Financial Risk
4-th quarter 2020/21
Lecture 3: ML inference, wind storm insurance

"As an alternative to the traditional 30-year mortgage, we also offer an interest-only mortgage, balloon mortgage, reverse mortgage, upside down mortgage, inside out mortgage, loop-de-loop mortgage, and the spinning double axel mortgage with a triple lutz."

Gudrun January 2005
326 MEuro loss
72 % due to forest losses
4 times larger than second largest

Maximum Likelihood (ML) inference (Coles p. 30-43)

Likelihood function = the function which shows how the "probability" (or likelihood) of getting the observed data depends on the parameters

$x_1, \dots x_n$ observations of i.i.d. variables $X_1, \dots, X_n$, density $f(x) = f(x; \theta)$

$\theta = (\theta_1, \dots, \theta_d)$ parameters

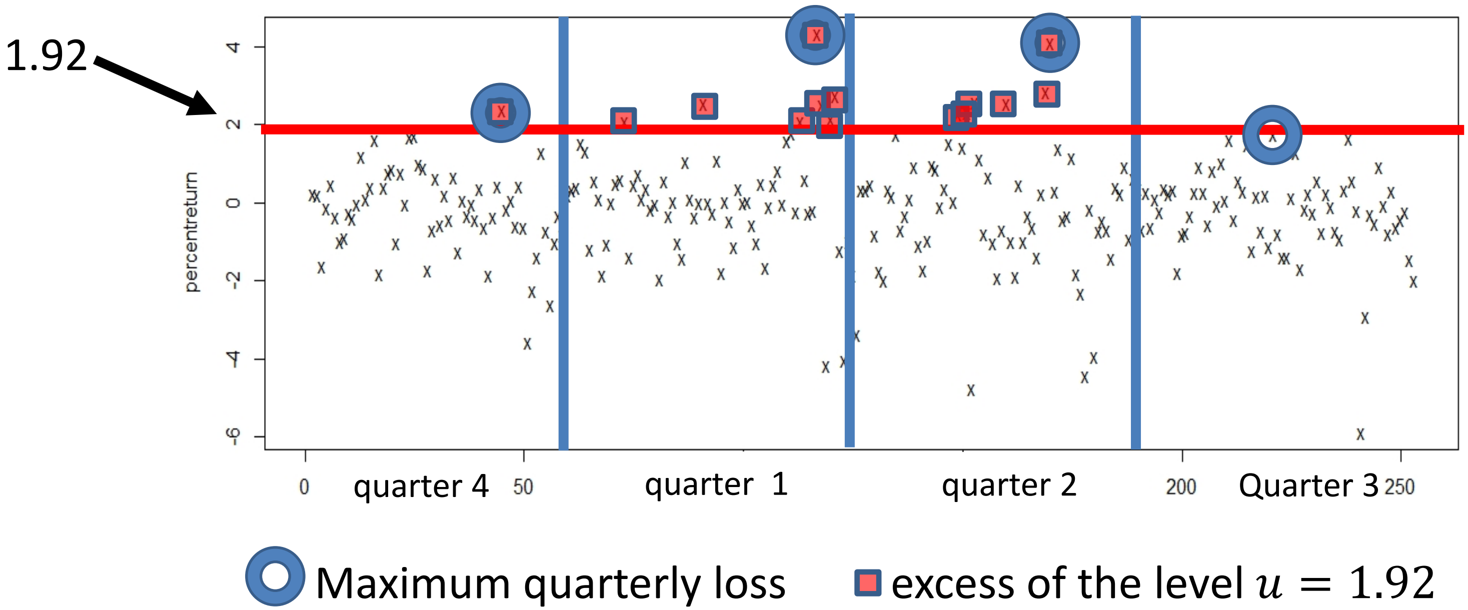
$L(\theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta)$ likelihood function

$\ell(\theta) = \log f(x_1; \theta) + \log f(x_2; \theta) + \dots \log f(x_n; \theta)$ log likelihood function

ML estimates = the value $\hat{\theta} = (\hat{\theta}_1 \dots \hat{\theta}_d)$ which maximizes the (log) likelihood function

- Ml estimates often have to be found through numerical maximization
- sometimes a maximum doesn't exist
- sometimes several local maxima (→ problem for numerical maximization)
- but typically no problems if the number of observations is "large"

