Risk Analytics

Extreme Value Theory
Block Maxima Approach

Frédéric Aviolat & Juraj Bodik

2025

Say you have observed outcomes $X_1, \ldots, X_n$ where higher values are worse, e.g.:

- Extreme temperatures (heatwaves, droughts, wild fires)
- Storms, extreme winds (hail)
- Heavy rainfalls, extreme discharges (floods)

Extreme value theory:

- How bad does it get?
- When should we ring the alarm? (see Challenger example)

How bad does it get?

"Study the worst case"

Two classical approaches:

- Block maxima: $\max(X_1, \ldots, X_n)$
- Peaks-over-Threshold: $\{X : X > u\}$ (for a high threshold u)

Study of Maxima

Let $X_1, X_2, \ldots$ be iid random variables with distribution function (df) F,

$$F(x) = \mathbb{P}(X \leq x).$$

Let $M_n = \max(X_1, \ldots, X_n)$ be worst–case value among $1, \ldots, n$.

$$\begin{aligned}
\mathbb{P}(M_n \leq x) &= \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) \\
&= \mathbb{P}(X_1 \leq x)\,\mathbb{P}(X_2 \leq x) \cdots \mathbb{P}(X_n \leq x) \\
&= F^n(x)\,.
\end{aligned}$$

It can be shown that, almost surely, $M_n \overset{n\to\infty}{\longrightarrow} x_F$, where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty$ is the right endpoint of F.

$\Rightarrow$ But what about normalized maxima?

Limiting Behaviour of Sums or Averages

We are familiar with the central limit theorem:

Let $X_1, X_2, \ldots$ be iid with finite mean μ and finite variance σ^2. Let $S_n = X_1 + X_2 + \ldots + X_n$. Then

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le x\right) \stackrel{n\to\infty}{\longrightarrow} \Phi(x)\,,$$

where Φ is the distribution function of the standard normal distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du\,.$$

The distribution of the normalized sum of the X_i converges to the standard normal distribution.

Limiting Behaviour of Sample Extrema

Let $X_1, X_2, \ldots$ be iid from F and let $M_n = \max(X_1, \ldots, X_n)$.

Suppose we can find sequences of real numbers $a_n > 0$ and b_n such that $(M_n - b_n)/a_n$, the sequence of normalized maxima, converges in distribution, i.e.

$$\begin{aligned}\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) &= \mathbb{P}(M_n \leq a_n x + b_n) \\ &= F^n(a_n x + b_n) \\ &\overset{n\to\infty}{\longrightarrow} H(x),\end{aligned}$$

for some non-degenerate distribution function $H(x)$.

We say that F is in the maximum domain of attraction of H, or $F \in \mathrm{MDA}(H)$.

In fact, we can find the form of $H(x)$ (Fisher-Tippett theorem).

Generalized Extreme Value Distribution

The GEV has distribution function

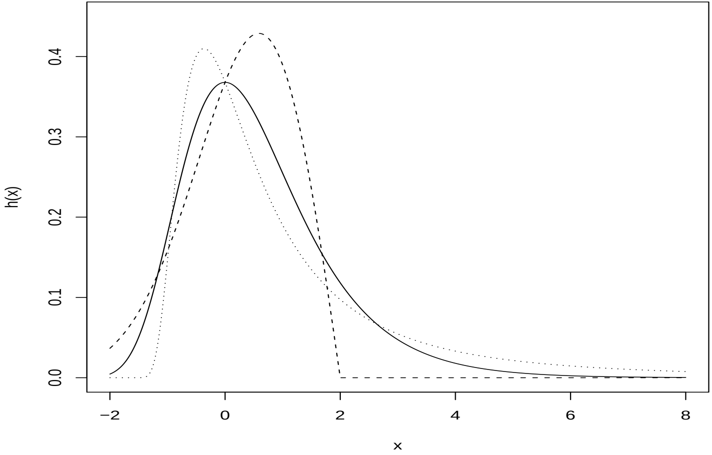
$$H_{\mu,\sigma,\xi}(x) = \begin{cases} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right) & \text{if } \xi \neq 0\,, \\ \exp\left(-e^{-(x-\mu)/\sigma}\right) & \text{if } \xi = 0, \end{cases}$$

where $1 + \xi(x-\mu)/\sigma > 0$.

