Generalized Additive Modelling for Sample Extremes

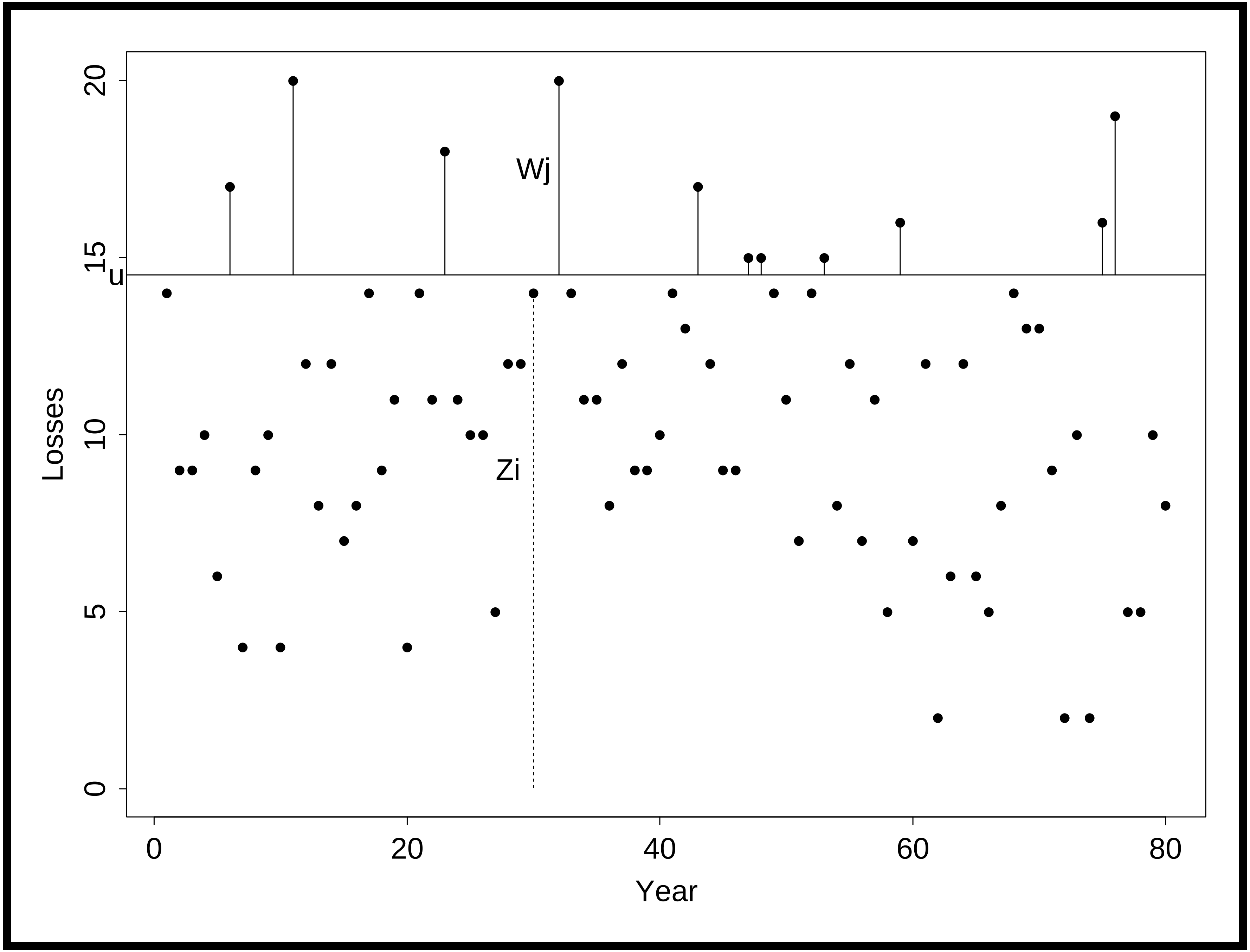
V. Chavez-Demoulin

ETHZ, Zürich

valerie@math.ethz.ch

Eurandom, 18th September 2003

- EVT: useful toolkit for describing non-standard price fluctuations
- Solution to the “heavy-tailed” reality in pricing, hedging, insurance regulation
- Aim to combine the EVT point process approach to exceedances with smoothing methods to give a flexible exploratory approach to modelling changes in extremes



Losses
Year
Wj
Zi
u
0
5
10
15
20
0
20
40
60
80

- $Z_1, \ldots, Z_q$ the ground up losses
- u (typically high) threshold
- n number of exceedances by $Z_1, \ldots, Z_q$ of u
- $W_1, \ldots, W_n$ excesses (loss exceeding u minus u)

