Extreme value statistics for financial risk

Lecture 1: Probability theory

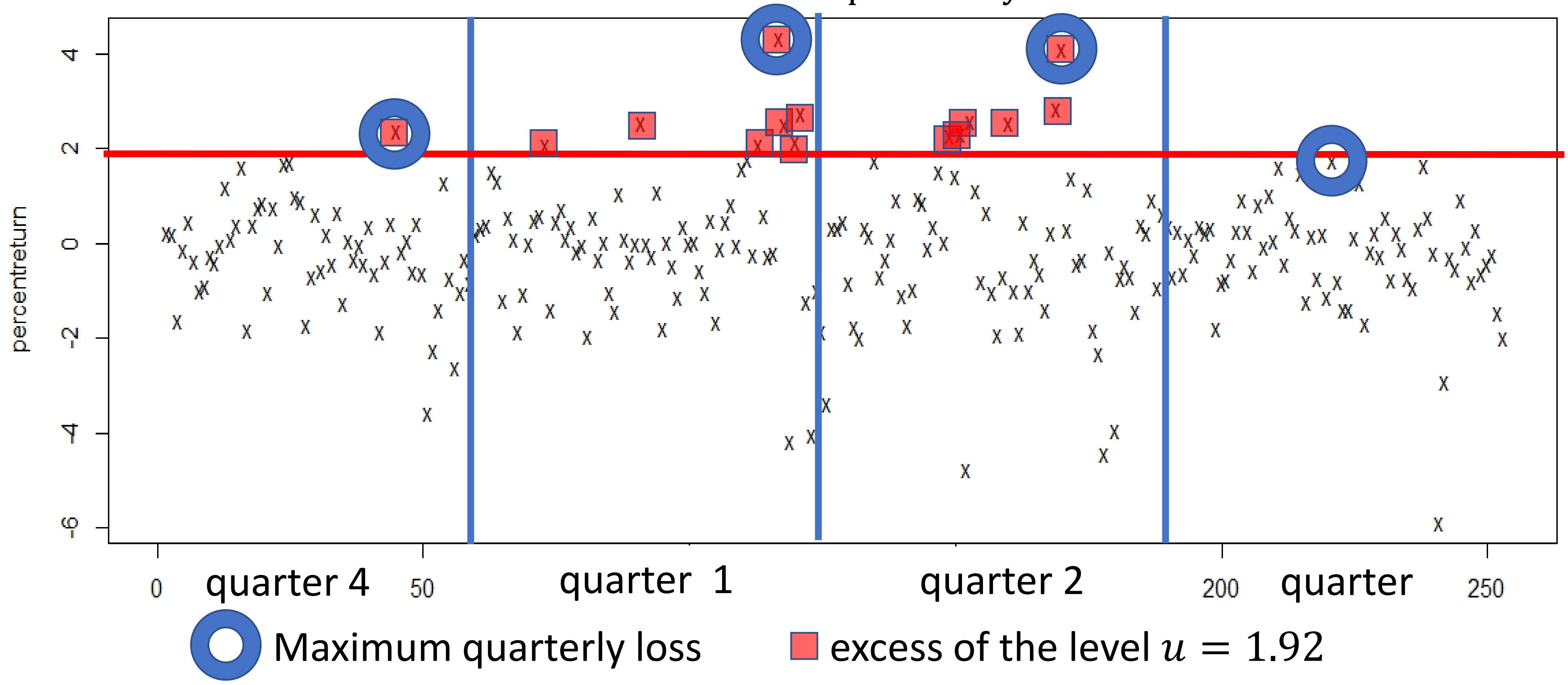
Lecture 2: Block maxima + PoT

Lecture 3: PoT + program packages

Lecture 4: Programming + multivariate block maxima

Lecture 5: Multivariate PoT

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

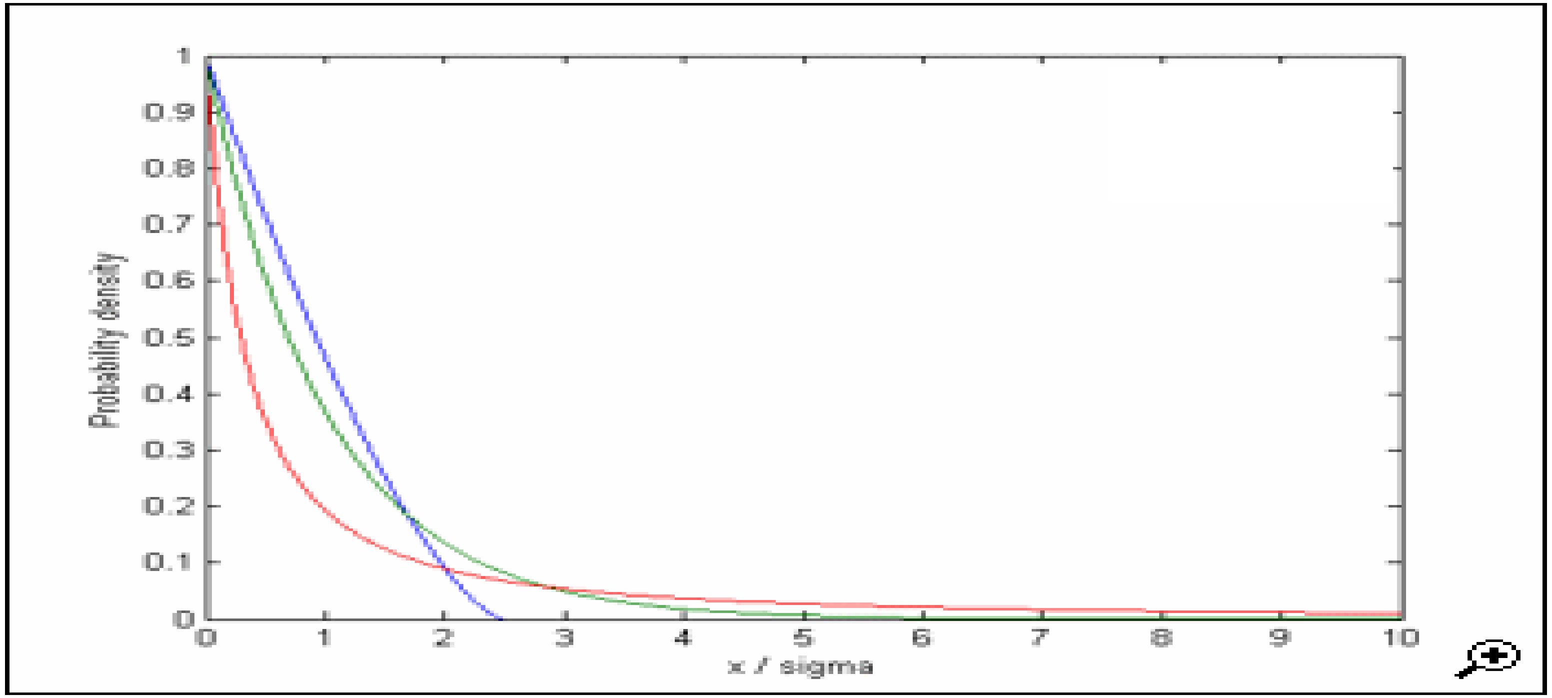


