Extreme value statistics for financial risk

Lecture 1: Probability theory

Lecture 2: Block maxima + PoT

Lecture 3: PoT + program packages

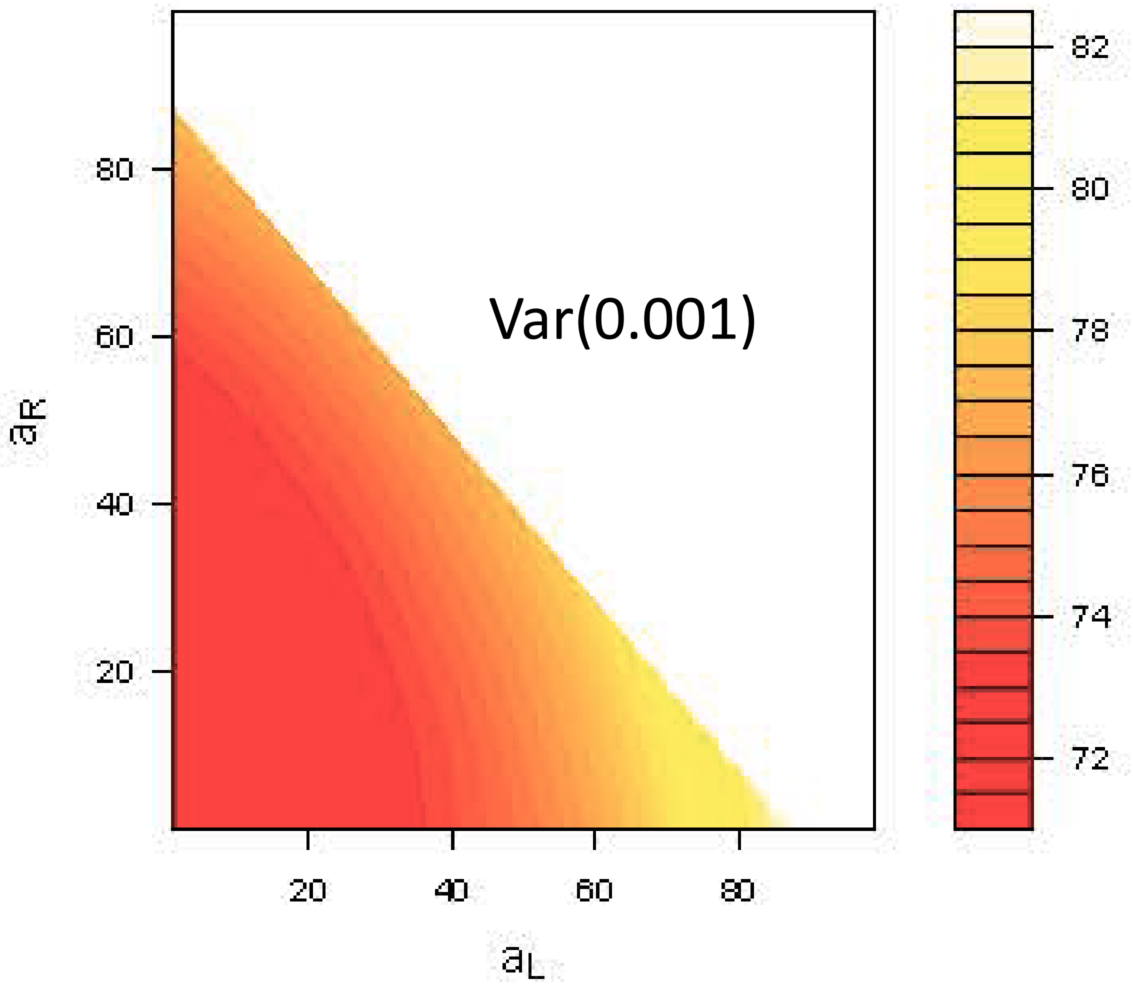
Lecture 4: Programming + multivariate block maxima