- u attachment point, lower level of an excess-of-loss reinsurance treaty,. . .
- $W_1, \ldots, W_n$ larger operational losses above a threshold u

- Inhomogeneous Poisson process with intensity λ for the number of exceedances combined with independent excesses W over the threshold
- Given u, the excesses are treated as a random sample from the GPD with scale parameter σ and shape parameter κ

Since each exceedance is associated with a specific event, it is possible to let the scale and shape parameters depend on covariates. For instance, insurance losses can be of different types, Operational losses will typically belong to various subclasses (fraud, system failures, backoffice errors, . . .) and their occurrence no doubt shows a non–constant (often stochastic) intensity, possibly depending on such factors as business cycles, transaction intensity etc. Large losses may become more or less frequent over time, or indeed they may become more or less severe. It is also well–known that in general, insurance and financial losses show cyclic behaviour.

- The natural variability of the exceedances tends to mask any trends or other dependence on time
- Variation due to the different covariates such as type of claims, of losses could be summarized parametrically
- Changes in time need not have a specific parametric form

Our approach combines the point process for exceedances with smoothing methods to give a flexible exploratory framework to model changes in large values for insurance or financial data.

Contents

- Brief description of the threshold method
- Implementation of spline smoothers
- Bootstrap uncertainty assessment
- Application
- Discussion

Threshold method

- Treat occurrences of events over (or under) threshold u as Poisson process
- Number of exceedances N over u follows homogeneous Poisson process, rate λ
- Exceedance sizes $W_j = Z_j - u$ are random sample from GPD

$$G(w) = \begin{cases} 1 - (1 - \kappa w/\sigma)_+^{1/\kappa} & \text{if} \quad \kappa \neq 0 \\ 1 - \exp(-w/\sigma) & \text{if} \quad \kappa = 0 \end{cases}$$

- Use orthogonal parametrization $\kappa, \nu = \sigma(1+\kappa)$ (to avoid undesirable convergence problems)
- Log likelihood for data splits into two parts:

$$l(\lambda, \kappa, \nu) = l_N(\lambda) + l_W(\kappa, \nu)$$

Semiparametric model

- Generalize previous approach
- Take λ to be time-varying, where

$$\lambda = \exp\left\{x^T\alpha + f(t)\right\}$$

- Take exceedances to be GPD with

$$\kappa = x^T\beta + g(t), \quad \nu = \exp\left\{x^T\eta + s(t)\right\}$$

- f, g and s are smooth functions of time t
- Penalize roughness of f, g and s through second derivatives

Penalized log likelihoods

- For rate of exceedances λ, maximize

$$l_N(\lambda) - \frac{1}{2}\gamma_\lambda \int f''(t)^2 dt,$$

equivalent to fitting standard generalized additive model

- For sizes of exceedances, maximize

$$l_W\left\{\kappa(\beta,g),\nu(\eta,s)\right\}-\frac{1}{2}\gamma_\kappa\int g''(t)^2dt-\frac{1}{2}\gamma_\nu\int s''(t)^2dt$$

If g, s are cubic splines, equivalent to maximizing

$$l_W\left\{\kappa(\beta,g),\nu(\eta,s)\right\}-\frac{1}{2}\gamma_\kappa g^TKg-\frac{1}{2}\gamma_\nu s^TKs$$

over β, η, g, s and leads to generalized ridge regression

- Parameters γ_λ, γ_κ and γ_ν control smoothness of f, g and s

Vector of covariate x can depend on time also, taking into account possible discontinuities in λ, κ and ν: model factors with different levels (due for example to a worldwide crisis, announcement of economic policy decisions)

Seasonal effects can be explored by nonparametric part of the model, by increasing the degrees of freedom

Methodology

- Choose forms for λ, κ and ν
- Choose smoothing parameters γ_λ etc. using AIC
- Use likelihood ratio statistics/AIC for model comparisons
- When model correct, residuals

$$R_j = -\hat{\kappa}_j^{-1} \log\left\{1 - \hat{\kappa}_j W_j (1 - \hat{\kappa}_j) \hat{\nu}_j\right\}$$