- μ is the location parameter.
- σ is the scale parameter.
- ξ is the shape parameter.

$$\text{For } \left\{ \begin{array}{c} \xi > 0 \\ \xi = 0 \\ \xi < 0 \end{array} \right\} \text{ we also say that H is } \left\{ \begin{array}{c} \text{Fréchet} \\ \text{Gumbel} \\ \text{Weibull} \end{array} \right\}.$$

GEV Densities



Solid line corresponds to $\xi = 0$ (Gumbel); dotted line is $\xi = 0.5$ (Fréchet); dashed line is $\xi = -0.5$ (Weibull). $\mu = 0$ and $\sigma = 1$.

Fisher–Tippett Theorem (1928)

Theorem If $F \in \mathrm{MDA}(H)$ then H is of the type $H_{\mu,\sigma,\xi}$ for some μ, σ, ξ.

"If suitably normalized maxima converge in distribution to a non–degenerate limit, then the limit distribution must be an extreme value distribution."

Remark Essentially all commonly encountered continuous distributions are in the maximum domain of attraction of an extreme value distribution.

Fisher-Tippett: Examples

Recall: $F \in \mathrm{MDA}(H_\xi)$, iff there are sequences a_n and b_n with

$$\mathbb{P}\left\{(M_n - b_n)/a_n \leq x\right\} = F^n(a_n x + b_n) \stackrel{n\to\infty}{\longrightarrow} H_\xi(x).$$

We have the following examples:

- The exponential distribution,

$$F(x) = 1 - e^{-\lambda x}, \lambda > 0, x \geq 0$$

  is in $\mathrm{MDA}(H_0)$ (Gumbel). Take $a_n = 1/\lambda$, $b_n = (\log n)/\lambda$.

- The Pareto distribution,

$$F(x) = 1 - \left(\frac{\kappa}{\kappa + x}\right)^\alpha, \quad \alpha, \kappa > 0, \quad x \geq 0,$$

  is in $\mathrm{MDA}(H_{1/\alpha})$ (Fréchet). Take $a_n = \kappa n^{1/\alpha}/\alpha$, $b_n = \kappa n^{1/\alpha} - \kappa$.

Using Fisher–Tippett on data: Block Maxima Method

If you are given n values, use the limiting distribution to model the max M_n: first find the a_n and b_n from the theorem, and then

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx H_{0,1,\xi}(x)\,,$$

or $\mathbb{P}(M_n \leq y) = H_{b_n,a_n,\xi}(y)$.

+ All that's left is to estimate three parameters: ξ, b_n and a_n
− Need repeated values of $M_n \Rightarrow$ required data is a multiple of n

Remark

The values b_n and a_n are equivalent to the parameters μ and σ in the formula.

ML Inference for Maxima

We have block maxima data $\mathbf{y} = \left(M_n^{(1)}, \ldots, M_n^{(m)}\right)$ from m blocks of size n. We wish to estimate $\boldsymbol{\theta} = (\xi, \mu, \sigma)$. We construct a log–likelihood by assuming we have independent observations from a GEV with density $h_{\boldsymbol{\theta}}$,

$$l(\boldsymbol{\theta}; \mathbf{y}) = \log \left\{ \prod_{i=1}^{m} h_{\boldsymbol{\theta}} \left(M_n^{(i)}\right) \mathbb{1}_{\left\{1+\xi\left(M_n^{(i)}-\mu\right)/\sigma>0\right\}} \right\},$$

and maximize this w.r.t. $\boldsymbol{\theta}$ to obtain the MLE $\widehat{\boldsymbol{\theta}} = (\widehat{\xi}, \widehat{\mu}, \widehat{\sigma})$.

In defining blocks, bias and variance must be traded off:

- we reduce bias by increasing the block size n;
- we reduce variance by increasing the number of blocks m.

Return level and return period

Return level

Let X be a random variable with df F. The return level $R_\alpha = F^{-1}(\alpha)$ is the α-quantile, for probability α.

Return level of empirical data

Let $X_1 <= \ldots <= X_n$ be a series of ordered observations with df F. The return level $R_k = F^{-1}(1 - \frac{1}{k})$ is the value exceeded in $1/k$ observations, on average.

Return period

Let X be a random variable with df F. The return period $P(x) = \frac{1}{1-F(x)} = \frac{1}{\bar{F}(x)}$ is the expected number of observations to get a value greater than x.

Return period (of empirical data)

Let $X_1 <= \ldots <= X_n$ be a series of ordered observations with df F. One can check that $P(R_k) = k, \forall k$ and $R_{P(x)} = x, \forall x$.

