a level u, then $\sigma_{u+v} - v\gamma = \sigma_u$.

Let M_T be the largest observation during a time period T years. An easy "Poisson" calculation gives that this maximum had the Extreme Value distribution

$$P[M_T \leq u + v] = \exp\left(-\lambda T (1 + \gamma \frac{v}{\sigma})_{+}^{-1/\gamma}\right). \tag{3.4}$$

The first step in a PoT analysis is to select a suitable threshold u which specifies if a storm loss is sufficiently large to be included in the analysis. The threshold selection is a trade-off between bias and variance. A high threshold gives less bias but also high variance due to few excesses. Traditionally the selection of the threshold is based in inspection of various diagnostic plots, such as parameter stability plots, mean/median excess plots, QQ-plots, and quantile plots. These plots also give a visual impression of the goodness of fit of the PoT model. Here we use this method and complement with simulated reference plots – i.e. the same plots based on samples simulated from the fitted distribution.

Once the threshold has been selected we use standard Maximum Likelihood (ML) methodology to estimate parameters, to compute confidence intervals, and to test submodels (for motivation, see Rootzén and Tajvidi (1997)).

Prediction of M_T, the largest storm loss in a future time interval [0,T] is at the center of interest for the insurance companies. A one-side prediction interval for M_T with coverage probability $1-p$ is given by

$$l(p,T) = [0, x_{T,p}],$$

where $x_{T,p}$ is the p-th upper quantile of the distribution of M_T. From Equation (3.4) follows that

$$x_{T,p} = u + \frac{\sigma}{\gamma}\left(\frac{(\lambda T)^{\gamma}}{(-\log(1-p))^{\gamma}} - 1\right). \tag{3.5}$$

By inserting ML estimates of the parameters into (3.5) one obtains a likelihood prediction interval for PML. In Hall et al. (2002) such prediction intervals are called "naive" since they do not take parameter uncertainty into account, and accordingly may have a coverage probability different from p. The Hall et al. paper instead proposes a bootstrap calibration procedure to adjust for parameter uncertainty, and shows that in many cases it improves coverage accuracy substantially. Here we used this approach. To be precise about which variant of the method we applied, we have included a step-by-step description of it in an Appendix 1. By way of further comment it is our experience that one should use a larger number of bootstrap resamples than in Hall et al. (2002), especially for more extreme events.

The total loss in a wind storm event is determined by the sizes of the claims and the number of claims in the event. We think of the distribution of the individual claim sizes as described by three parts, (i) the large claims; modelled as a GP distribution, (ii) the small claims; modelled nonparametrically by their means and variances, and (iii) the relative proportions of small and large claims.

4 Results of the univariate analysis

Threshold choice: Figure 4.1 shows stability plots for the parameters γ and $\sigma - \gamma u$ and for the ML estimates of 0.99-quantile of the excess distribution. Since the variation is quite large for $u > 7$ we for reasons of presentation have truncated the plots at $u = 7$.

If the data indeed have a GP distribution the first two of these plots should show a constant mean plus some random variability, while the last one should show a linear trend plus random variation. There is a rather high variability in the plots. However, the region $u = (2, 3)$ appears reasonably stable, and we decided to choose the threshold $u = 2$. This gives 80 storm events.

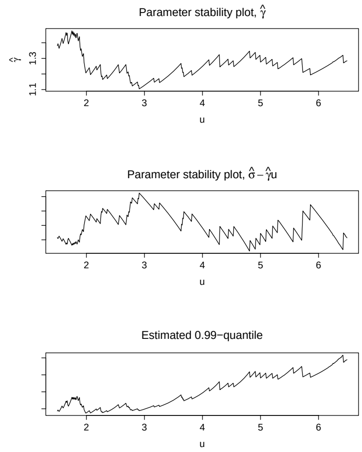

Figure 4.1: Parameter stability plot. Top: $\hat{\gamma}$. Middle: $\hat{\sigma} - u\hat{\gamma}$. Bottom: Estimated 0.99-quantile.

Period	$\hat{\gamma}$	$\hat{\lambda}$	u
1982-2005	1.21	3.33	2
1982-2004	0.99	3.35	2
R & T	0.7	3.8	1.4

Table 4.1: Estimated parameters for storm losses. The Rootzén and Tajvidi (R&T) estimates used data from 1982 to 1993.

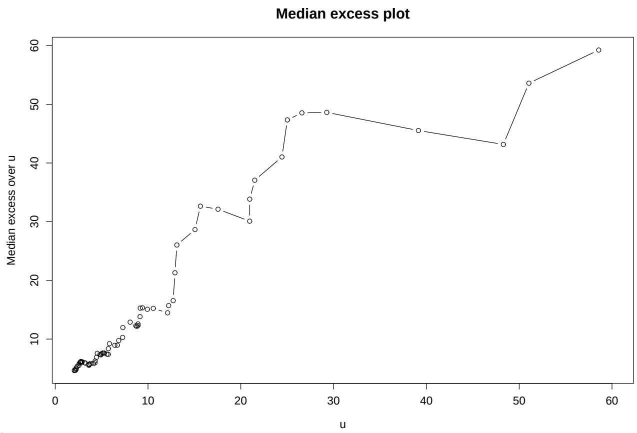

Figure 4.2: Median excess plot of the wind storm events 1982-2005.

Parameter estimates: The parameters of the PoT model were estimated for three different parts of the data, see Table 4.1. The data from the Rootzén and Tajvidi (1997) study were inflation adjusted to 2005 prices, but not adjusted for changes in exposure. The main feature of the table is that the estimated γ increases: it is 0.70 for the R&T data, 0.99 for the time period 1982-2004 and 1.21 for the full 1982-2005 data set, which also includes the storm Gudrun.

Goodness of fit: The median excess plot, Figure 4.2, indicates linearity and is consistent with a GP distribution. In the plot we only included u-values where the median estimate was based on seven or more windstorm losses.