Lecture 5: Multivariate PoT

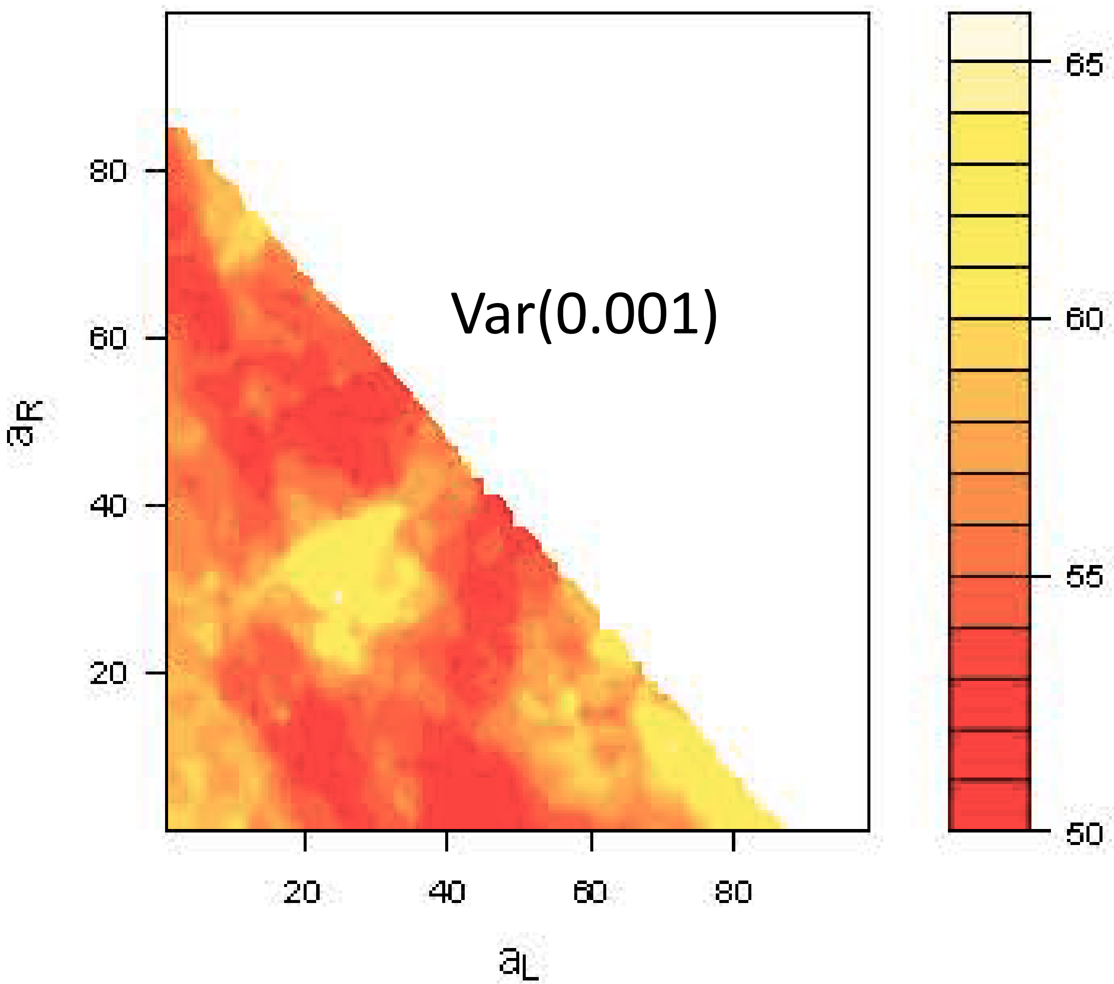
Multivariate peaks over thresholds modelling and likelihood inference
Or:
News from the exciting exploration of GP-territory

Joint with Anna Kiriliouk, Johan Segers, Maud Thomas, Jennifer Wadsworth

Portfolio: HCBC 10%, Loyds x-axis (%), RBS y-axis (%), Barclays 100-10-x-y%



4-dimensional GP model



1-dimensional GP model fitted separately to each portfolio

The philosophy is simple

- Extreme episodes are often quite different from ordinary everyday behavior, and ordinary behavior then has little to say about extremes, so that only other extreme events give useful information about future extreme events

- Operationally: choose thresholds $u_1, \ldots, u_d$, say that an extreme episode occurs if at least one of the components of the observation $\boldsymbol{Y} = (Y_1, \ldots, \mathrm{Y}_d)$ exceeds its threshold, only model times of occurence and undershoots and overshoots, $\mathbf{X} = \boldsymbol{Y} - \boldsymbol{u}$, of extreme episodes

- Theory: times of occurrence follows a Poisson process, $\mathbf{X}$ is a generalized Pareto (GP) distribution

Theoretically motivated models give much better possibilities to learn from experience than if everyone uses their own ad hoc method

Notation

Bold symbols are d-variate vectors. For instance, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)$ and $\mathbf{0} = (0, \ldots, 0) \in R^d$. Operations and relations are componentwise, with shorter vectors recycled.

For instance $\boldsymbol{a}\boldsymbol{x} + \boldsymbol{b} = (a_1x_1 + b_1, \ldots, a_dx_d + b_d)$, $\boldsymbol{x} \leq \boldsymbol{y}$ if $x_j \leq y_j$ for $j = 1, \ldots, d$, and $t^{\boldsymbol{\gamma}} = (t^{\gamma_1}, \ldots, t^{\gamma_d})$

Multivariate Generalized Pareto(GP) distributions

$$H(\boldsymbol{x}) = \begin{cases} \dfrac{1}{-\log G(\mathbf{0})} \log \dfrac{G(\boldsymbol{x})}{G(\boldsymbol{x} \wedge \mathbf{0})}, & \text{if } \boldsymbol{x} > \boldsymbol{\eta} \\ 0, & if\ x_i < \eta_i, \text{some } j = 1. \dots d \end{cases}$$

where G is a multivariate GEV cdf with $G(\mathbf{0}) > 0$

The GP distributions are the limit distributions threshold excesses: Let $\boldsymbol{X} \sim F$. If there exist continuous threshold and scaling functions $\boldsymbol{u}_t$ and $\boldsymbol{s}_t > 0$, with $F(\boldsymbol{u}_t) < 1$ and $F(\boldsymbol{u}_t) \to 1$ as $t \to \infty$, such that

$$Pr\left(\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t) \vee \boldsymbol{\eta} \leq \boldsymbol{x} \mid \boldsymbol{X} \nleq \boldsymbol{u}_t\right) \to_d H(\boldsymbol{x}), \qquad \text{as } n \to \infty,$$

where H has non-degenerate margins, then H is a GP distribution. Conversely all GP distributions can be obtained in this way.