Apple losses (= - $100 \times \frac{\text{price tomorrow} - \text{price today}}{\text{price today}}$) one year back



How large is the risk of a big quarterly loss? BM

How large is the risk of a big loss tomorrow? PoT

Example: ML estimation of the parameters in the PoT model

T = length of observation period
N = number of observed excesses (random variable!)
$x_1, \ldots x_N$ observed excess sizes
$\theta = (\sigma, \gamma, \lambda)$ parameters

The probability of observing *N* excesses is $\frac{(\lambda T)^N}{N!} \exp\{-\lambda T\}$
independence ⇒

$$L(\theta) = L(\lambda, \sigma, \gamma) = \frac{(\lambda T)^N}{N!} \exp\{-\lambda T\} \prod_{i=1}^{N} \frac{1}{\sigma} \left(1 + \frac{\gamma}{\sigma} x_i\right)_+^{-1/\gamma - 1}$$

$$\begin{aligned} \ell(\lambda, \sigma, \gamma) &= N \log(\lambda) + N \log(T) - \log(N!) - \lambda T \\ &- N \log(\sigma) - \sum_{i=1}^{N} (1/\gamma + 1) \log\left(1 + \frac{\gamma}{\sigma} x_i\right)_+ \end{aligned}$$

$$\frac{\partial}{\partial \lambda} \ell(\lambda, \sigma, \gamma) = \frac{N}{\lambda} - T = 0 \quad \text{so that} \quad \hat{\lambda} = \frac{N}{T}$$

$\hat{\sigma}, \hat{\gamma}$ obtained from numerical maximization of the second part of $\ell(\lambda, \sigma, \gamma)$

ML inference: asymptotic properties

$\mathcal{I}(\theta) = E_\theta\big(\big(- \frac{\partial^2}{\partial\theta_i \partial\theta_j}\ell(\theta)\big)\big)$ expected Fisher information matrix, estimated by $\mathcal{I}(\hat{\theta})$ or by $I(\theta)$ where $I(\theta) = (-\frac{\partial^2}{\partial\theta_i \partial\theta_j}\ell(\theta))$ is the the observed Fisher information matrix. (In the expected Fisher information matrix, the observations are replaced by the corresponding random variables when the expectations are computed. Numerical optimization programs typically compute the Hessian, $-I(\theta)$)

$\hat{\theta} = (\hat{\theta}_1, \ldots \hat{\theta}_d)$ asymptotically has a d-dimensional multivariate normal distribution with mean θ and variance $\mathcal{I}(\theta)^{-1}$

In particular, the variance of $\hat{\theta}_i$ may be estimated by $(\mathcal{I}(\hat{\theta})^{-1})_{ii}$ = the *i*-th diagonal element of $\mathcal{I}(\hat{\theta})^{-1}$), or by $(I(\hat{\theta})^{-1})_{ii}$ The latter is often more accurate.

k_α = the α-th quantile of the standard normal distribution ($k_{0.975} = 1.96$)

$(\hat{\theta}_i - k_{1-(1-\alpha)/2}\sqrt{(I(\hat{\theta})^{-1})_{i,i}},\ \hat{\theta}_i + k_{1-(1-\alpha)/2}\sqrt{(I(\hat{\theta})^{-1})_{i,i}}\ )$ asymptotic $100\alpha\%$ confidence interval

Exercise: Compute a confidence interval for the parameters in a Poisson process

t = length of observation period = 5 years

n = number of observed excesses = 31

λ = parameter (= yearly intensity =expected number of excesses per year) of Poisson process

Solution:

$$\mathrm{L}(\lambda) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

$$\log(\ell(\lambda)) = n\log(\lambda) + n\log(t) - \log(n!) - \lambda t$$

$$\frac{d\log(\ell(\lambda))}{d\lambda} = \frac{n}{\lambda} - t = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{n}{t} = \frac{31}{5} = 6.25$$

$$I(\lambda) = -\frac{d^2\log(\ell(\lambda))}{d\lambda^2} = \frac{n}{\lambda^2}$$

Estimated variance of $\hat{\lambda}$ is $-\frac{1}{I(\hat{\lambda})} = \frac{\hat{\lambda}^2}{n}$