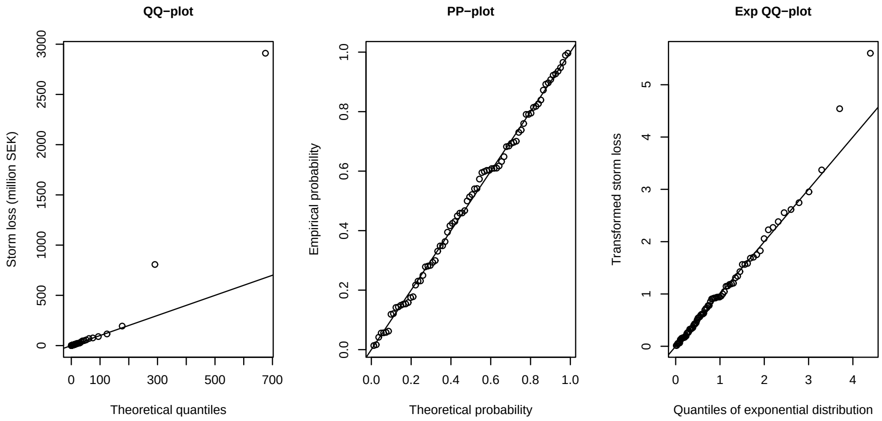

Figure 4.3: Model validation plots for the wind storm losses larger than 2 MSEK, 1982-2005, against the fitted GP distribution. Left: QQ-plot. Middle: PP-plot. Right: QQ-plot of the data transformed to exponential distribution using the estimated parameters, against an exponential distribution.

Figure 4.3 shows that the QQ-plot, PP-plot, and ”exponential QQ-plot” (i.e. the QQ-plot with data and estimated distribution transformed to an exponential scale) are as could be expected for a heavytailed GP distribution. Visually the two largest values deviate from the fitted line in the QQ-plot. However, these values are well inside the the pointwise confidence intervals limits (left out for reasons of presentations), and are in fact just as could be expected from a heavytailed distribution such as the fitted one.

To illustrate the fit further we have constructed ”simulated reference plots” (Figures 4.4 and 4.5), i.e. plots constructed from simulated samples of the same size from the fitted model. There is no indication that the time series plots and the parameter stability plots for the the simulated and real data comes from substantially different distributions.

Prediction intervals: We have used the method specified in the appendix, with 10000 resamples, to construct bootstrap calibrated prediction intervals for the maximum future loss. The prediction intervals based on the data for the entire period, 1982-2005, are substantially wider than the intervals based on the ”pre-Gudrun” period 1982-2004, see Figure 4.6. We have scaled the results to a ”dummy” currency to keep the confidentiality agreeed with Länsförsäkringar . From the figure it can also be seen that naive prediction intervals are much narrower than the bootstraped ones.

Viewed from the 2004 horizon, Gudrun with a loss of approximately 2,900 MSEK was a rather extreme event, as can be seen from the the prediction intervals we computed. From the perspective of Rootzén and Tajvidi (1997) Gudrun was even more extreme, both because the R&T intervals where ”naive” rather than bootstrapped and because they didn't included the 1997 event, which was about 5 times larger than the maximum event during 1982-1993. Nevertheless, even according to the R&T estimates there was

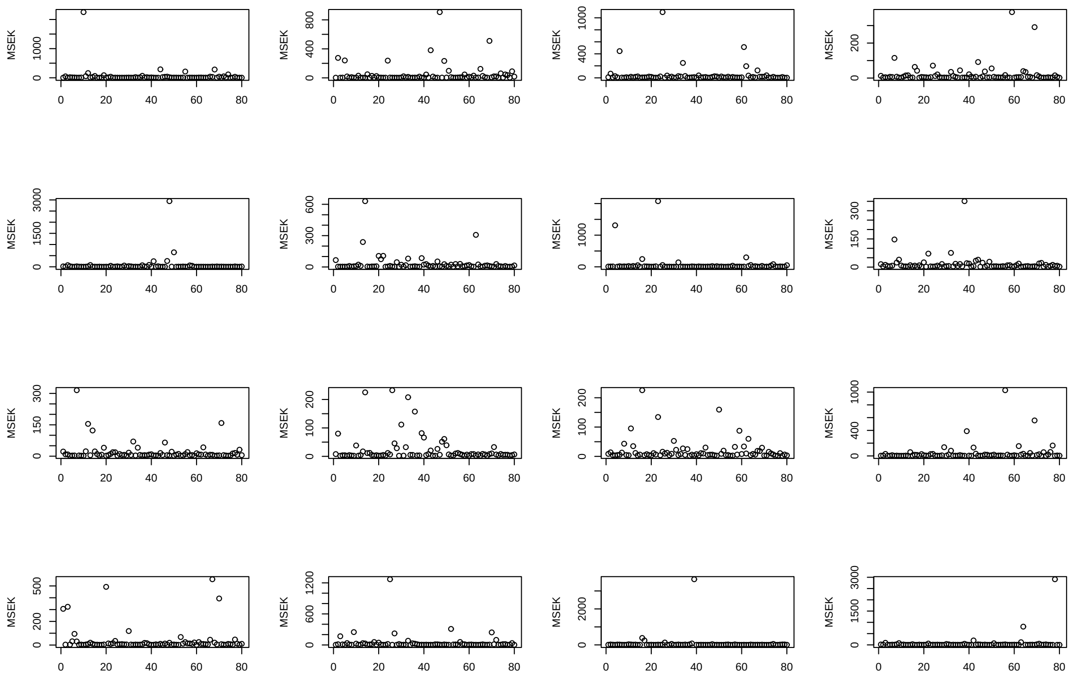

Figure 4.4: Simulated time series plots. Data generated from a GP distribution with parameters $u = 2$, $\gamma = 1.21$, and $n = 80$. Note the different vertical scales. The last plot shows the real data.