- maxima of F are in “the domain of attraction of a GEV cdf” if and only if threshold excesses of F are in “the threshold domain of attraction” of a GP cdf H"

The GP distributions are the threshold-stable distribution: Let $\boldsymbol{X} \sim H$, with H nondegenerate. If there exist continuous scaling and threshold functions $\boldsymbol{s}_t \geq \boldsymbol{0}, \boldsymbol{u}_t$ with $F(\boldsymbol{u}_t) < 1$ and $F(\boldsymbol{u}_t) \to 1$ as $t \to \infty$, such that

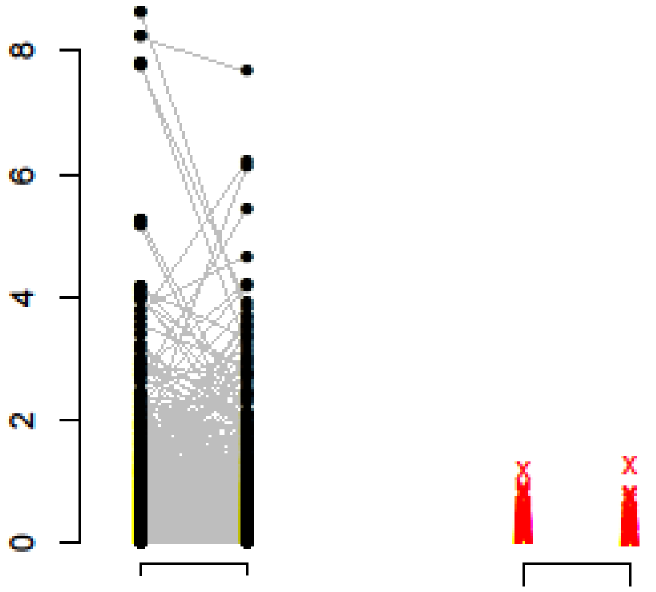
$$Pr\left(\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t) \leq \boldsymbol{x} \mid \boldsymbol{X} \not> \boldsymbol{u}_t\right) = H(\boldsymbol{x}), \qquad \text{for all } t \geq 1,$$

then H is a GP distribution. Conversely all GP distributions has this property.

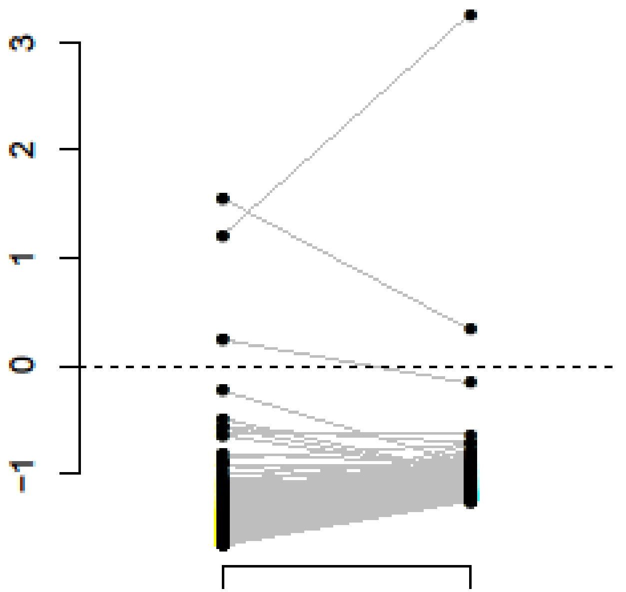
Multivariate GP distributions

- Marginal distributions conditional of being positive are 1-dimensional GP-s
- No finite-dimensional parametrization
- Likelihood inference requires development of realistic and computationally tractable parametric subclasses
- The possibility of mass on lower boundaries leads to new and not well understood modelling challenges
- If no mass on lower boundaries, then ***full asymptotic dependence*** → this is assumed in the sequel