Return level for GEV

- The return level for probability α of $H_{\mu,\sigma,\xi}$ is

$$R_\alpha = H^{-1}_{\mu,\sigma,\xi}(\alpha) = \begin{cases} \mu - \frac{\sigma}{\xi}\left(1 - (-\log(\alpha))^{-\xi}\right) & \text{if } \xi \neq 0 \\ \mu - \sigma \log(-\log(\alpha)) & \text{if } \xi = 0 \end{cases}$$

- For a return period $T = 1/p$, take $\alpha = 1 - p$. The return level is the value expected to be exceeded on average once every T periods.
- More precisely, R_{1-p} is exceeded by the maximum in any particular period with probability p.
- For example, if we have annual maxima, R_{1-p} is the value expected to be exceeded on average once every T years, and it is exceeded by the annual maximum in any particular year with probability p.

Return level plot

- If we set $y_p = -\log(1-p)$,

$$R(y_p) = \begin{cases} \mu - \frac{\sigma}{\xi}\left(1 - y_p^{-\xi}\right) & \text{if } \xi \neq 0 \\ \mu - \sigma \log(y_p) & \text{if } \xi = 0 \end{cases}$$

The plot $\{(\log(y_p), R(y_p)) : 0 < p < 1\}$ is the return level plot of $H_{\mu,\sigma,\xi}$.

- **Remarks**
  - If $\xi = 0$, the plot is linear.
  - If $\xi < 0$, the plot is concave with $\lim_{p \to 0} R(y_p) = \mu - \sigma/\xi$.
  - If $\xi > 0$, the plot is convex and has no finite bound.

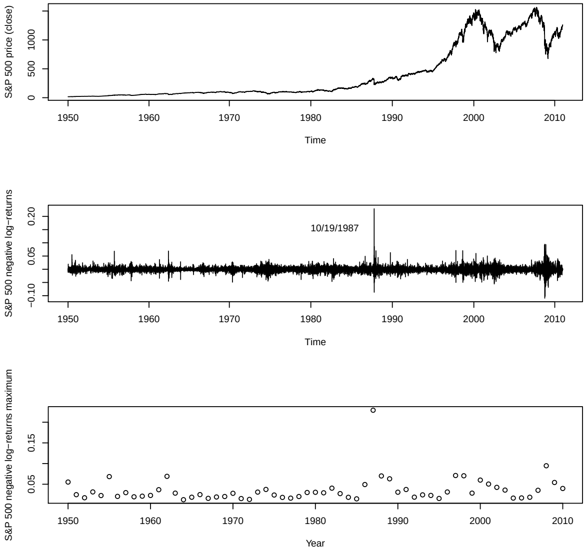
An Example: S&P 500

At our disposal are all daily closing values of the index since 1950. We analyse annual maxima of daily negative log-returns in the index. These values $M_{260}^{(1)}, \dots, M_{260}^{(61)}$ are assumed to be iid from $H_{\xi,\mu,\sigma}$.

Remark

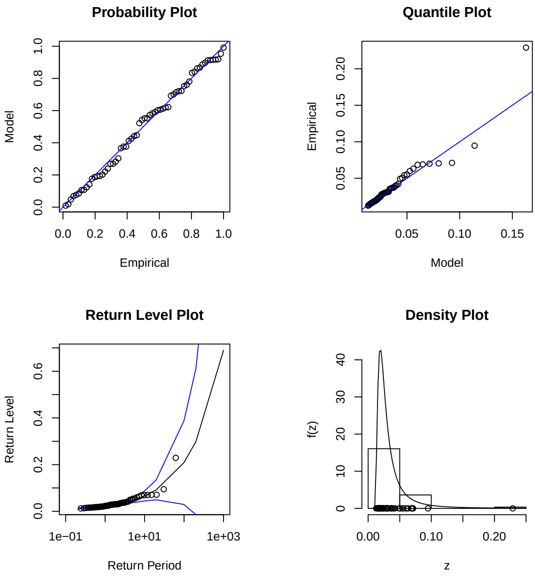
Although we have only justified this choice of limiting distribution for maxima of iid data, it turns out that the GEV is also the correct limit for maxima of stationary time series, under some technical conditions on the nature of the dependence. These conditions are fulfilled, for example, by GARCH processes.