95% confidence interval:

$$(\hat{\lambda} - \mathrm{k}_{0.975}\sqrt{\hat{\lambda}^2/n}, \hat{\lambda} + \mathrm{k}_{0.975}\sqrt{\hat{\lambda}^2/n})$$

$$= (6.25 - 1.96\frac{6.25}{\sqrt{31}}, 6.25 + 1.96\frac{6.25}{\sqrt{31}})$$

ML inference: the delta method

$\eta = g(\theta) = g(\theta_1, \ldots \theta_d)$ function of the parameters

$\hat{\eta} = g(\hat{\theta}) = g(\hat{\theta}_1, \ldots \hat{\theta}_d)$ estimate of the function of the parameters

$\nabla(\theta) = (\frac{\partial}{\partial\theta_1} g(\theta), \ldots \frac{\partial}{\partial\theta_d} g(\theta))$ gradient, $\nabla(\hat{\theta})$ estimate of gradient

$\hat{\eta}$ asymptotically normal with mean η and variance $\nabla(\theta)\mathcal{I}(\theta)^{-1}\nabla(\theta)^t$ (which e.g. can be estimated by $\nabla(\hat{\theta})I(\hat{\theta})^{-1}\nabla(\hat{\theta})^t$) .

From this one can construct confidence intervals for η in the same way as the confidence intervals for θ on the previous page.

Works well if g is approximatly linear, not so well otherwise

Exercise: Compute a confidence interval for the 95% quantile in a GP distribution based on observations $x_1, x_2, \dots x_N$

The GP density is $h(x;\sigma,\gamma) = \frac{1}{\sigma}\left(1 + \frac{\gamma}{\sigma}x\right)^{-\frac{1}{\gamma}-1}$

$$\log h(x;\sigma,\gamma) = -\log\sigma - \left(\frac{1}{\gamma} + 1\right)\log\left(1 + \frac{\gamma}{\sigma}x\right)$$

$$\frac{d}{d\sigma}\log h(x;\sigma,\gamma) = -\frac{1}{\sigma} + \frac{1+\gamma}{\sigma^2}\frac{x}{\left(1 + \frac{\gamma}{\sigma}x\right)}$$

$$\frac{d}{d\gamma}\log h(x;\sigma,\gamma) = \frac{1}{\gamma^2}\log\left(1 + \frac{\gamma}{\sigma}x\right) - \left(\frac{1}{\gamma} + 1\right)\frac{1}{\sigma}\frac{x}{\left(1 + \frac{\gamma}{\sigma}x\right)}$$

log likelihood function

$$\ell(\sigma,\gamma) = \sum_{i=1}^{N} \log h(x_i;\sigma,\gamma)$$

ML-estimates $\hat{\sigma}, \hat{\gamma}$ obtained as solutions to the equations

$$\sum_{i=1}^{N} \frac{d}{d\sigma} \log h(x_i;\sigma,\gamma) = 0$$

$$\sum_{i=1}^{N} \frac{d}{d\gamma} \log h(x_i;\sigma,\gamma) = 0$$

Observed information matrix obtained by inserting estimates into

$$I(\sigma,\gamma) = \begin{pmatrix} -\frac{d^2}{d\sigma^2}\ell(\sigma,\gamma) & -\frac{d^2}{d\sigma d\gamma}\ell(\sigma,\gamma) \\ -\frac{d^2}{d\sigma d\gamma}\ell(\sigma,\gamma) & -\frac{d^2}{d\gamma^2}\ell(\sigma,\gamma) \end{pmatrix}$$

The 95% quantile $x_{.95}$ in a GP distribution is obtained by solving

$$H(x_{.95}) = 1 - \left(1 + \frac{\gamma}{\sigma} x_{.95}\right)^{-\frac{1}{\gamma}} = 0.95$$

The solution is

$$g(\sigma, \gamma) = x_{.95} = \frac{\sigma}{\gamma}(0.05^{-\gamma} - 1\,)$$

so the maximum likelihood estimate is

$$\hat{x}_{.95} = \frac{\hat{\sigma}}{\hat{\gamma}}\left(0.05^{-\hat{\gamma}} - 1\right)$$

$$\nabla(\sigma, \gamma) = \left(\frac{d}{d\sigma} g(\sigma, \gamma), \frac{d}{d\gamma} g(\sigma, \gamma)\right)$$

$$= \left(\frac{1}{\gamma}(0.05^{-\gamma} - 1\,), -\frac{\sigma}{\gamma^2}(0.05^{-\gamma} - 1\,) - \frac{\gamma}{\sigma}(-\log 0.05)0.05^{-\gamma}\right)$$