Data	Next year	Next 5 years	Next 15 years
1982-2004	9.18%	6.87%	5.34%
1982-2005	9.18%	7.37%	5.82%

Table 4.2: The naive prediction probability associated with the interval for the one-sided bootstrap calibrated 10% prediction intervals.

more than a 1 % risk that a storm like Gudrun would occur in a 15-year period, so Gudrun was not completely unthinkable, in particular if one took statistical uncertainty into account.

In Table 4.2 we illustrate the size of the correction made by replacing naive prediciton intervals with bootstraped calibrated ones.

Trend analysis: As a first crude analysis, there are 5 records out of 80 losses in the windstorm data. This agrees well with distribution of the number of records for 80 i.i.d. observations, which has expected value 4.97 and standard deviation 1.83, see Embrechts et al. (1997).

Fitting a GP model with constant scale parameter and a linear trend $\gamma = \alpha + \beta t$ to the 80 wind storm losses gave the estimates $\alpha = 0.74$ and $\beta = 0.038$. A likelihood ratio test of the hypothesis $\beta = 0$ had p-value 0.83. The model $\sigma = \exp(\alpha + \beta t)$ and γ constant

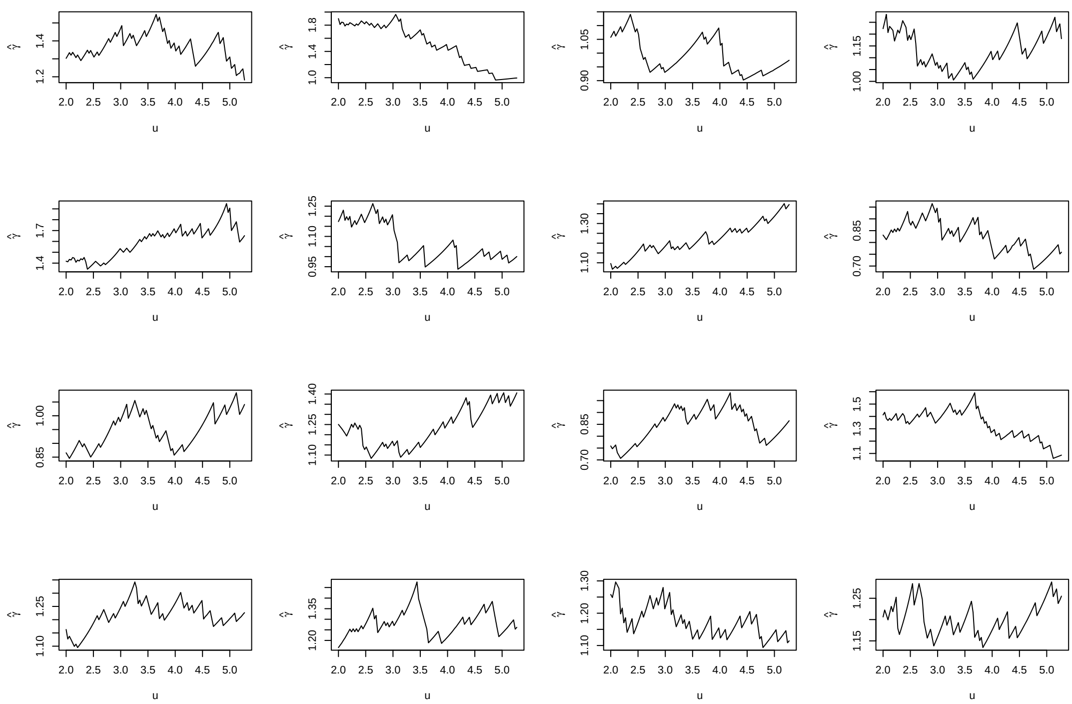

Figure 4.5: Simulated parameter stability plots for γ. Data generated from GP distribution with parameters $u = 2$, $\gamma = 1.21$, and $n = 80$. Note the different vertical scales. The last plot is based on the real data.

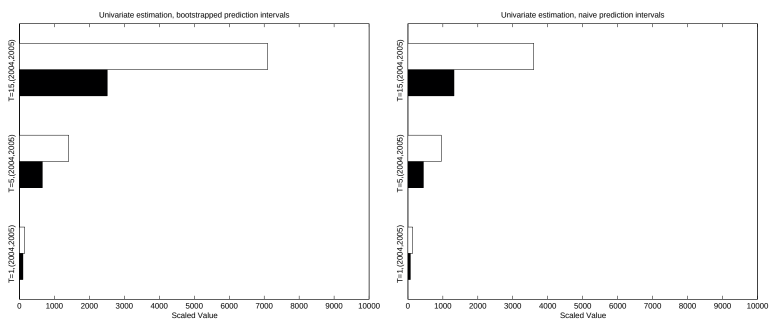

Figure 4.6: Estimated one-sided prediction intervals based on univariate analysis. Risk level 10%. Dark is 1982-2004 and white is 1982-2005 data respectively. Left: Bootstrap calibrated prediction intervals. Right: Naive prediction intervals.

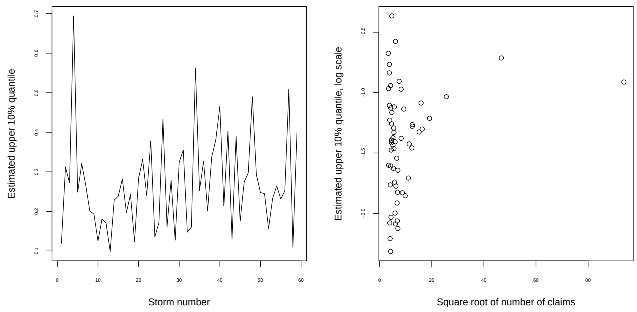

Figure 4.7: Estimated 10% quantile (MSEK) of individual excesses of 0.06 MSEK for storms with at least 10 individual claims larger than 0.06 MSEK. Left: Against storm number. Right: On log scale, against square root of the number of excesses.

lead to the estimates $\hat{\gamma} = 1.21$, and $\hat{\beta} = -0.016$. A likelihood ratio test of $\beta = 0$ gave the p-value 0.37.