How large is the risk of a big quarterly loss?

How large is the risk of a big loss tomorrow?

Generalized Pareto(GP) distributions

$$H(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)_{+}^{-1/\gamma}$$



$\gamma > 0$ left endpoint 0, right endpoint ∞, heavytailed

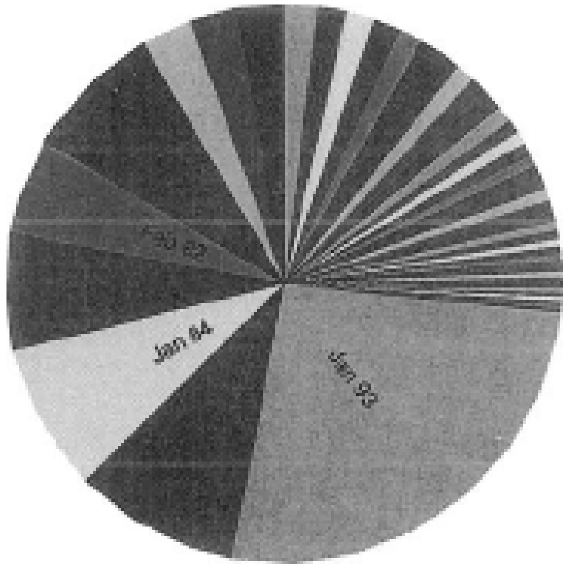
$\gamma = 0$ cdf $H(x) = 1 - e^{-x/\sigma}$, exponential