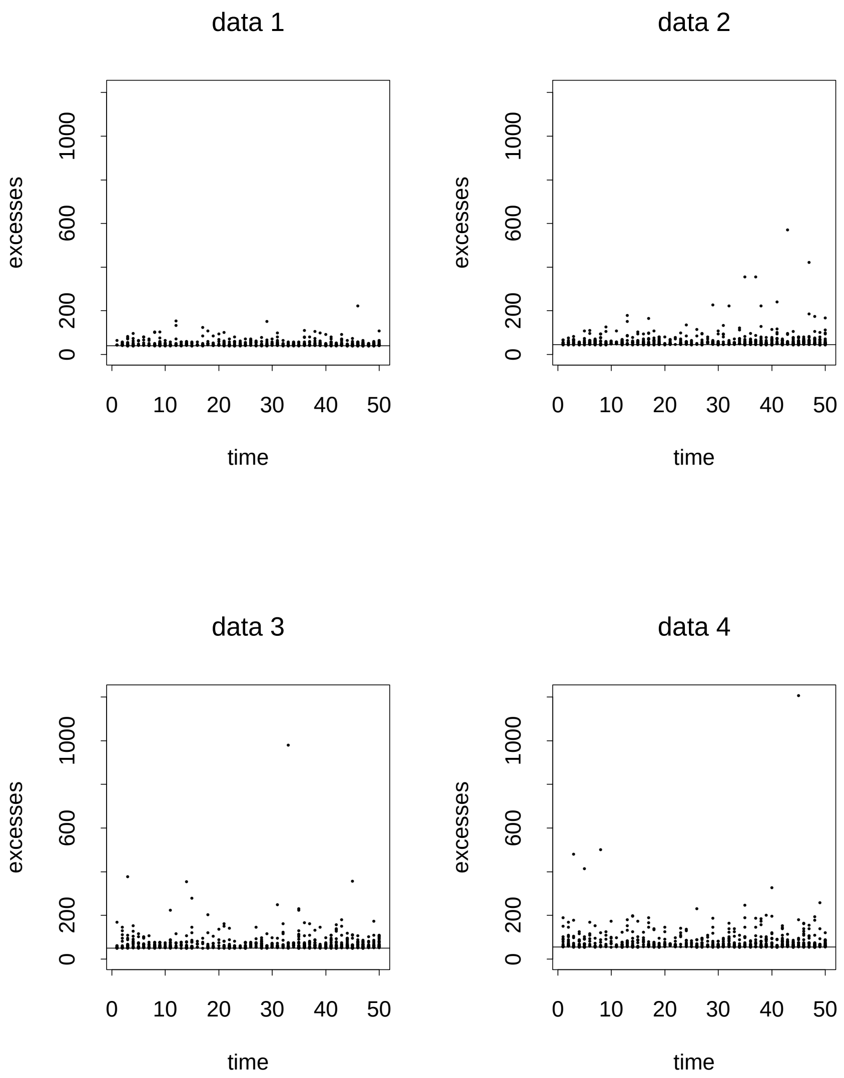
are approximately independent unit exponential variables

Bootstrap uncertainty assessment

- Need model-robust assessment of uncertainty
- Use bootstrap, either resampling the R_j
  - computed from undersmoothed curves
  - added to oversmoothed curves
- Either yields percentile confidence intervals/pointwise bands