As the first analysis of the individual claims in the storm events we for each storm event fitted a separate GP distribution to the excesses of .06 MSEK. Figure Figure 4.7, left, shows estimated 0.1 upper quantiles computed from the fitted GP distributions. Visually there is a weak positive trend, but formal test did not indicate significance. The right plot indicates that the the individual claims in the largest storm events tended to have higher quantiles.

A linear regression analysis of the averages of the claims which were smaller than MSEK 0.06 (see Figure 4.8), weighted by the number of claims in the storm event lead to an estimated trend of SEK 753 per decade, with associated p-value less than 0.001. The trend was mainly caused by the 15 first storms. The right plot in Figure 4.8 gives no indication that the means of the small claims dependend on the number of claims.

Weighted linear regression of the proportions of the individual claims which exceeded MSEK 0.06 resulted in a estimated increase of 5% per decade, with p-value smaller than .001.

A constant distribution for the excesses of MSEK 0.06 plus an increasing trend in the means of the small claims and in the proportions exceeding should lead to increasing trends in the quantiles for the entire distribution of individual claims. This can in fact be seen in Figure 4.9. Weighted linear regression analysis of the empirical 0.7- and 0.9-quantile for storms with more than 100 claims showed positive slopes, both with p-value less than .001. We didn't see any dependence between these quantiles and the number of claims.

To check for trends in the yearly numbers of storm events (with loss exceeding 2 MSEK), we used a generalized linear model with Poisson response distribution and inten-

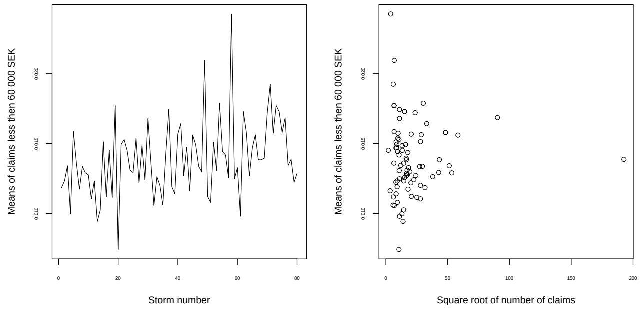

Figure 4.8: Averages (in MSEK) of individual claims smaller than MSEK 0.06. Left: Against storm numbers. Right: Against square root of number of excesses.

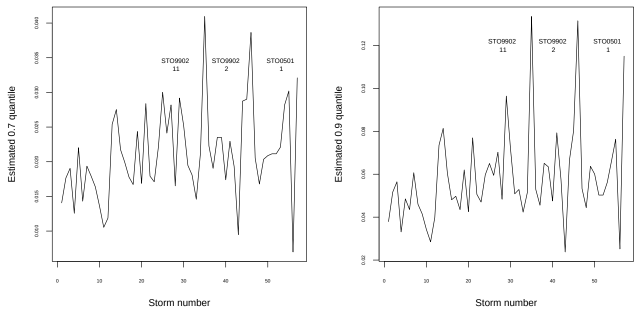

Figure 4.9: Empirical quantiles of wind storm losses, for events with more than 100 claims. Left: 0.7-quantile. Right: 0.9-quantile.

sity $\log(\lambda) = \alpha + \beta \times$ years. The p-value for the test of $\beta = 0$ was 0.29. The autocorrelation for the yearly numbers of storm events was -0.35. If the data had been normal, this would not be significantly different from 0.

5 Bivariate analysis

The previous sections have used an univariate approach. In the sequel we model the losses caused by damage to the forest industry and damage to buildings separately. There are two reasons for this. The first is that bivariate modeling gives a possibility to to study portfolio changes which affect building losses and forest losses differently. For instance, the major part of losses in the Gudrun event was due to damage to the forest industry. As forest was not heavily insured one could expect an increase in forest insurance after this event – and in fact one year after Gudrun, Länsförsäkringar had already experienced a 10% increase in forest insurance. Thus, using the same simple portfolio correction as in Section 2, and with the loss in a wind storm event caused by damage to buildings B and the loss caused by damage to forests denoted F, then for risk predictions $B + aF$, for some $a > 1$, may be more interesting than $B + F$. However, it is only the latter sum, and multiples thereof, which may be studied through the univariate approach. The former quantity requires bivariate modeling.

The second reason, discussed in the introduction, is that univariate modeling "interpolates" between the less dramatic risks posed by damage to buildings, and the potentially extreme losses due to damage to forest. Thus univariate modeling may both overestimate the risks of moderate or large losses and miss the risks posed by the rare events which cause extremely large damage to the forest industry.

Bivariate PoT methods are still in a rather early stage of development. Bruun and Tawn (1998) compared a univariate and a bivariate EVS approach to coastal flooding. A conclusion was that the two methods sometimes produced different risk levels, with the bivariate ones more conservative. The authors also argued that the bivariate method provided better extrapolation and design information. A main issue for PoT estimation is how to handle observations where one component is extreme and the other not, cf. Rootzén and Tajvidi (2006). This paper derived the general form of multidimensional generalized Pareto distribution, but did not develop the statistical aspects. In a series of papers, Tawn, Ledford and coworkers, develop and uses heuristically motivated statistical methods which cover a much larger class of situations, where the components also may be asymptotically independent, but which don't explicitly use multivariate GPD's. Here we apply, for the first time, the method of Rootzén and Tajvidi (2006).