S&P 500

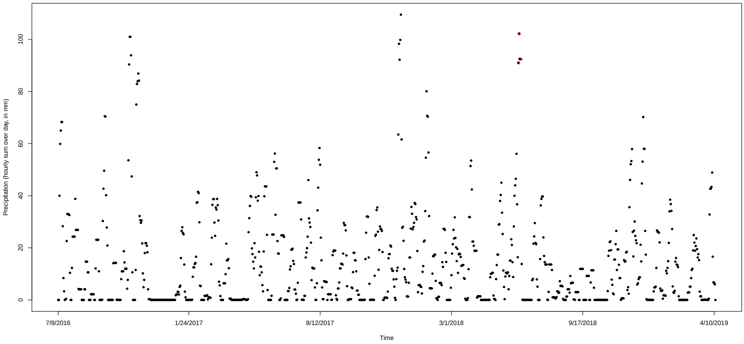


S&P 500 price (close)
0
500
1000
1950
1960
1970
1980
1990
2000
2010
Time
S&P 500 negative log−returns
−0.10
0.05
0.20
10/19/1987
1950
1960
1970
1980
1990
2000
2010
Time
S&P 500 negative log−returns maximum
0.05
0.15
1950
1960
1970
1980
1990
2000
2010
Year

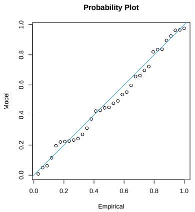
S&P 500 diagnostic graphs

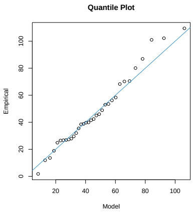


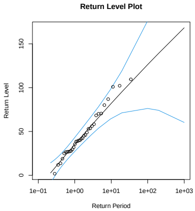
Heavy rainfall in Lausanne

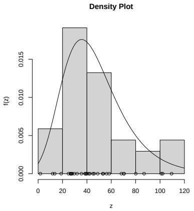


Heavy rainfall in Lausanne









Risk Analytics

Extreme Value Theory
Peaks-over-Threshold Method

Frédéric Aviolat & Juraj Bodik

2025

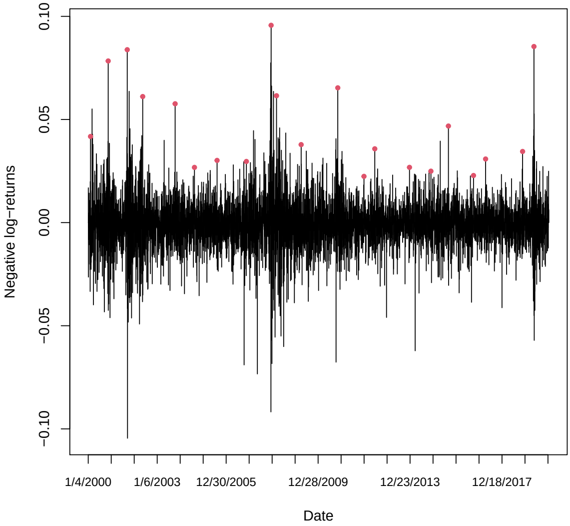
Previous Approach

- We saw how to analyse extreme values by their maxima.
- But this replaced a whole block of n values by a single one.
  - Lots of data loss, since n has to be “large enough”
  - Ignores behaviour inside a block (iid)

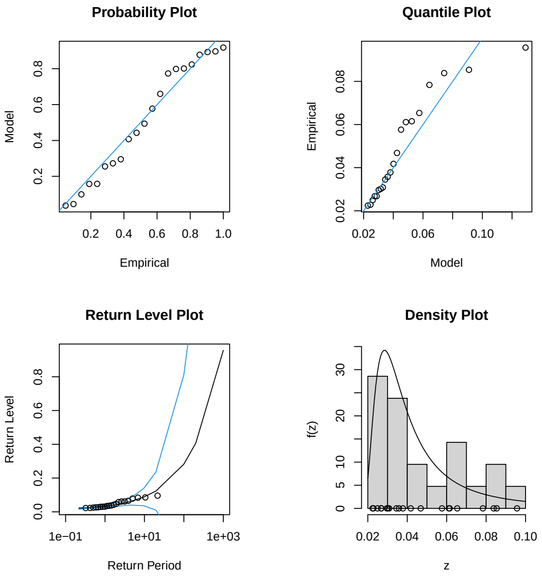
Threshold Methods

- Idea: instead of looking only at the highest, use all the values which are "high enough":
- The modelled values are $\{X_i : X_i > u\}$ for a high threshold u.