$\gamma < 0$ left endpoint 0, right endpoint $\sigma/|\gamma|$

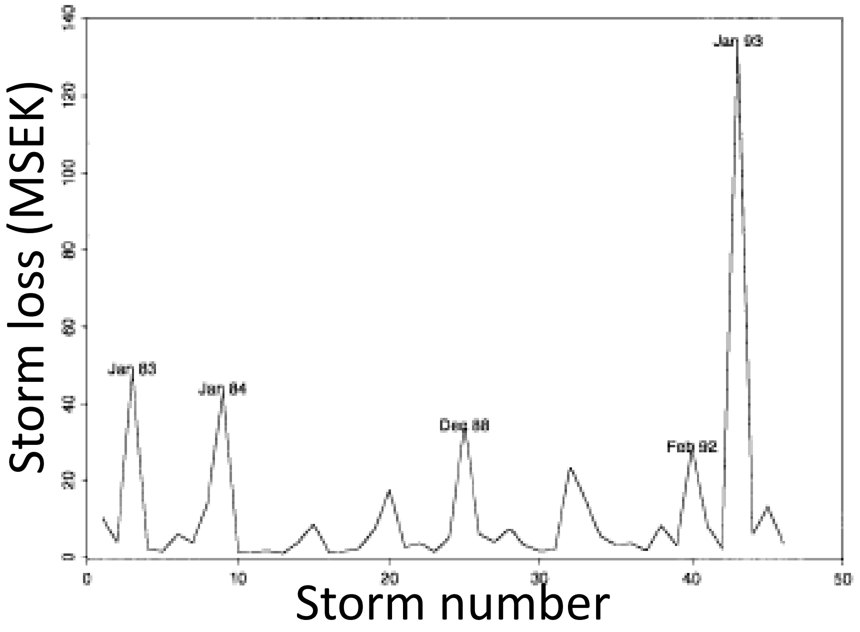
Peaks over thresholds (PoT) method

- Choose (high) threshold u and from i.i.d observations $Y_1, \dots, Y_n \sim F$ obtain N threshold excesses $X_1 = Y_{t_1} - u, \dots, X_N = Y_{t_N} - u$, where $t_1, \dots, t_N$ are the times of threshold exceedance
- Assume $X_1, \dots, X_N$ are i.i.d and follow a GP distribution and that $t_1, \dots, t_N$ are the occurrence times of an independent Poisson process, so that N has a Poisson distribution
- Use $X_1, \dots, X_n$ to estimate the parameters of the GP distribution and N to estimate the mean of the Poisson distribution
- Estimate tail $\bar{F}(x) = 1 - F(x) = \bar{F}(u)\bar{F}_u(x-u)$, where $\bar{F}_u(x-u)$ is the conditional distribution of threshold excesses, by