We used the particular bivariate symmetric logistic GPD (BGPD) given in Example 1 of Rootzén and Tajvidi (2006) (see Appendix 2) to model the exceedances $(X, Y) = (B - u_B, F - u_f)$ of the thresholds $u_B = 2$ and $u_F = 0.14$. Then, in the bivariate PoT model storm events occur in time as a Poisson process, and the losses in such storm events are supposed to be i.i.d., independent of the Poisson process, and to follow the symmetric logistic BGPD. Storm events are specified by the the requriement that either building loss, or forest loss, or both, exceed their threshold. Our analysis of the bivariate storm loss data set using this model consisted of two steps: parameter estimation and quantile computation.

In the estimation of the BGPD parameters, if one of the losses in a storm event was below its threshold, we replaced it by zero. Thus, for estimation we replaced negative components in the observations (x, y) by 0, i.e. we based the estimation on (x_+, y_+). The BGPD likelihood function then consisted of three parts. The first one, for the case when both components were positive, was obtained by (symbolic) differentiation of the first expression for the distribution function in Appendix 8. The second component, corresponding to $x_+ = 0, y_+ > 0$, was obtained by setting $x = 0$ in this expression and differentiating with respect to y. The third component of the likelihood function, for $x_+ > 0, y_+ = 0$ was obtained correspondingly. ML parameter estimates were then obtained by maximizing this likelihood function numerically.

For the computation step, we in the general case are interested in the total loss when the forest exposure in the portfolio has increased by a factor a, i.e. when the total loss caused by a storm event is $B + aF$. Let M_T be the largest such loss during a time period of length T. Suppose $v = u - u_B - au_F > 0$. By the same Poisson argument as for (3.4) we then have that

$$P[M_T \leq u] = \exp\left(-\lambda T P[X + aY > v]\right), \tag{5.1}$$

where λ is the intensity of the Poisson process of exceedances and (X, Y) has the BGPD.

Estimates of quantiles of M_T where obtained by inversion of estimates of $P[M_T \leq u]$. Such estimates were in turn obtained by replacing the parameters in the right hand side of (5.1) by their estimates. Here λ was estimated by the number of storm events divided by the length of the observation period. Further $P[X + aY > v]$ was estimated by the same probability in the BGPD where the true parameters were replaced by their ML estimates. Since there are no analytical expression for $P[X + aY > v]$ this probability had to be computed numerically. We did this in Mathematica: first we used symbolic differentiation of the distribution function in Appendix 2, different in the three different regions of definition for the BGPD, to get the density of the BGPD. Then, to compute the desired probability, one integrates this density over the region

$$x + ay > v, \quad x > max\{-u_B, -\sigma_x/\gamma_x + \mu_x\}, \quad y > max\{-u_F, -\sigma_x/\gamma_x + \mu_x\}.$$

Here the restriction $x > -u_B$ is because losses cannot be negative, and the restriction $x > -\sigma_x/\gamma_x + \mu_x$ is since the support of the BGPD is the rectangle $(-\sigma_x/\gamma_x + \mu_x, \infty) \times (-\sigma_y/\gamma_x + \mu_y, \infty)$. The restrictions on y are for the same reasons.

We found that the numerical calculations of the probabilities where non-trivial and should be carefully treated. The reason is the small magnitude of the tail of the distribution.

As seen above, in the estimation step we only trusted the BGPD to give a good fit in the rectangle $(u_B, \infty) \times (u_F, \infty)$. This is because we felt that situations where one component of loss was just over its threshold and the other component was below were not asymptotic enough for the model to give good fit for the second component. On the other hand, in the computation step interest was centered on quite large losses, and we thought it reasonable to use the BGPD model for the entire area of interest.

We used the univariate model checking tools discussed in Section 3 to see how well the marginal GP distributions fit. To check the fit of the dependence structure of the model, we compared the model estimate of Pickand's dependence function with a nonparametric estimate of the same quantity, see Section 8.2.5 in Beirlant et al. (2004).

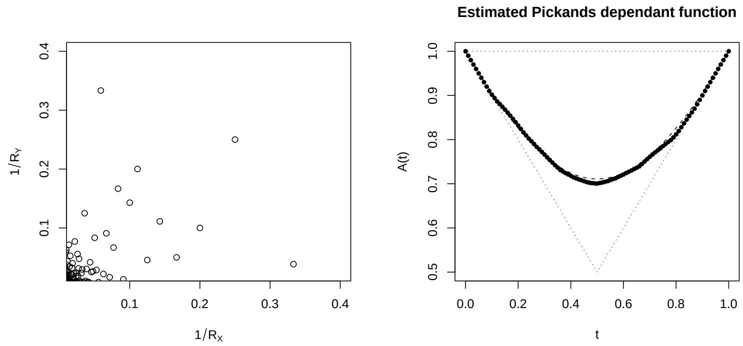

Figure 6.1: Left: Reciprocal of ranks (counted form top) of forest losses, $(1/R_F)$, and building losses, $(1/R_B)$. Figure truncated for reasons of presentation. Right plot: Estimated Logistic dependence function (dotted) and the non-parametric estimated Pickands' dependence function for forest and building losses.

	1982 - 2005					1982 - 2004				
Parameters	λ	r	γ_x	γ_y	LLH	λ	r	γ_x	γ_y	LLH
Estimates	3.92	2.09	1.86	1.09	-425.98	3.96	2.03	1.63	1.05	-399.99
Initial values		2.03	1.85	1.09	-428.73		1.93	1.57	0.91	-402.29

Table 6.1: Estimated parameters for the bivariate GP model, for $u_B = 2.00,\ u_F = 0.14$. Small forest losses in 2005 not included.

6 Results of bivariate analysis

Parameter estimates: In the numerical optimization of the likelihood function we used the marginal GPD parameters estimated separately for forest loss and for building loss as initial values. The initial r, μ_x and μ_y values were obtained by holding these marginal parameters fixed and maximizing over r, μ_x and μ_y with initial values $(2, 0, 0)$. The final estimates were reasonably close to the estimated ones (Table 6.1).