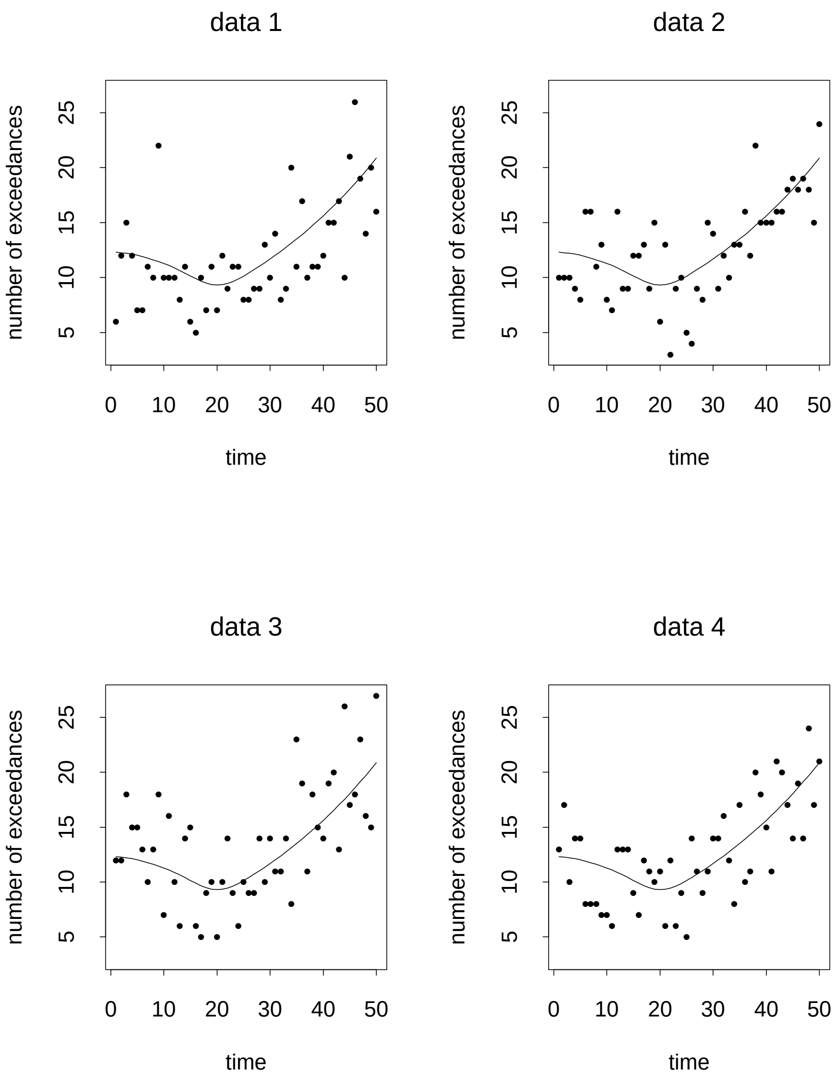
Simulated data

- How the methodology can be turned into a real analysis
- $1/p$ year return level: $y_{1-p} = u - \frac{\sigma}{\kappa}\left\{(\lambda/p)^{-\kappa} - 1\right\}$ changing in time?



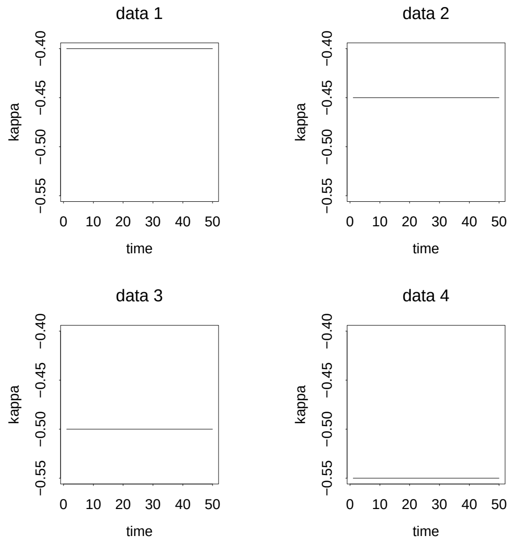
Simulated datasets of excesses against time

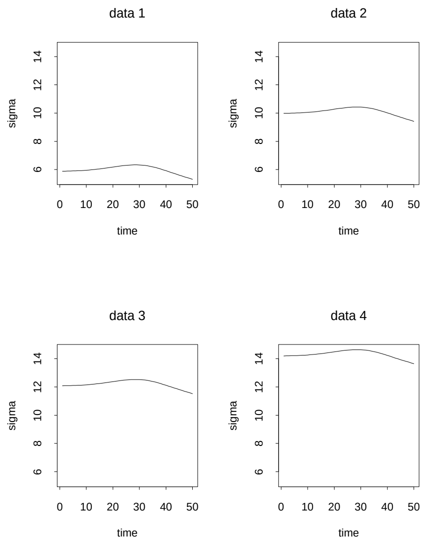
- Could be operational risk losses of $k = 4$ types over 50 years
- Each recorded as an excess above some threshold value u_i, $i = 1, \ldots, k$
- Is y_{1-p} constant or changing in time? If so how does y_{1-p} change with time?