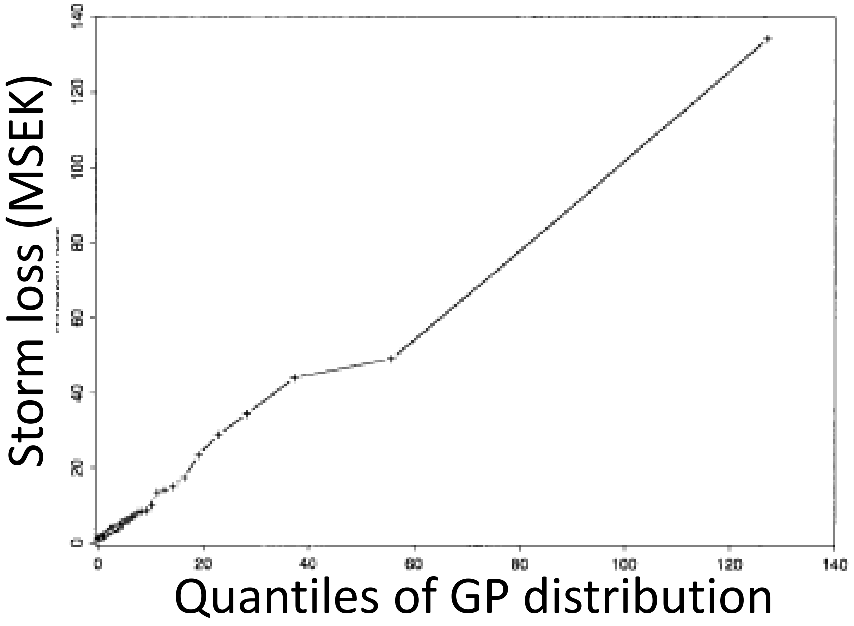
$$\hat{\bar{F}}(x) = \frac{N}{n}\,\hat{H}(x-u)$$