Measuring riskiness by the size the shape parameter γ, building losses were less risky than the univariate estimate of riskiness, and forest losses where more risky (Tables 4.1 and 6.1). Interestingly, this is also observed for the 1982-2004 data on the wind storms before Gudrun. Hence, the danger posed by the possibility of very large forest losses could have been detected also in 2004.

Goodness of fit: Figure 6.1, left plot, indicates that on the Frechet scale dependence is symmetric, and the right plot shows a good agreement between the non-parametric and parametric estimate of the Pickands' dependence function. Figure 6.2 shows the fit of the marginal GP distributions. We also fitted a GP distribution to the 1982-2004 forest loss data. Gudrun was outside the 95% pointwise confidence intervals for this fit. Otherwise

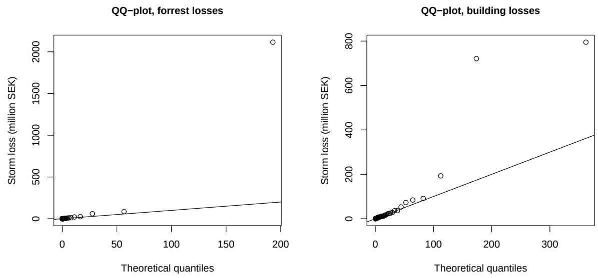

Figure 6.2: Left plot: QQ-plot against GP distribution, forest losses. Right plot: QQ-plot against GP distribution, building losses.

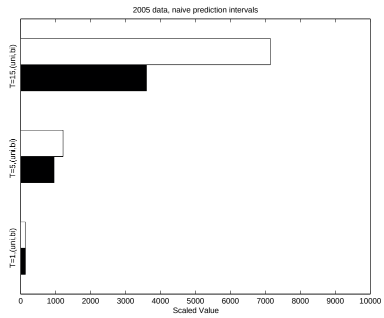

Figure 6.3: Estimated naive one-sided prediction intervals based on bivariate model and univariate model. Risk level 10%. Dark is univariate model and white is bivariate model.

the fit seemed good.

Prediction intervals: We used numerical integration in Mathematica with the method *GaussKronrod* to calculate the probabilities in Equation (5.1), followed by a minimization to find the quantile. This was rather time consuming and the minimization required good initial guesses.

The bivariate prediction intervals were wider than the naive univariate ones for extreme quantiles but smaller for moderate quantiles, see Figure 6.3. Presumably, this is because the univariate analysis tries to average between the rare but potentially very large forrest losses and the more common losses due to damage to buildings, and then overestimates the risk of moderate, but underestimates the risk of extremely large losses. Further, the bivariate prediction intervals are not bootstrap calibrated. This makes it possible, from the experience of Section 4, that calibrated prediction intervals would be wider. If one compares the the bivariate result with the bootstrap calibrated univariate ones they are similar in magnitude, see Figure 6.4. We do not want to speculate of the magnitude of a bivariate result corrected for statistically uncertainty.

An further point is that if the 6 small forest losses from 2005 are included in the analysis the resulting estimates indicate a smaller dependence between building and forest losses and less risky forest losses. This indicates the importance of selection procedure used to

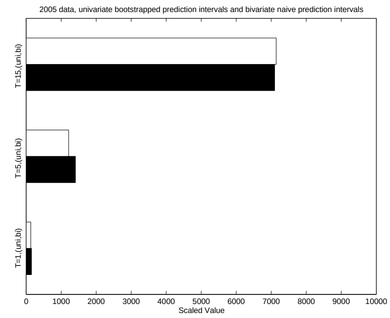

Figure 6.4: Estimated one-sided prediction intervals based on bivariate model (naive) and univariate model (bootstrapped). Risk level 10%. Dark is univariate model and white is bivariate model.

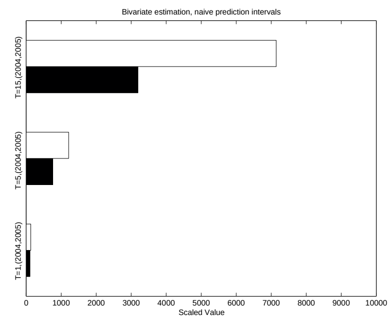

Figure 6.5: Estimated naive one-sided prediction intervals based on bivariate model. Risk level 10%. Dark is based on 1982-2004 data and white on 1982-2005 data.

produce storm loss data sets.

The bivariate model estimates the probability of a storm of Gudruns magnitude or greater within next year to be 1.6% and 1.1% for the 1982-2005 and 1982-2004 data respectively. See Figure Figure 6.5 for a comparison of risk between the 1982-2004 and 1982-2005 data sets. As discussed in the introduction, there has been three storms with similarly extreme magnitude during the 20th century. This makes these estimated probabilities quite reasonable. Note that even if the two data sets produced prediction intervals of different magnitudes, the probabilities are much more similar. This will be further discussed in Section 7.

Increase in forest exposure: We considered two cases: that after the Gudrun storm the forest insurance portfolio increases by 20% or 50% while the building portfolio stays constant. We used the same somewhat simplistic portfolio correction again, i.e. we assumed that the total loss after the portfolio changes are $B + 1.2F$ and $B + 1.5F$, respectively, with B and F the forest and building losses without portfolio change.

Figure 6.6 shows the that the impact of the changes is moderate for the 10% risk level – the most extreme quantile increased by about 30% and the others by less. That the impact was largest for the most extreme quantile is as could be expected, since the forest

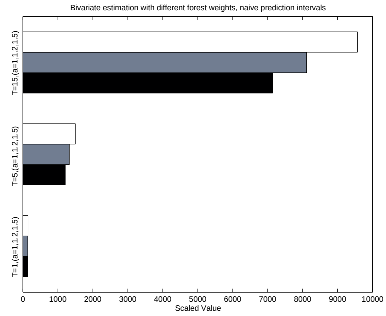

Figure 6.6: Estimated naive one-sided prediction interval (MSEK) for the 10% risk level for different forest insurance exposures, based on the 1982-2005 data. Dark is for no change in exposure, grey is assuming forest exposure has increased by 20% and white is assuming forest exposure has increased by 50%. intervals.

losses are more heavy tailed than the building losses.