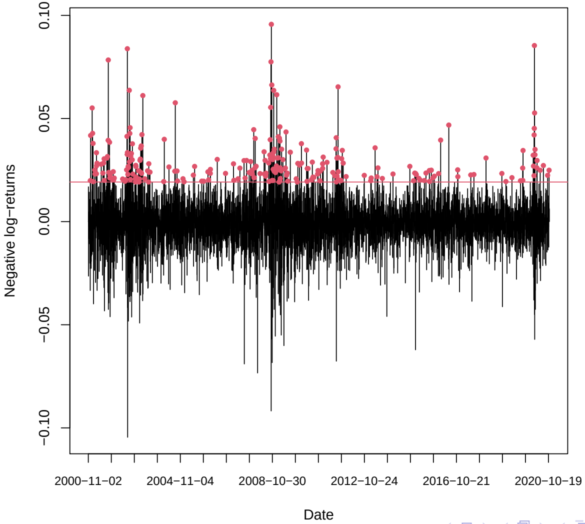
Nestlé negative log-returns (block maxima approach)



Nestlé negative log-returns (block maxima approach II)



Nestlé negative log-returns (Peaks-Over-Threshold approach)



Generalized Pareto Distribution

- The GPD is a two parameter distribution with df

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1+\xi x/\beta)_+^{-1/\xi} & \xi \neq 0\,, \\ 1 - \exp(-x/\beta) & \xi = 0\,, \end{cases}$$

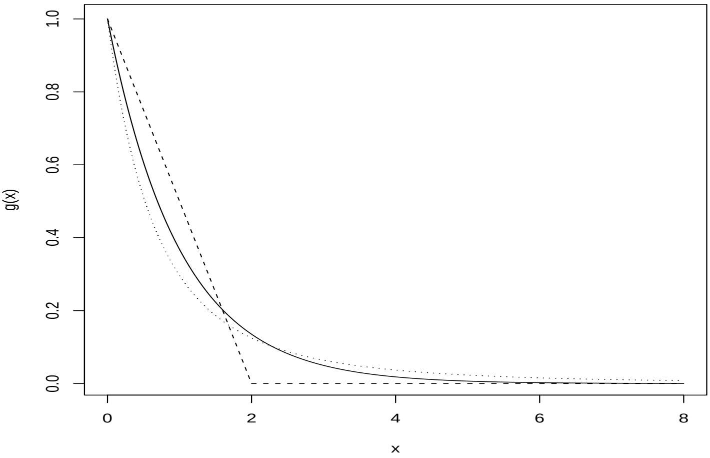
  where $\beta > 0$ and $a_+ = \max(a, 0)$.
  The support of $G_{\xi,\beta}$ is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

- Particular cases:

$$\left\{ \begin{array}{l} \xi > 0 \\ \xi = 0 \\ \xi < 0 \end{array} \right\} \text{ is equivalent to } \left\{ \begin{array}{l} \text{Pareto} \\ \text{Exponential} \\ \text{Pareto type II} \end{array} \right\}.$$

- **Moments.** For $\xi > 0$, the distribution is heavy tailed.
  $\mathbb{E}\left(X^k\right)$ does not exist for $k \geq 1/\xi$.

GPD Densities



Solid line corresponds to $\xi = 0$ (exponential); dotted line is $\xi = 0.5$ (Pareto); dashed line is $\xi = -0.5$ (Pareto type II). $\beta = 1$.

Peaks-over-threshold method

- **The excess distribution:** Given that a loss exceeds a high threshold, by how much can the threshold be exceeded?
- Let u be the high threshold and define the excess distribution above the threshold u to have the df

$$F_u(x) = \mathbb{P}(X - u \leq x \mid X > u) = \frac{F(x+u) - F(u)}{1 - F(u)},$$

for $0 \leq x < x_F - u$ where $x_F \leq \infty$ is the right endpoint of F.
- Extreme value theory suggests the GPD is a natural approximation for this distribution.

Examples

1. **Exponential.** $F(x) = 1 - e^{-\lambda x}$, $\lambda > 0$, $x \geq 0$. We find that

   $$F_u(x) = F(x)\,, \quad x \geq 0\,.$$

   The "lack–of–memory" property.

2. **GPD.** $F(x) = G_{\xi,\beta}(x)$. This time

   $$F_u(x) = G_{\xi,\beta+\xi u}(x)\,,$$

   where $0 \leq x < \infty$ if $\xi \geq 0$ and $0 \leq x < -\beta/\xi - u$ if $\xi < 0$. The excess distribution of a GPD remains a GPD with the same shape parameter; only the scaling changes.

What about the general case?