Pie chart of LFAB wind-storm losses 1982-1993

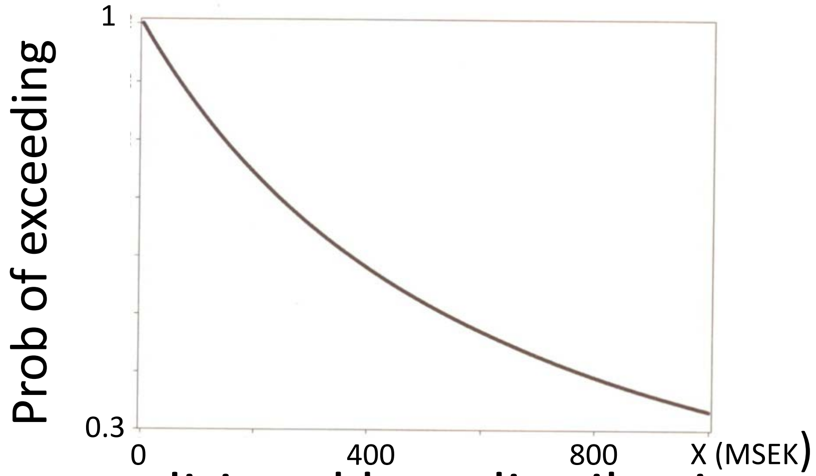


LFAB windstorm losses 1982-1993



qq-plot against GP distribution

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



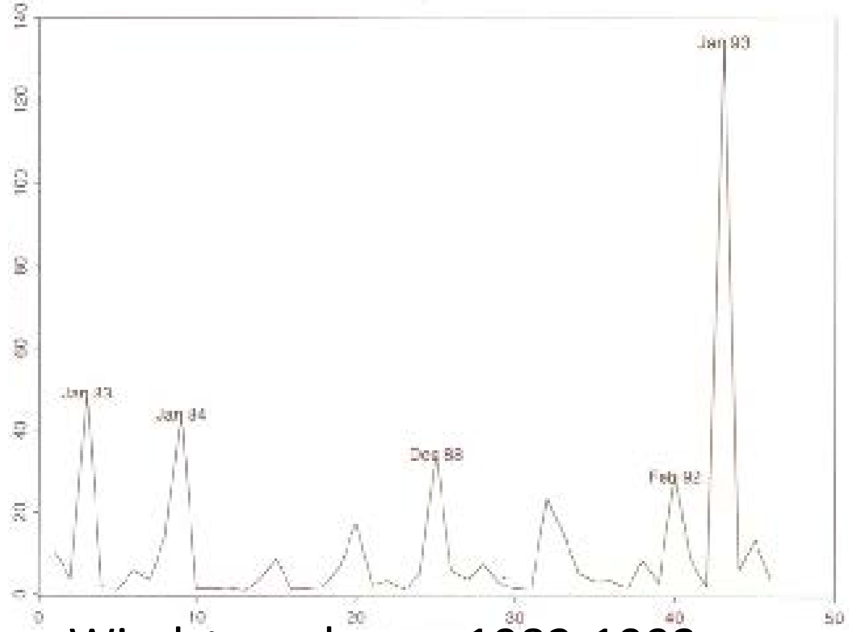
conditional loss distribution

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

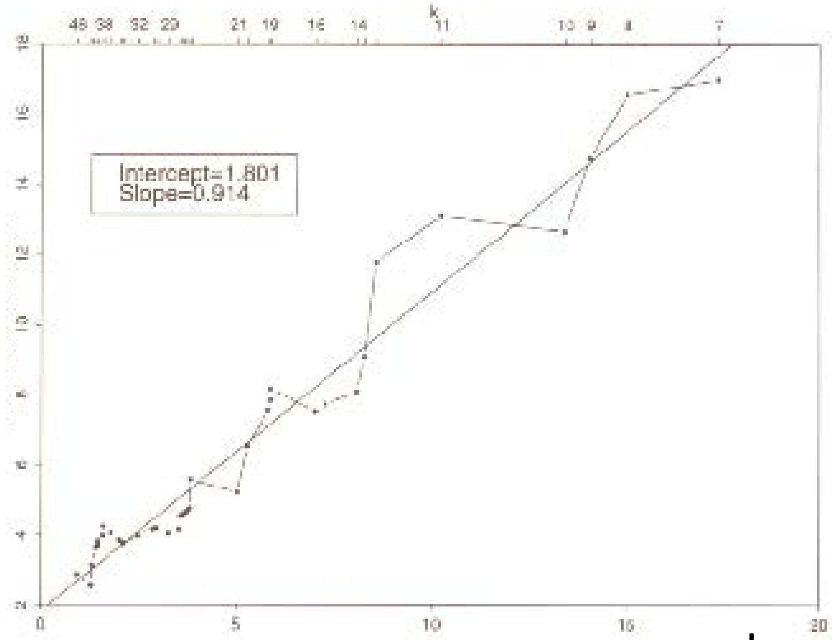
Threshold choice compromise between low bias (= good fit of model), which requires high threshold/few order statistics, and low variance, which requires low threshold/many order statistics. Tools aiding choice include

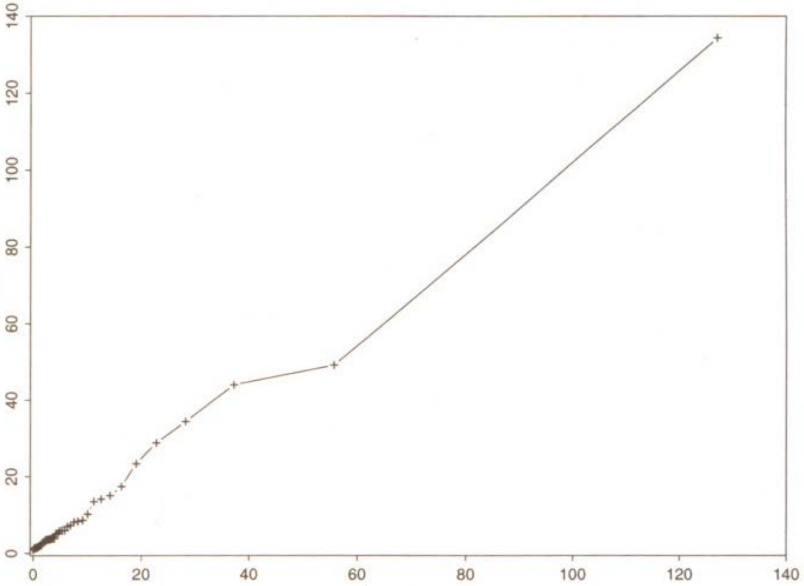
- mean excess plots (high variability for heavy tails)
- median excess plots
- plots of parameter estimates as function of threshold/number of order statistics
- qq- and pp-plots

automatic threshold selection procedures exist, but perhaps not always reliable (“optimal” threshold depends on the underlying unknown distribution)