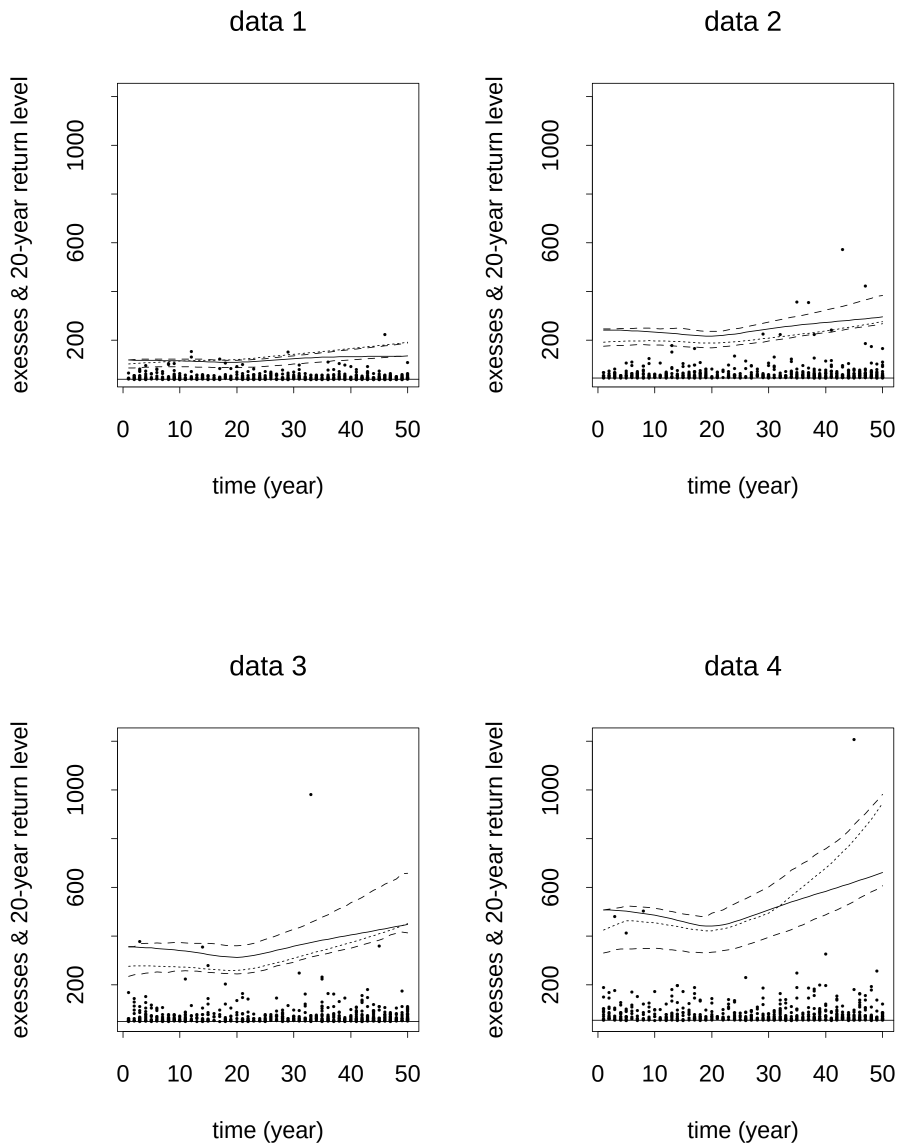
Numbers of exceedances simulated from the Poisson process of intensity shown by the solid line

- $\lambda(t) = f(t)$ nonparametric form
- given these numbers of exceedances, their sizes are simulated from a GPD with
  - $\kappa(t) = x^T\beta$ where x is any explanatory var. which differs for each dataset (heavy-tailed losses as $\kappa < 0$)
  - $\sigma(t) = \exp\{x^t\eta + s(t)\}$ semi-parametric form


data 1
data 2
data 3
data 4
kappa
time
−0.40
−0.45
−0.50
−0.55
0
10
20
30
40
50


data 1
data 2
data 3
data 4
sigma
time
14
12
10
8
6
0
10
20
30
40
50

- Different models for λ, κ and ν are compared using tests based on the likelihood ratio statistics
- Estimation of the 20-year return level $y_{1-0.05}$



data 1
data 2
data 3
data 4
exesses & 20-year return level
200
600
1000
0
10
20
30
40
50
time (year)

- Keep in mind that this is a simulated example
- Across the four rating classes there is a clear difference in quantile curves
- Estimated curves catch some of the main features of this nonstationarity

Discussion

- Exceedances over/under thresholds
  - widely-used approach with natural interpretation
  - exceedance times modelled using existing code (GAM)
  - fitting for exceedance sizes more involved but quite feasible

- Smoothing extremes by penalized log likelihood
  - convenient and rapid exploration technique
  - highlights features of underlying distribution
- Also applies to maximum and r-largest approaches

References

- Chavez-Demoulin, V. and Davison, A. C. (?2004) *Applied Statistics*
- Chavez-Demoulin, V. and Embrechts P. (2004) *to appear in in the Journal of Risks and Insurance*
- Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values* Springer
- Davison, A. C. and Ramesh, N. I. (2000) *JRSS,B*
- Davison, A. C. and Smith, R. L. (1990) *JRSS,B*

- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance* Springer
- Green, P. J., Silverman, B. W. (1994) *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach* C&H
- Pauli, F. and Coles, S. G. (?2001) *JRSS,B*