Asymptotics of Excess Distribution

Theorem. (Pickands–Balkema–de Haan (1974/75)) We can find a function $\beta(u)$ such that

$$\lim_{u \to x_F} \sup_{0 \leq x < x_F - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0 \,,$$

if and only if $F \in \mathrm{MDA}\,(H_\xi)$, $\xi \in \mathbb{R}$.

Essentially all the common continuous distributions used in risk management or insurance mathematics are in $\mathrm{MDA}\,(H_\xi)$ for some value of ξ.

Using Pickands–Balkema–de Haan on data

- "For a wide class of distributions, the distribution of the excesses over high thresholds can be approximated by the GPD."
- This result suggests we choose u high and assume the limit result is more or less exact

$$F_u(x) \approx G_{\xi,\beta}(x)\,,$$

  for some ξ and β.
- To estimate these parameters we fit the GPD to the excess amounts over the threshold u, i.e. to $\{X_i - u \,|\, X_i > u\}$.
- Standard properties of maximum likelihood estimators apply if $\xi > -0.5$
- To implement the POT method we must choose a suitable threshold u, using data–analytic tools (e.g. mean excess plot) to help.

When does $F \in \text{MDA}(H_\xi)$ hold?

- Fréchet case: $F \in \text{MDA}(H_{\xi>0})$
  If the tail of the df F decays like a power function then $F \in \text{MDA}(H_{\xi>0})$. Heavy-tailed distributions such as Pareto, Burr, log-gamma, Cauchy and t–distributions as well as various mixture models. Not all moments are finite.
- Gumbel Case: $F \in \text{MDA}(H_0)$
  Essentially it contains distributions whose tails decay roughly exponentially and we call these distributions light–tailed. All moments exist for distributions in the Gumbel class. Examples are the normal, log-normal, exponential and gamma.
- Weibull Case: $F \in \text{MDA}(H_{\xi<0})$
  If the tail of the df F is bounded above then $F \in \text{MDA}(H_{\xi<0})$. Examples are the uniform, beta.

The Mean Excess Plot

The mean excess function of a random variable X is

$$e(u) = \mathbb{E}(X - u \mid X > u).$$

It is the mean of the excess distribution function $F_u(x)$ above the threshold u, expressed as a function of u.

GPD Model:
Excess losses over threshold u are exactly GPD with $\xi < 1$, i.e., $X - u \mid X > u \sim \mathrm{GPD}(\xi, \beta)$. It is easily shown that for any higher threshold $v \geq u$

$$e(v) = \mathbb{E}(X - v \mid X > v) = \frac{\beta + \xi(v - u)}{1 - \xi},$$

so the mean excess function is linear in v.

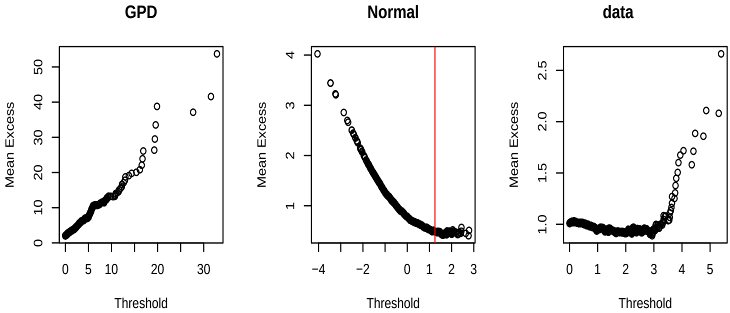
Choosing a threshold in practice

- Given some data, what is an appropriate threshold?
- The sample mean excess plot estimates $e(u)$ in the region where we have data:

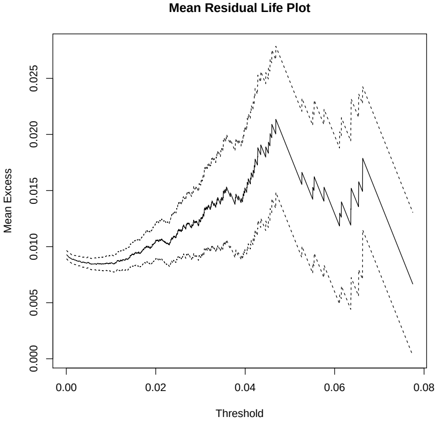
$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)_+}{\sum_{i=1}^{n} \mathbb{1}_{(X_i > u)}},$$

  We seek a threshold u, above which the plot is roughly linear.
- If we can find such a threshold, the result of theorem Pickands-Balkema-De Haan could be applied above that threshold.
- The plot is erratic for large u, when the averaging is over very few excesses. It is often better to omit these from the plot.
- **Bias-variance tradeoff:**
  - threshold too low $\Rightarrow$ bias because of the model asymptotics being invalid;
  - threshold too high $\Rightarrow$ variance is large due to few data points.