The standard error of $\hat{x}_{.95}$ is estimated as

$$\mathrm{SE}(\hat{x}_{.95}) = \sqrt{\nabla(\hat{\sigma},\hat{\gamma}) \quad I(\hat{\sigma},\hat{\gamma})^{-1}\nabla(\hat{\sigma},\hat{\gamma})^T}$$

and a 95% asymptotic confidence interval is

$$(\hat{x}_{.95} - 1.96\mathrm{SE}(\widehat{x_{.95}}), \qquad \hat{x}_{.95} + 1.96\mathrm{SE}(\widehat{x_{.95}}))$$

ML inference: Likelihood Ratio (LR) tests

$\theta = (\theta_1, \theta_2)$ partition of θ into two vectors θ_1 and θ_2 of dimensions $d - p$ and p. $\hat{\theta}_2^*$ maximizes $l(\theta_1, \theta_2)$ over θ_2, for θ_1"kept fixed" (so function of θ_1)

$2(\ell(\hat{\theta}) - \ell(\theta_1, \hat{\theta}_2^*))$ asymptotically has a χ^2 distribution with $d - p$ degrees of freedom if θ_1 is the true value → LR test:

Reject $\mathrm{H}_0 : \theta_1 = \theta_1^0$ at the significance level α % if

$2\left(l(\hat{\theta}) - l\left(\theta_1^0, \hat{\theta}_2^*\right)\right) > \chi_\alpha^2(d-p)$, where $\chi_\alpha^2(d-p)$ is the $(1-\alpha)$-th quantile of the χ^2 distribution with $d - p$ degrees of freedom

ML inference: profile likelihood confidence intervals

(often more accurate than delta method intervals, ***plots from Coles***)

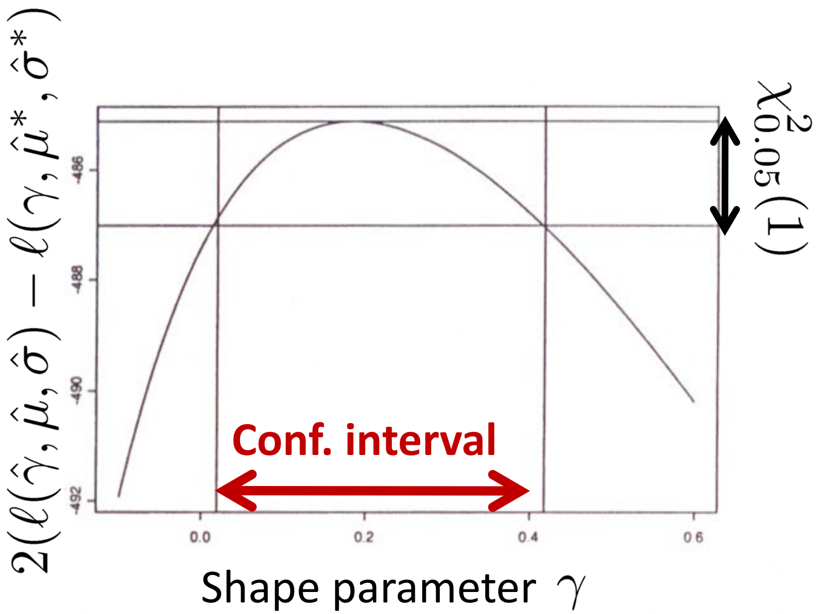


FIGURE 4.3. Profile likelihood for ξ in threshold excess model of daily rainfall data.