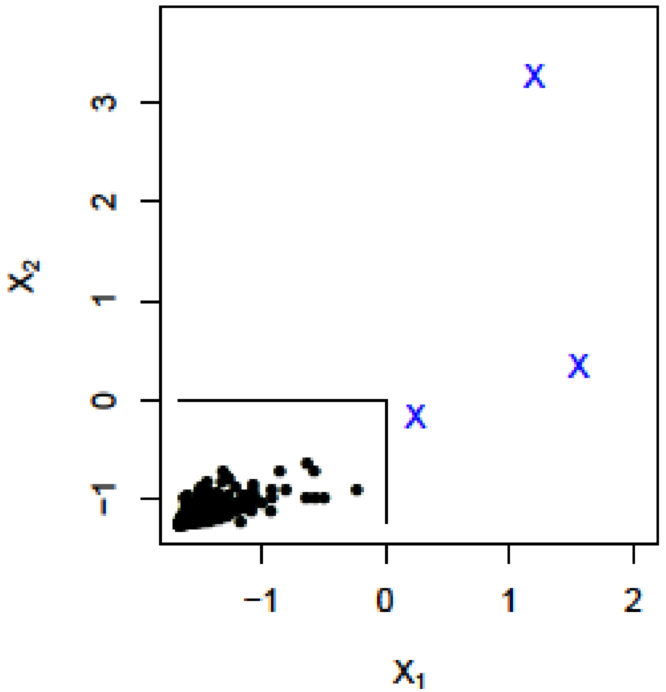
A general construction



R_1 R_2 $T^{-\gamma}$



$$X_i = R_i T^{-\gamma} - \frac{\sigma_i}{\gamma_i}, i = 1,2$$



Components

$R = (R_1, R_2)$ “any” vector
T “uniformly distributed on $(0, \infty)$”

Extreme event

GEV = distribution of vector consisting of the two maxima

GP distribution = conditional distribution of distance from blue crosses to threshold

The mechanism behind the GEV and GP models and distributions

four general representations of GP distributions

R) From previous slide: constructed on real scale; indata $\mathbf{R}, \boldsymbol{\sigma}, \boldsymbol{\gamma}$; distribution function

$$H_R(\boldsymbol{x}) = \frac{\int_0^\infty \left\{F_{\boldsymbol{R}}\left(t^{\gamma}\left(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\right)\right) dt - F_{\boldsymbol{R}}\left(t^{\gamma}\left(\boldsymbol{x} \wedge \mathbf{0} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\right)\right)\right\} dt}{\int_0^\infty \bar{F}_{\boldsymbol{R}}(t^{\gamma} \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) dt}$$

U) Use “R-representation” on standardized scale $\boldsymbol{\gamma} = \mathbf{0}$ to get $\mathbf{X}_0$; indata $\boldsymbol{U}, \boldsymbol{\sigma}, \boldsymbol{\gamma}$; transform to real scale via

$$\boldsymbol{X} = \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}(\boldsymbol{e}^{\gamma \boldsymbol{X}_0} - \mathbf{1})$$

S) Ferreira –de Haan: indata $\mathbf{S}, \boldsymbol{\sigma}, \boldsymbol{\gamma}$, with $\max\{S_1, S_2, \dots S_d\} = 0$, put $\boldsymbol{X_0} = \boldsymbol{S} + E$, with E independent of $\boldsymbol{S}$ and exponentially distributed, get general distribution through

$$\boldsymbol{X} = \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}(\boldsymbol{e}^{\gamma X_0} - \mathbf{1})$$

T) indata $\mathbf{T}\ \boldsymbol{\sigma}, \boldsymbol{\gamma}$; set $\boldsymbol{S} = \mathrm{T} - \max\{T_1, T_2, \dots T_d\}$, use Ferreira de Haan with this $\mathbf{S}$

R) can respect physical constraints, prediction natural, discontinuous at $\gamma = 0$