Examples of Mean Excess Plots

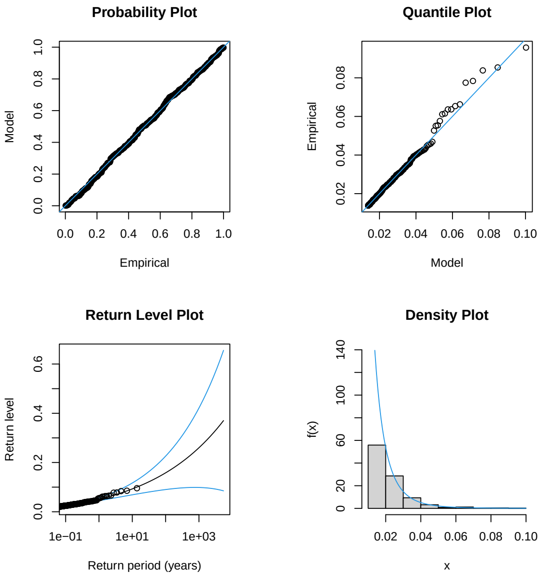


Nestlé negative log-returns: MRL-plot



Threshold could be put at 0.01? 0.04 looks maybe a bit high.

Nestlé negative log-returns: MRL-plot



Modelling Tails of Distributions

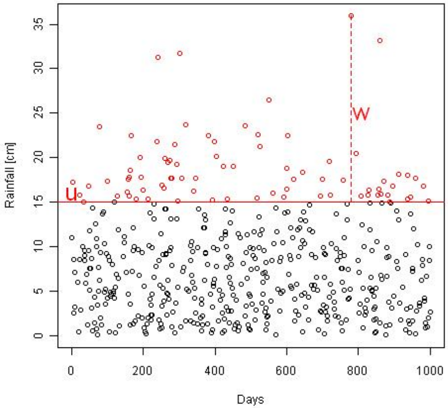
- Under our assumption that $F_u = G_{\xi,\beta}$ for some u, ξ, β we have, for $x \geq u$,

$$\begin{aligned}\bar{F}(x) &= \mathbb{P}(X > u)\mathbb{P}(X > x \mid X > u) \\ &= \bar{F}(u)\mathbb{P}(X - u > x - u \mid X > u) \\ &= \bar{F}(u)\bar{F}_u(x - u) \\ &= \bar{F}(u)\left(1 + \xi\frac{x - u}{\beta}\right)^{-1/\xi}\end{aligned}$$

  which if we know $\bar{F}(u)$, gives us a formula for estimating tail probabilities.

- This formula may be used to derive formulas for risk measures like VaR and expected shortfall (cf. Risk Measures Module).

Peaks-Over-Threshold approach in practice (I)



Peaks-Over-Threshold approach in practice (II)

Suppose $X_1, \ldots, X_n$ a sample of iid data coming from the distribution F. Considering exceedances above some high threshold u, we have that

- The number of exceedances N_u above the high threshold u follows a Poisson distribution with parameter λ, and, independently,
- the sizes $W_i = X_i - u$, $i = 1, \ldots, N_u$ of these N_u exceedances follow a GPD with parameters ξ and β.

In practice, we estimate the three parameters λ, ξ, β by maximizing the likelihood. Inference can be done separately for λ (from the Poisson likelihood) and for ξ, β (from the GPD likelihood).
The MLE of λ is $\hat{\lambda} = N_u/n$.

Dependence and clustering (I)

- Independence of widely separated extremes seems reasonable in most applications.
- But they almost always display short-range dependence in which clusters of extremes occur together (flow level maxima often occur during the strongest storm of the year).
- In these cases, it seems unrealistic to assume independence within each year.
- In the threshold method, the usual solution is to fit the point process model to the maxima of the clusters.

Dependence and clustering (II)

- An important practical problem is the identification of clusters from data.
- In this course both the choice of a suitable high threshold and of a method to identify independent clusters are based on a runs approach.
- Suppose that a series $X_1, \ldots, X_n$ has short range dependence, so extremes occur in clusters of mean size $1/\theta$, where $0 < \theta \leq 1$; θ is called the "extremal index".
- $\theta = 1$ corresponds to asymptotically independent clusters of size 1, whereas $\theta \approx 0$ means extremes tend to cluster.