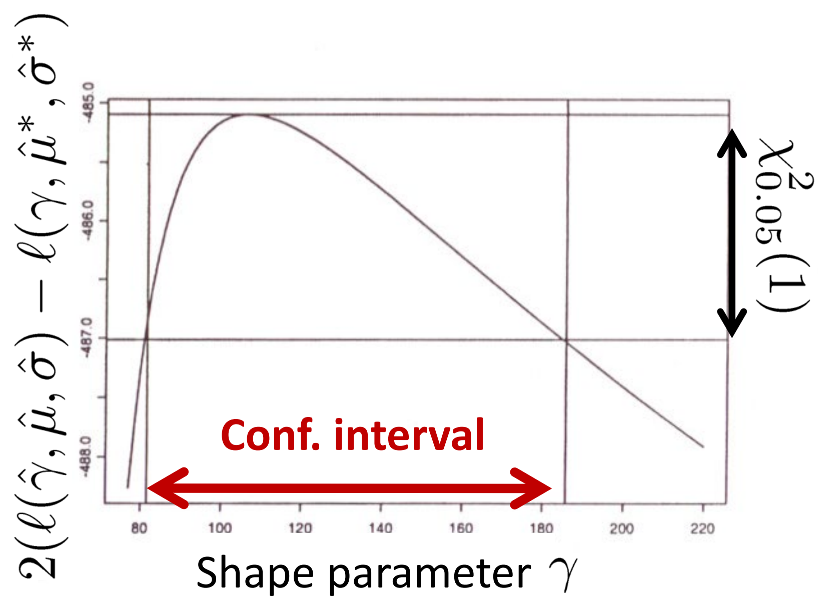


FIGURE 4.4. Profile likelihood for 100-year return level in threshold excess model of daily rainfall data.

Profile likelihood confidence intervals for the shape parameter in the Block Maxima model. The delta method would give similar interval in the left case, but not in the right.

$$x_p = VaR_p(L) \; = \; p\text{-th quantile of distribution of loss L}$$
$$= \text{solution to } \; F_L(x_p) = p$$

$$ES_p(L) = E\left(L \middle| L > VaR_p(L)\right) = \text{Expected Shortfall}$$

For the PoT model with threshold u suppose that $P(L > u) = p_u$. Then

$$VAR_p(L) = \frac{\sigma}{\gamma}\left\{\left(\frac{1-p}{p_u}\right)^{-\gamma} - 1\right\} + u, \qquad \text{for } p > p_u$$

provided this value is greater than u, and

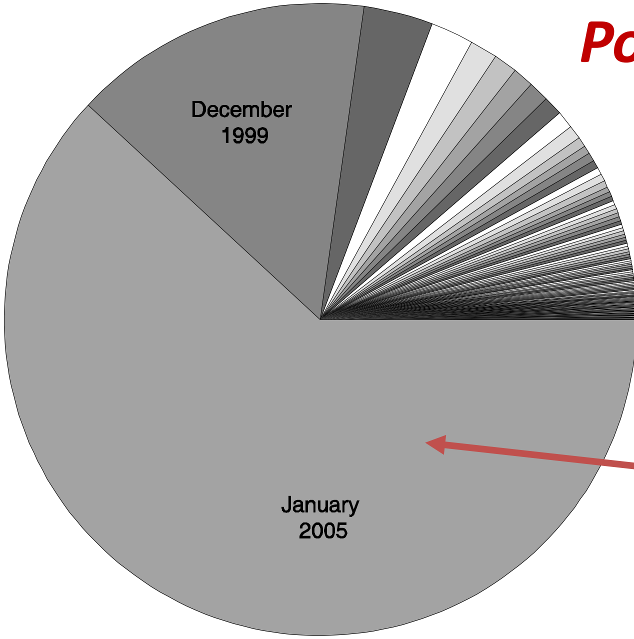
$$ES_p(L) = VAR_p(L) + \frac{\sigma + \gamma(VAR_p(L) - u)}{1-\gamma}$$

Important exercise: Check if the formulas on the previous page are correct.

$VaR_p(L)$ and $ES_p(L)$ are estimated by replacing σ, γ in the formulas on the previous page by their estimates $\hat{\sigma}, \hat{\gamma}$ and replacing p_u by its estimate

$$p_u^* = \frac{\#\text{ excesses of } u}{\#\text{observations}}$$

PoT: windstorm insurance (Rootzén& Tajvidi)



Windstorm losses for Länsförsäkringar 1982-2005

Gudrun January 2005
326 MEuro loss
72 % due to forest losses
4 times larger than second largest

The real problem!

The problems

How much reinsurance should LFAB buy?

Should LFAB worry about windstorm losses getting worse?

How should LFAB adjust if its forest insurance portfolio grows?

and:

Can detailed modeling give better risk estimates?

Are windstorms becoming more frequent?