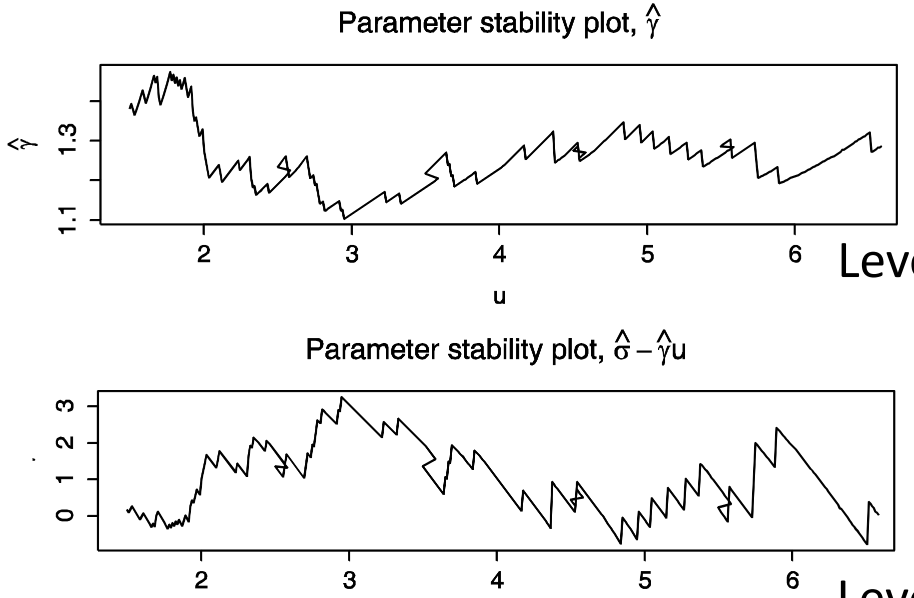


Windstorm losses 1982-1993
3-day excesses of 0.9 MSEK





Quantiles of GP distribution



parameter stability plots

Dependence

- *Block maxima method*: If blocks are long, then block maxima (typically) are approximately independent, and one can proceed as for i.i.d. observations
- *PoT:* Standard method is to use i.i.d. PoT method with excesses replaced by cluster maxima, and exceedance times replaced by the times when cluster maxima occur. (Though sometimes it may be better to use all excesses: work in progress with H. Drees, A. Janssen)
- *PoT:* Cluster identification partly "skill", using plotting, subject matter knowledge, … some automatic methods have been proposed
- *PoT:* Estimate extremal index by $\hat{\theta} = \frac{\#\text{clusters}}{\#\text{exceedances}}$
- PoT: Use $P\,r\left(\vee_{i=1}^{n} \mathrm{X}_i \leq x\right) \approx F(x)^{n\theta}$ to switch between PoT and block maxima

99% ***Value at Risk*** = 99% VaR = $.99^{th}$ quantile of the distribution of the losses (L) = solution $x_{0.99}$ to the equation $F_L(x_{0.99}) = 0.99$, where F_L is the estimated cdf of the losses (which can be daily negative returns, weekly negative returns, maximum weekly negative returns, total insured loss in a windstorm, ...)

- *Block maxima:* $F_L(x) = G(x; \hat{\mu}, \hat{\sigma}, \hat{\gamma})$, where G is the GEV cdf and $\hat{\mu}, \hat{\sigma}, \hat{\gamma}$ are the estimated parameters

- *PoT:* $F_L(x) = 1 - \frac{N}{n}(1 - H(x - u;\ \hat{\sigma}, \hat{\gamma}))$, where H is the GP cdf, $\hat{\sigma}, \hat{\gamma}$ are the estimated parameters, u is the threshold, n is the number of observations, and N is the number of excesses

99% ***Expected Shortfall*** = $ES_{0.99}(L) = E(L|L > VaR_{0.99}(L))$. Use of ES is motivated by axioms – but for heavytailed distributions ES estimates are more variable, and the axioms are not always satisfied in practice. ES is given by similar formulas as above

From: E. Gilleland (2015) Computing software, Ch 25 in the 2016 book “Extreme Value Modeling and Risk Analysis: Methods and Applications” by Dipak K. Dey, Jun Yan, Chapman and Hall/CRC. New survey will presumably appear in the journal extremes in 2019 or 2020