U) prediction natural, continuous at $\gamma = 0$

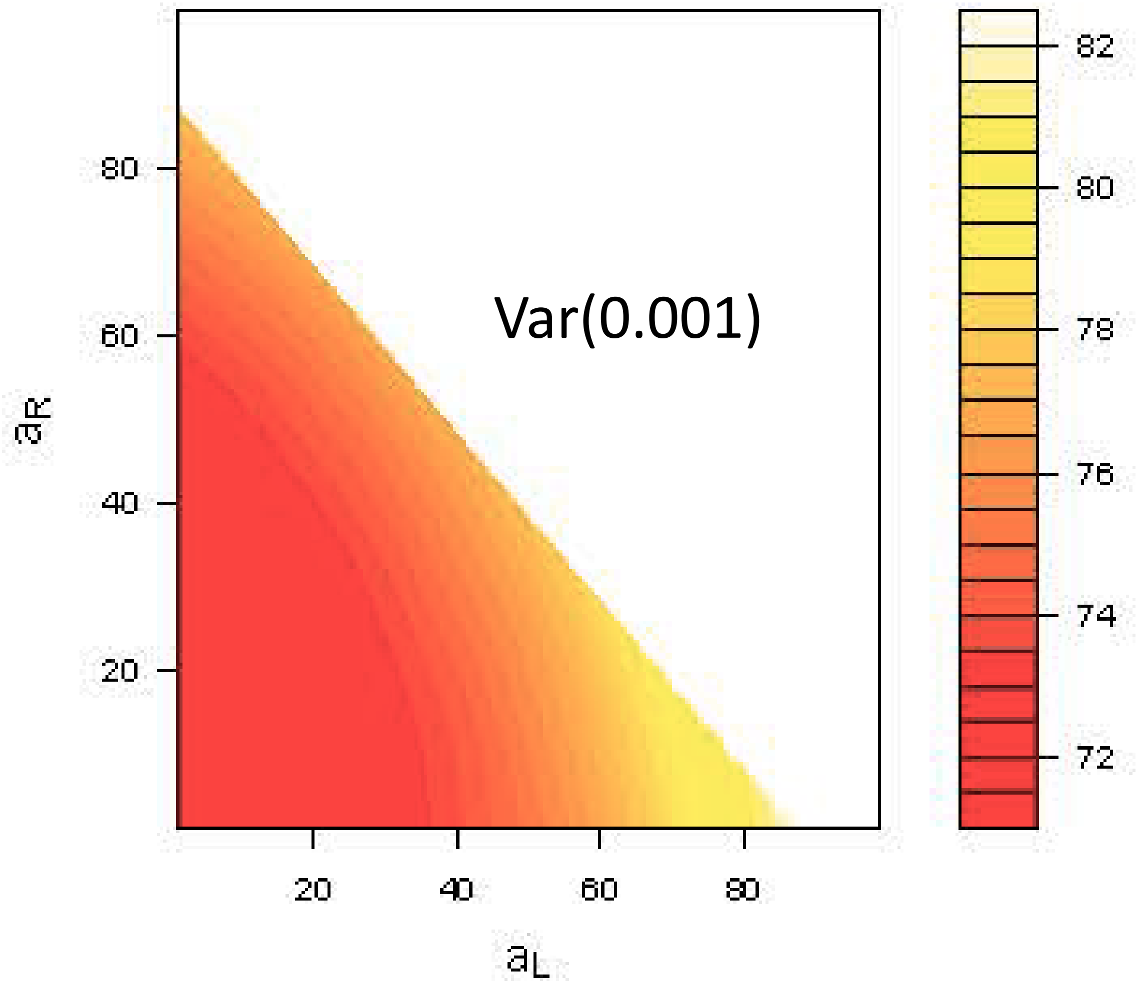
S) computations simplest, continuous at $\gamma = 0$; prediction & lower-dimensional margins unnatural, parametric models not obvious

T) continuous at $\sigma = 0$; prediction & lower-dimensional margins unnatural

Densities

- Explicit expressions for densities for all representations
- Expressions in terms of one-dimensional integrals of densities and c.d.f.-s of the $\boldsymbol{R}, \boldsymbol{U}, \boldsymbol{S}, \text{or } \boldsymbol{T}$ cdf-s
- Sometimes possible to get explicit expressions.
- Censored likelihood contributions computable
- Still, computation is a challenge

Portfolio: HCBC 10%, Loyds x-axis (%), RBS y-axis (%), Barclays 100-10-x-y%

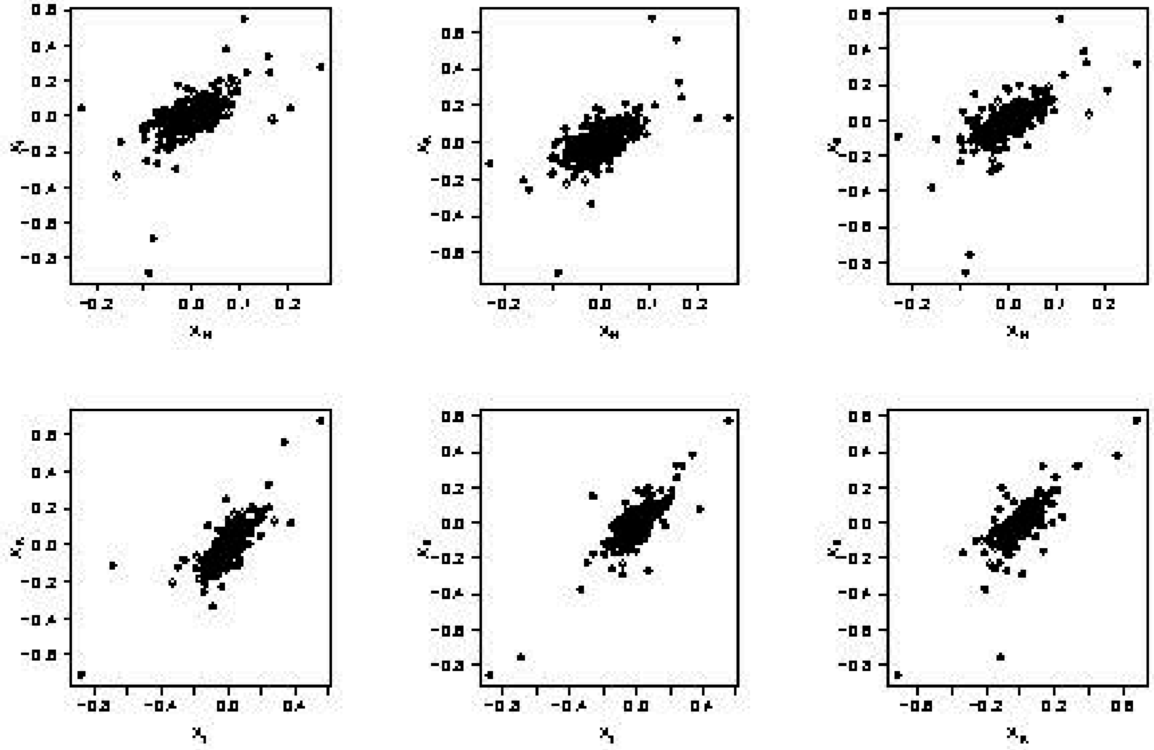


4-dimensional GP model

How was this done?