1994 PoT analysis of 1982-1993 LFAB data (the basic method, more sophisticated analysis of 1982-2005 data in later paper)

Risk (MSEK)	next year	next 5 years	next 15 years
10%	66	215	473
1%	366	1149	2497

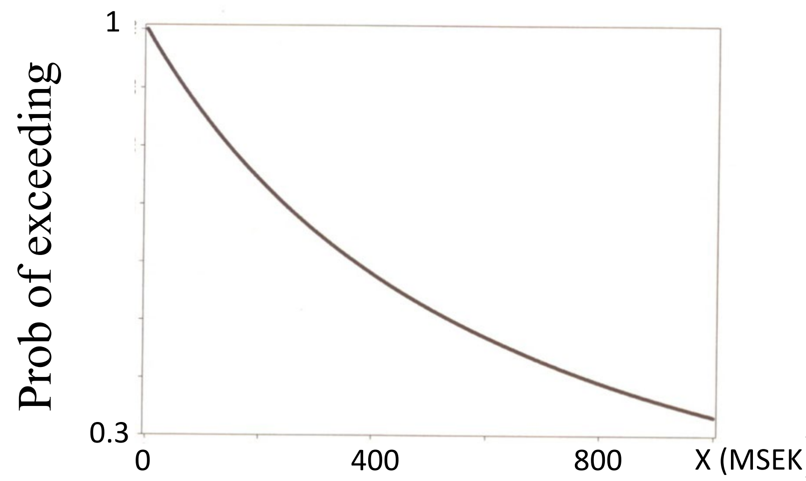
$X_i \;\; \mathrm{GP}(y\,;\sigma_t,\gamma)$

$\sigma_t = \exp(\alpha + \beta t)$

$\hat{\alpha} = 0.93$

$\hat{\beta} = .013 \pm .013$

no evidence of trend in extremes



conditional probability that a loss in excess of the reinsurance level 850 MSEK exceeds x

Gudrun: 2912 MSEK, 12 years later

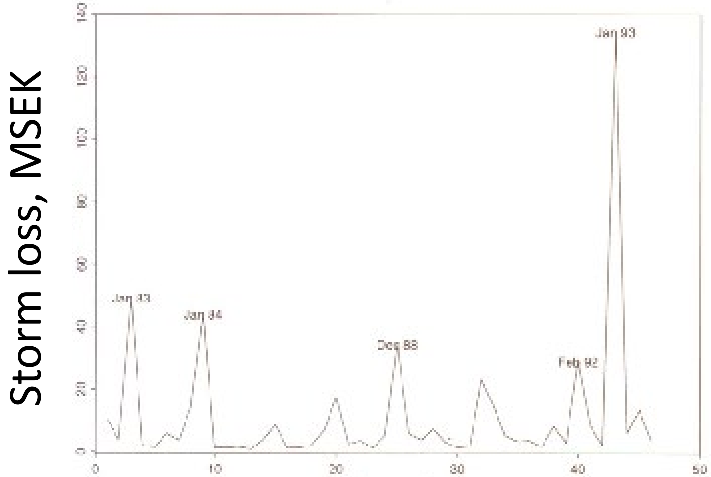
Windstorms of 1902 and 1969 probably comparable to Gudrun

Choice of threshold/number of order statistics in PoT, model diagnostics

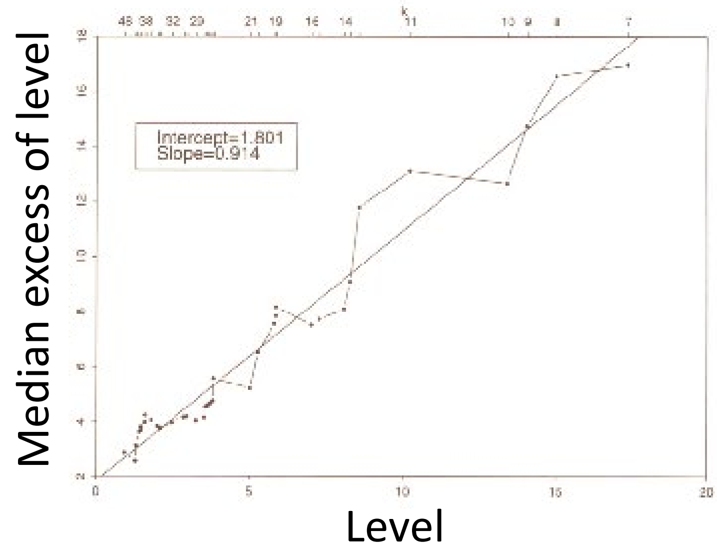
Threshold choice compromise between low bias (= good fit of model): requires high threshold/few order statistics, and low variance: requires low threshold/many order statistics