7 An alternative presentation of risk

Above we have quantified risk by computing prediction intervals for quantiles (or PML's) for different risk levels and time periods. For small risk levels and long time periods the PML-s are so large that reinsurance up to these levels might be prohibitively costly and make it impossible for companies to provide insurance.

The number of windstorm events which the intervals are based is moderate, and one explanation of the sizes of the endpoint of the prediction intervals could be that they are caused by statistical uncertainty. However, we instead believe the reason is that losses have heavy tails and that very large losses in fact are possible.

One consequence of the heavy tails is that a small alteration in risk level causes a large change in the prediction interval. A more calm alternative description is to estimate the probabilities of loss larger that suitably selected levels. These levels e.g. could correspond to possible reinsurance amounts.

A drawback with this approach is that it does not in an obvious way take statistical uncertainty into consideration. One possibility is to add confidence intervals to the estimated probabilities, but again, how should one then combine the risk measured by the PML estimate and the risk of non-coverage of the confidence interval?

For this investigation we consider a very extreme storm, a loss which is some multiple of the loss caused by Gudrun. For simplicity, denote the amount as X in our dummy currency. We are interested to compare estimated probabilities for a loss of this magnitude on different time horizons. Also, we compare with the probabilities for a loss of magnitude $2X$

Table 7.1 shows that point estimates of such probabilities in fact may be easier to

	a=1.0			a=1.2			a=1.5		
Loss event	In 1Y	In 5Y	In 15Y	In 1Y	In 5Y	In 15Y	In 1Y	In 5Y	In 15Y
X	0.8 %	3.7 %	10.7 %	0.8%	3.9%	11.4 %	0.9 %	4.3 %	12.5%
$2X$	0.5 %	2.5 %	7.4 %	0.6%	2.7%	7.9%	0.6 %	2.9%	8.7%

Table 7.1: Probability for a loss which is larger than a specified loss size X or 2X.

digest. We also give the same probabilities for the forest loss portfolio increased by 20% and 50% respectively.

8 Conclusions

In this paper we selected and developed univariate and bivariate threshold methods for analysis of wind storm loss data and applied them to a data set consisting of the large wind storm losses for the Swedish insurance group Länsförsäkringar during the period 1982-2005. The models were used to construct prediction intervals for future maximum losses and to assess the effects of a possible increase in the number of forest insurance contracts. We further made detailed investigations of the quality of fit of the models and of possible time trends in the data. Our main conclusions were

- There was a weak positive trend in the sizes of the individual claims in the storm events. This could be due to people building more exclusively and/or in more exposed areas. However, the sizes of high level excesses of the individual claims did not seem to increase.
- No other trends in were found in the data. In particular we found little evidence that storms in Sweden have become more frequent or more severe over time.
- The univariate peaks over threshold model fitted the data well. The storm Gudrun influenced predictive risk estimates markedly.
- A bivariate model also fitted well. It pointed to somewhat higher risks of extremely large losses, and smaller risks of moderate and large losses. We believe the bivariate method may be the most realistic ones. A further advantage with the bivariate approach is that it makes it possible to study the effects of changes in the composition of the insurance portfolio.
- Predicted losses were rather insensitive to changes in portfolio size.
- It would have been possible to detect the risks of losses of the size caused by Gudrun from the 1982-2004 data, and even earlier univariate analyses didn't rule out the possibility of events like Gudrun.

- Companies should develop systematic ways of thinking about "not yet seen" disasters.

To elaborate on the last point: as argued above, it would have been possible to predict Gudrun from earlier data. However this would have required companies to have procedures which detected and put the focus on potential new "loss modes". Of course this is a general problem – companies need to have good and structured ways of thinking of not only what has happened but on what could happen.

As a final comment, the Swedish regulatory agency asks for "1 in 200 year" solvency estimates, and other countries have similar requirements. We believe that such very extreme estimates are fraught with too large uncertainties, and too much a result of the assumptions companies put into the analysis. A better approach could be to use, say "1 in 50" years estimates as a basis for solvency requirements, and then ask companies to complement this with well thought out plans for what to do if even larger losses occur. One possibility for such plans could be to change contracts so that policyholders only get a part of their claims covered if the total loss to a company exceeds some very large amount – nobody will profit from insurance companies going into bankruptcy.

Acknowledgement We want to thank Slobodanka Mijalkovic, Nader Tajvidi, and Håkan Pramsten for help with the Länsförsäkringar data bases and for helpful comments.

Appendix 1

Consider wind storm losses $X_1, \ldots, X_n$ incurred at times $T_1, \ldots, T_n$. As discussed in Section 3, the X_i's are assumed to be iid and have Generalized Pareto distributions and the $T_i - T_{i-1}$ are assumed to be mutually independent and independent of the X-es, and exponentially distributed. A "naive likelihood prediction interval" is obtained by letting the estimated quantile be the right endpoint of the interval. The bootstrap calibrated prediction interval is constructed as follows (see Hall et al. (1999, 2002)):

Select p, T and N. Here $1-p$ is the desired coverage probability for the prediction interval, T is the time horizon and N is the number of bootstrap resamples.

1. Estimate $\hat{\gamma}$, $\hat{\sigma}$ and $\hat{\lambda}$ from the storm data.
2. Simulate from the GP and exponential distribution, using $\hat{\gamma}$, $\hat{\sigma}$ and $\hat{\lambda}$, a new storm sample with the same time length as the original time series and estimate new $\hat{\gamma}^*$, $\hat{\sigma}^*$ and $\hat{\lambda}^*$.
3. Calculate a naive prediction interval, $l^*(p,T)$, from $\hat{\gamma}^*$, $\hat{\sigma}^*$ and $\hat{\lambda}^*$.
4. Denote M_T^* a future maximum value given $\hat{\gamma}$, $\hat{\sigma}$ and $\hat{\lambda}$. Compute $\pi(1-x)_i = P[M_T^* \in l^*(x,T)]$ for x in the vicinity of p. This can be done analytically.