Banks

- Data: weekly negative returns of four largest UK banks, October 2007 - April 2016 (444 observations)



Pairwise scatterplots

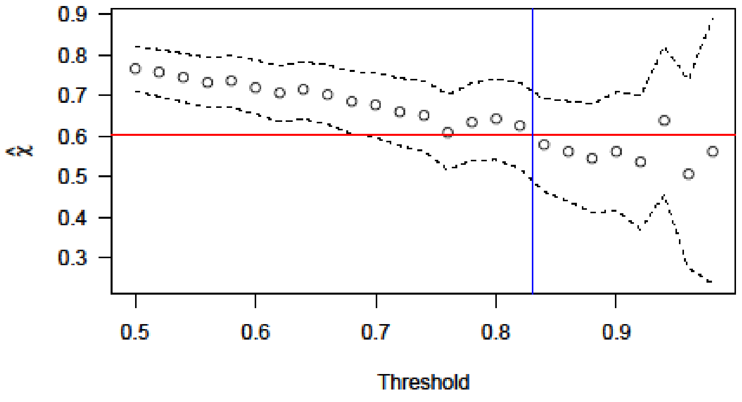
Selected "best" model from five:

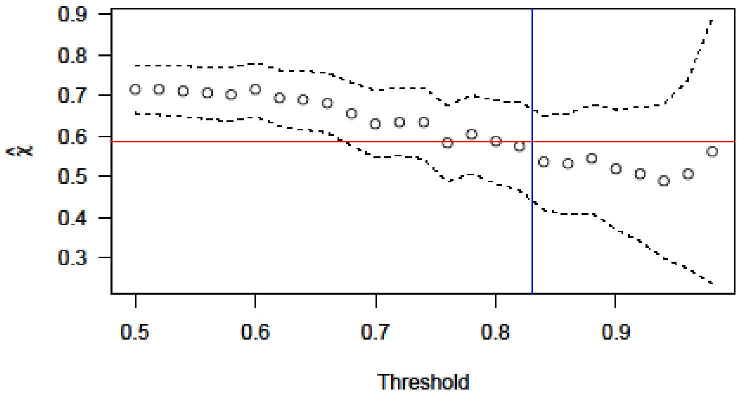
1. T-model with independent Gumbel components for $\boldsymbol{T}$
2. U-model with independent Gumbel components for $\boldsymbol{U}$
3. T-model with independent reverse exponential components for $\boldsymbol{T}$
4. U-model with independent reverse exponential components for $\boldsymbol{U}$
5. U-model with independent normal components for $\boldsymbol{T}$

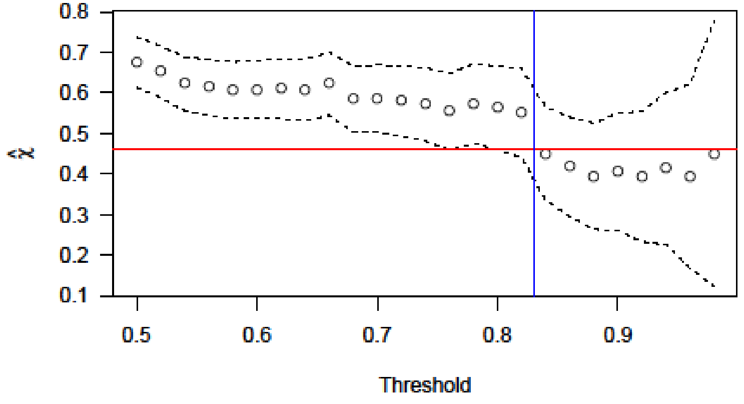
(Remember, in R-model the GP distribution is "the distribution of $\boldsymbol{R} \times T^{-\gamma}$", so the GP distribution has dependent components even if $\boldsymbol{R}$ has independent components. Dependence stronger the smaller the variation of the components in $\boldsymbol{R}$. Same for the other three models.)