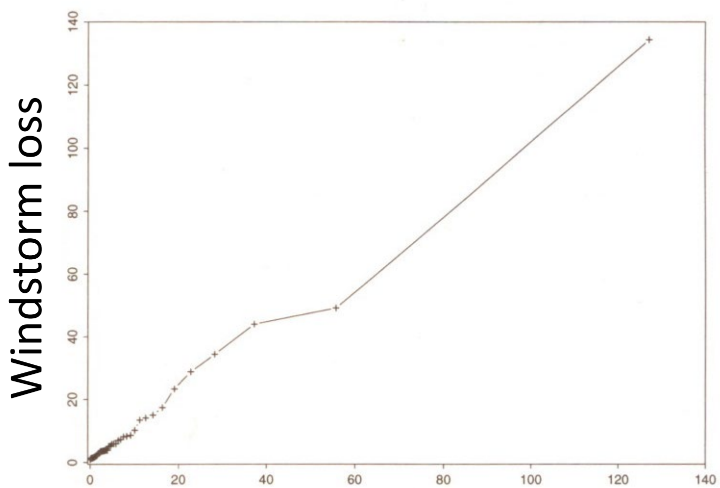
- mean excess plots (high variability for heavy tails)
- median excess plots
- plots of parameter estimates as function of threshold/number of order statistics
- qq- and pp-plots

automatic threshold selection procedures exist, and are getting better, but still “optimal” threshold depends on the unknown underlying distribution which has to be estimated.

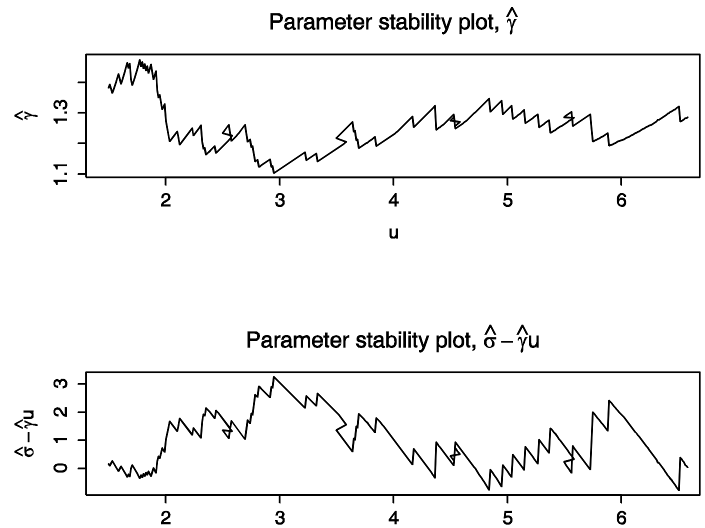


Windstorm losses 1982-1993
excesses of 0.9 MSEK





Quantiles of GPD



Windstorm losses 1982-2005
parameter stability plots

Some conclusions

- risk cannot be summarized into one number
- extreme value statistics provide the simplest methods (but other methods may sometimes be needed)
- didn’t find clear trends
- meteorological data didn’t help
- don’t trust computer simulation models unless statistically validated
- companies should develop systematic techniques for thinking about “not yet seen” catastrophes

- put contractual limits to aggregate exposure

A step in another direction:
catastrophe risks

BIG --- ”happens only once”

- can’t adjust and improve as experience is gained
- methods based on means, variances, central limit theory have little meaning
- difficult to keep in mind that catastrophes can (and will!) occur

a gamble --- find the odds of a gamble!

O. Perrin, H. Rootzén, and R. Taessler, A discussion of statistical methods used to estimate extreme wind speeds. *Theor. Appl. Climatol. 85*, 203–215 (2006)

H. Rootzén and N. Tajvidi (1997). Extreme value statistics and wind storm losses: a case study. *Scand. Actuarial J.,* 70-94, reprinted in "Extremes and integrated risk management", Risk Books 2000.

H. Rootzén, and N.Tajvidi (2000). Can losses caused by wind storms be predicted from meteorological observations? *Scand. Actuarial J.,* 162-175.

H. Rootzén, and N.Tajvidi (2006). The multivariate Generalized Pareto Distribution. *Bernoulli 12*, 917-930

E. Brodin and H. Rootzén (2009). Modelling and predicting extreme wind storm losses. *Insurance, Mathematics and Economics, 44* , 345-356