R packages for univariate EVS

- evd (Stephenson, 2002)
- evdbayes (Stephenson and Ribatet, 2014)
- evir (Pfaff and McNeil, 2012)
- **extRemes (Gilleland and Katz, 2011a, 2015)**
- in2extRemes (Gilleland and Katz, 2011a, 2015) **GUI**
- extremevalues (van der Loo, 2010a,b) **GUI**
- fExtremes (Wuertz et al., 2013)

R packages for univariate EVS, ct’d

- ismev (Heffernan and Stephenson, 2012)
- lmom (Hosking, 2014a)
- lmomRFA (Hosking, 2014b)
- lmomco (Asquith, 2014)
- texmex (Southworth and Heffernan, 2013)
- VGAM (Yee, 2013; Yee and Stephenson, 2007)

R packages for multivariate EVS

- copula (Hofert et al., 2014; Hofert and Maechler, 2011; Kojadinovic and Yan, 2010; Yan, 2007)
- evd (Stephenson, 2002)
- evir (Pfaff and McNeil, 2012)
- lmomRFA (Hosking, 2014b)
- lmomco (Asquith, 2014)
- **SpatialExtremes (Ribatet et al., 2013)**
- texmex (Southworth and Heffernan, 2013)

MATLAB packages

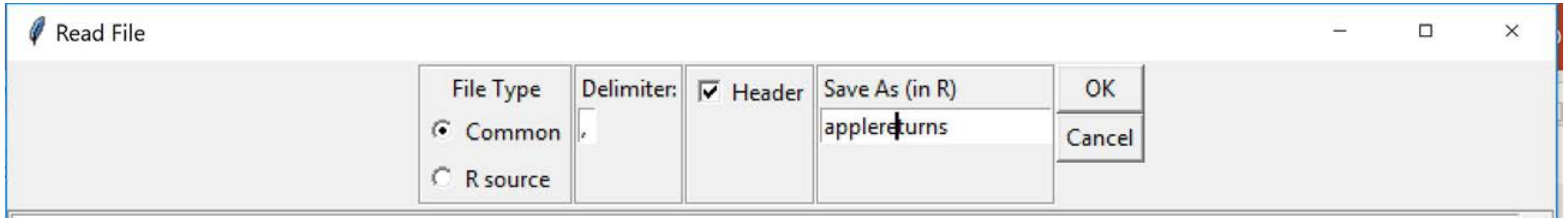
- EVIM, MATLAB 1 package (Gencay et al., 2001, available at: http://www.sfu.ca/˜rgencay/) for univariate EVS
- NEVA (Cheng et al., 2014) package for univariate EVS
- The WAFO (Brodtkorb et al., 2000) MATLAB and Python package focused on wave analysis for oceanography, sea modeling, andfatigue analysis. Also contains the customary block maxima and POT approaches
- MATLAB itself also has some EVS capabilities

Standalone Packages

- EXTREMES (available at: http://extremes.gforge.inria.fr/) point-and-click software for univariate EVS
- GLSnet and peakFQ, freely available Fortran packages from the U.S. Geological Survey's Water Resources Applications Software (http://water.usgs.gov/software). Aimed at modeling flow characteristics, do not contain any extreme value distribution functionality.
- Package HYFRAN and HYFRAN-PLUS (El Aldouni et al., 2008) are commercial products originally designed for hydrological frequency analysis, but has grown to include tools for fitting numerous distributions with different tail weights to i.i.d. data..
- Wallingford HydroSolutions provides various commercial software products
- A Windows package, Xtremes Version 3.1, accompanies the EVT book Reiss and Thomas, 2007. Can performing most EVA tasks. The STABLE routines written by Nolan (2007) are included with Xtremes.

Use of in2extRemes

- Download R
- Click on "packages", click on "install package(s)", choose in2extRemes from list
- Click on "packages", click on "load package" choose in2extRemes from list
- Type in2extRemes() on workspace to open GUI
- Import .cls (Excel can easily be converted to cls) files with column headers by choosing "file" and then "read data" and do as in picture



- Then you can try out the functionalities in in2extRemes

If $F(x_p) = 1 - p$ then x_p is called the "***return level***" and $1/p$ is the "***return period***"