Threshold choice

- first standard 1-d marginal threshold choice for each margin
- then use χ-statistics to check if thresholds have to be increased
- (other threshold selection methods developed)





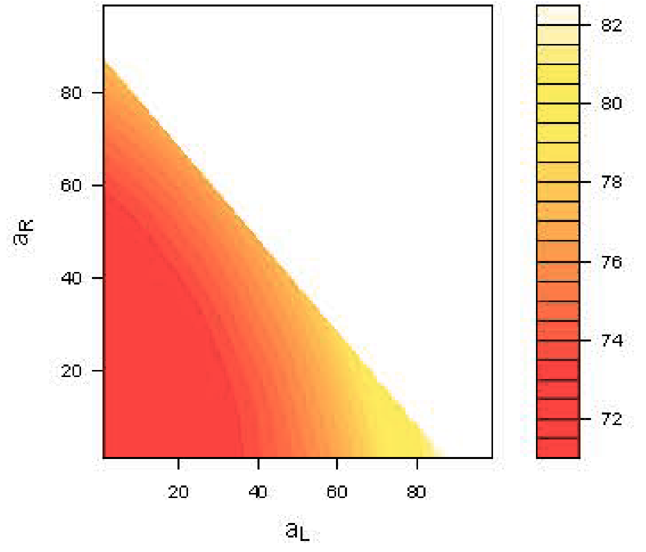


Model choice

- standardize data to common GP margins with the probability integral transform using the empirical distribution function
- fit the most general standardized model within each model class to the standardized data, using censored likelihood estimation
- select the standardized model with smallest AIC
- fit the full final model including marginal parameters to the original (non-standardized) data, using censored likelihood estimation
- LR-test for simplifications in the marginal parameterization.

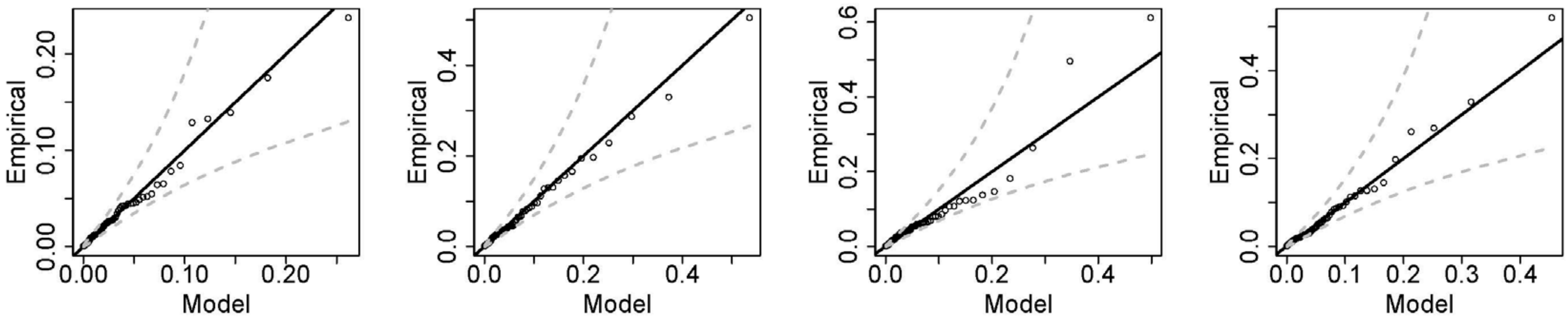
Result

- The standardized T-model with independent Gumbel components for $\boldsymbol{T}$ had smallest AIC
- Fit of this model to non-standardized data and LR-testing for dimension reduction led to a final 5-parameter model
- Computation of portfolio VaR simplified by (unexpected?) probabilistic result: sums of multivariate GP components conditional on being positive are 1-dimensional GP