5. Repeat step 1-4 N times and compute the mean of $\pi(1-x)_i$, $i = 1, \ldots, N$ denoted by $\hat{\pi}(1-x)$.

6. Define $\hat{\alpha}$ to be the solution of $\hat{\pi}(1-\alpha) = 1-p$. Then $l(\hat{\alpha}, T)$ is the bootstrap calibrated prediction interval.

Appendix 2

The distribution function of the symmetric logistic GPD is zero in the third quadrant. The expressions for the other quadrants are as follows.

$$1 - \frac{((1+\frac{\gamma_x(x-\mu_x)}{\sigma_x})_+^{-r/\gamma_x} + (1+\frac{\gamma_y(y-\mu_y)}{\sigma_y})_+^{-r/\gamma_y})^{1/r}}{((1-\frac{\gamma_x\mu_x}{\sigma_x})^{-r/\gamma_x} + (1-\frac{\gamma_y\mu_y}{\sigma_y})^{-r/\gamma_y})^{1/r}}, \quad x > 0,\ y > 0,$$

$$\frac{((1+\frac{\gamma_x(x-\mu_x)}{\sigma_x})_+^{-r/\gamma_x} + (1-\frac{\gamma_y\mu_y)}{\sigma_y})^{-r/\gamma_y})^{1/r} - ((1+\frac{\gamma_x(x-\mu_x)_+}{\sigma_x})^{-r/\gamma_x} + (1+\frac{\gamma_y(y-\mu_y)_+}{\sigma_y})^{-r/\gamma_y})^{1/r}}{((1-\frac{\gamma_x\mu_x}{\sigma_x})^{-r/\gamma_x} + (1-\frac{\gamma_y\mu_y}{\sigma_y})^{-r/\gamma_y})^{1/r}}, \quad x > 0,\ y \leq 0,$$

$$\frac{((1-\frac{\gamma_x\mu_x}{\sigma_x})^{-r/\gamma_x} + (1+\frac{\gamma_y(y-\mu_y)}{\sigma_y})_+^{-r/\gamma_y})^{1/r} - ((1+\frac{\gamma_x(x-\mu_x)_+}{\sigma_x})^{-r/\gamma_x} + (1+\frac{\gamma_y(y-\mu_y)_+}{\sigma_y})^{-r/\gamma_y})^{1/r}}{((1-\frac{\gamma_x\mu_x}{\sigma_x})^{-r/\gamma_x} + (1-\frac{\gamma_y\mu_y}{\sigma_y})^{-r/\gamma_y})^{1/r}}, \quad x \leq 0,\ y > 0.$$

The parameters must satisfy, $\mu_x < \sigma_x/\gamma_x$ and $\mu_y < \sigma_y/\gamma_y$ when $\gamma_x, \gamma_y > 0$ as is the case in this paper. The parameter r is governs dependency: $r = 1$ gives independence while $r = \infty$ corresponds to complete dependence.

References

[1] Bruun, J.T. and Tawn, J.A (1998) Comparison of approaches for estimating the probability of coastal flooding. *J. Roy. Statist. Soc. Ser. C* **47**, Part 3, 405-423.

[2] Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values.* Springer, London.

[3] Cossette, H., Duchesne, T., and Marceau, E. (2003) Modeling catastrophes and their impact on insurance portfolios. *North American Actuarial Journal* **7**,1-22.

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

[5] Hall, P, Peng, L., and Tajvidi, N. (1999) On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika* **86**, 871-880.

[6] Hall, P, Peng, L., and Tajvidi, N. (2002) Effect of extrapolation on coverage accuracy of prediction intervals computed from Pareto-type data. *Ann. Statist.* **30**, 875-895.

[7] Holmberg (2005) Sammanställning av stormskador på skog i Sverige under de senaste 210 åren. *Report no. 9, Skogsvårdsstyrelsen*, Skogsstyrelsens förlag 551 83 Jönköping.

[8] Jaffee, D.M. and Russell, T. (1996) Catastrophe insurance, capital markets and uninsurable risk. In: Cummins, J.D. (Ed.) *Proceedings of the Conference on Risk Management in Insurance Firms.* Financial Institutions Center, The Wharton School, Philadelphia.

[9] Lescourret, L. and Robert, C. (2006) Extreme dependence of multivariate catastrophic losses. *Scand. Actuarial J.* **4**, 203-225.

[10] Ledford, A.W. and Tawn, J.A. (1997) Modelling dependance within joint tail regions. *J. R. Statist. Soc. Ser. B* **59**, 475-499.

[11] Lies, M. (2000) Storm over Europe An underestimated risk. *Swiss Re.*

[12] Rootzén, H. and Tajvidi, N. (1997) Extreme value statistics and wind storm losses: a case study. *Scand. Actuarial J.* **1**, 70-94.

[13] Rootzén, H. and Tajvidi, N. (2001) Can wind storm losses be predicted from meteorological observations? *Scand. Actuarial J.* **5**, 162-175.

[14] Rootzén, H. and Taijvidi, N. (2006) Multivarate generalized Pareto distribution. *Bernoulli* **12**, 917-930.

[15] Smith, R.L. (1987) Estimating tails of the probability distributions. *Ann. Statist.* **15**, 1174-1207.

[16] Swedish Metorological and Hydrological Institute (SMHI) (2005) *Årsredovisning.*

[17] Valinger, E., Ottosson Lövenius, M., Johansson, U., Fridman, J., Claeson, S., Gustafsson, Å. (2006) Analys av riskfaktorer efter stormen Gudrun. *Report no. 8, Skogsvårdsstyrelsen*, Skogsstyrelsens förlag 551 83 Jönköping.