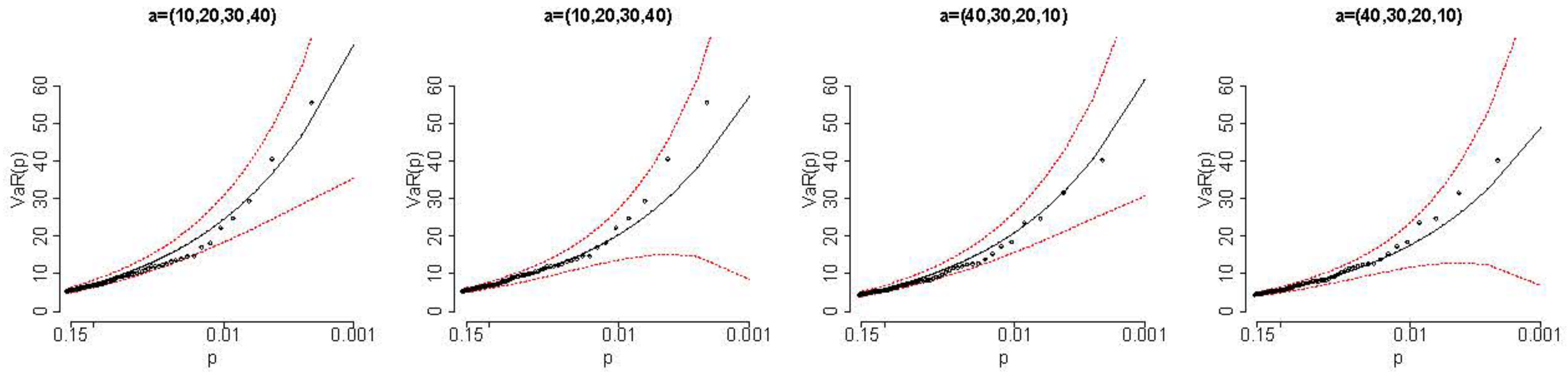


Model diagnostics

- χ-statistics also model diagnostics
- Marginal qq-plots



- VaR plots (=qq-plots) for different portfolios

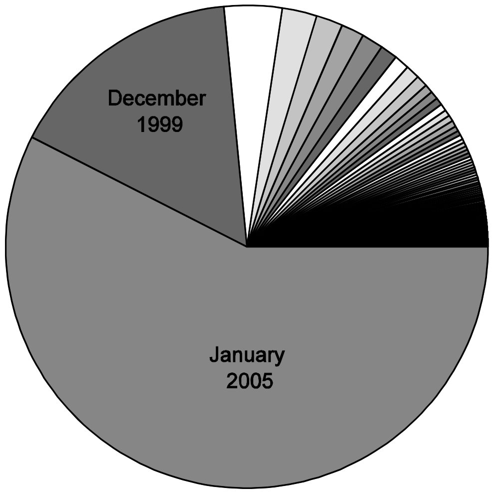


1-dim estimates 4-dim estimates

Before Gudrun

After Gudrun

Windstorm Gudrun, January 2005



SEK 2,8 billion loss (LF)
55% of total loss 1982-2005

Remember: analysis based on data 1982-1993: 1% chance biggest damage next 15-year period will be more than SEK 2,5 billion

???

Bivariate PoT analysis basered on data 1982-2005: 10% chance biggest damage next 15-year period will be more than SEK 7 billion (this time we didn’t miss damage to forrest – but did we miss something else?)

Take home

- Simultaneous occurrence of extremes important
- Intense development PoT and block maxima models for multivariate extremes
- Multivariate PoT even more recent than multivariate block maxima, but often most useful
- Likelihood inference gives possibility for flexible inclusion of covariates and trends
- Well understood how to include trends in marginal distributions. How to do it for dependence parameters not investigated (yet)
- Important computational challenges

A. Kiriliouk, H. Rootzén, J. Segers and J..L. Wadsworth: Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics, online* (2018)

H. Rootzén, J. Segers, J.L. Wadsworth: Multivariate peaks over thresholds models. *Extremes 21*, 115–145 (2018)

H. Rootzén, J. Segers, J.L. Wadsworth: Multivariate generalized Pareto distributions: parametrizations, representations, and properties *J. Multivariate Anal.* (2018)

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

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

Book on practical use of EVS

S. Coles: *An Introduction to Statistical Modeling of Extreme Values*. Springer (2001)

More theoretical books

P. Embrechts, C. Klüppelberg, T. Mikosch: *Modelling Extremal Events: for Insurance and Finance.* Springer, 2nd ed. (2013)

S. Resnick: Extreme values, regular variation, and point processes. Springer, 2nd ed. (2013)

R. Leadbetter, G. Lindgren, H. Rootzen: *Extremes and related properties of stationary sequences and processes.* Springer, 2nd ed. (2012)

J.-M. Azaïs and M. Wschebor: *Level Sets and Extrema of Random Processes and Fields.* Wiley (2009)

L. De Haan, A. Ferreira: *Extreme Value Theory: An Introduction.* Springer (2007)

J. Beirlant, Y. Goegebeur, J. Segers, J. L. Teugels*: Statistics of Extremes: Theory and Applications.* Wiley (2004)

J.-M. Azaïs and M. Wschebor: *Level Sets and Extrema of Random Processes and Fields.* Wiley (2009)

Recent paper on threshold selection

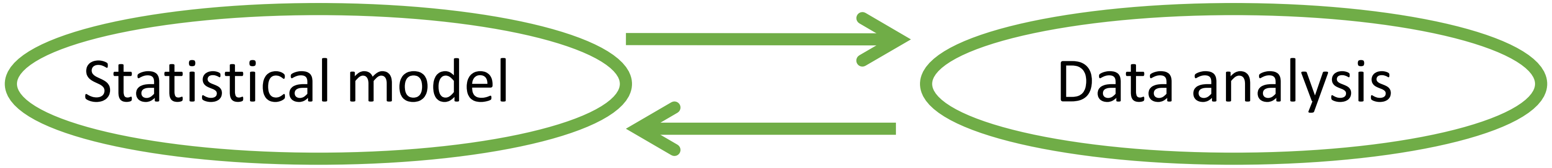
J. Wadsworth: Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics 58,* 116-126 (2016)

Recent paper on declustering

B Berghaus, A Bücher: Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist.* 2307-2335 (2018)

Model checking and adjustment

- Independence
- Stationarity
- Distributional fit
- Period and threshold choice



- Choice of distributions
- Estimation (often of quantiles)
- Confidence intervals
- “Prediction intervals”
- Hypothesis testing
- Regression and dependence modeling
- Prediction
- ***Understanding!!***