Declustering method

For a fixed threshold u, we identify different groups of exceedances over u as independent clusters only if there are at least v consecutive observations under u between them. An estimate number of independent clusters for a sample of n observations is

$$C_n(u, v) = \sum_{j=1}^{n-v} Z_j(1 - Z_{j+1}) \cdots (1 - Z_{j+v}), \tag{1}$$

where $Z_j = \mathbb{1}_{(X_j > u)}$. The estimate of extremal index is then $\hat{\theta} = C_n(u, v)/N_u$, where N_u is the number of exceedances over u. Two important remaining points are the choices of u and v. In applications, we use sensible choices based on the specific situation.

Declustering method example

We consider the first-order process ARMAX(ρ) with parameter $0 \leq \rho \leq 1$ so that

$$Y_i = \max(\rho Y_{i-1}, \epsilon_i),$$

where $\{\epsilon_i\}$ are iid from F_ϵ with $F_\epsilon(x) = \exp\{-(1-\rho)/x\}$, for $x > 0$. It can be shown that $\theta = 1 - \rho$.

cf. in R, see doc of function `decluster` *in package* `extRemes`.

Comments

- POT modelling is a bit more flexible than block maxima, because the threshold can be moved continuously.
- If there is seasonality, a moving threshold will be required, and this requires separate modelling.
- In practice the benefit of including more high values in a threshold approach is limited: often one sees that the standard errors are about the same as when using block maxima.
- The real potential benefit of threshold methods is the possibility of more detailed modelling of temporal evolution of extremes, by looking at clusters of extremes, which are missed using the block maximum approach.

Risk Analytics

Extreme value theory: summary and practice

Frédéric Aviolat & Juraj Bodik

2025

Block maxima

Let $X_1, \ldots, X_n$ be a series of iid data from F and let $M_n = \max(X_1, \ldots, X_n)$.

If $F \in \mathrm{MDA}(H)$, then:

- H is a Generalized Extreme Value Distribution (GEV), with parameters $\mu, \sigma,$ and ξ.
- The maxima $M_n \stackrel{n\to\infty}{\longrightarrow} H_{\mu,\sigma,\xi}$, for some μ, σ, ξ.

Peaks-over-threshold

The POT (peaks-over-threshold) model is a limit model for threshold exceedances in iid processes. The limit is derived by considering datasets $X_1, \ldots, X_n$ and thresholds u_n that increase with n and letting $n \to \infty$. The limit model says that:

- Exceedances occur according to a homogenous Poisson process in time.
- Excess amounts above the threshold are iid and independent of exceedance times.
- Distribution of excesses is Generalised Pareto (GPD).

Stationary series

Let $X_1, \ldots, X_n$ be a stationary time series.

- **Block maxima** The maxima can usually be considered independant. So asymptotic results are still valid.
- **Peaks-over-threshold**
  - Asymptotic results are still valid.
  - But clustering usually happens.
  - Use declustering.
  - Take into account the extremal index θ.
  - Clusters have mean size $1/\theta$.

Non-stationary series

Let $X_1, \ldots, X_n$ be a non-stationary time series.

- **Block maxima** Use time-varying parameters:

$$\begin{aligned} \mu(t) &= g(t, \boldsymbol{\alpha}) \\ \sigma(t) &= h(t, \boldsymbol{\alpha}) \\ \xi(t) &= q(t, \boldsymbol{\alpha}) \end{aligned}$$

- **Peaks-over-threshold** Use time-varying parameters:

$$\begin{aligned} u(t) &= v(t, \boldsymbol{\alpha}) \\ \beta(t) &= b(t, \boldsymbol{\alpha}) \\ \xi(t) &= q(t, \boldsymbol{\alpha}) \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_K)$ is a vector of parameters.

Return level and return period

Let X be a random variable with df F.

Return period

The return period $P(x) = \frac{1}{1-F(x)} = \frac{1}{\bar{F}(x)}$ is the expected number of observations to get a value greater than x.

Return level

The return level $R_\alpha = F^{-1}(\alpha)$ is the α-quantile, for probability α.

The k-period return level $R_k = F^{-1}(1 - \frac{1}{k})$ is the $(1 - \frac{1}{k})$-quantile.

Value-at-Risk and Expected Shortfall

Let X be a random variable with df F.

Value-at-Risk

$\mathrm{VaR}_\alpha(X) = F^{-1}(\alpha)$ is the α-quantile, for probability α.

Expected Shortfall

$\mathrm{ES}_\alpha(X) = \mathbb{E}\left(X \mid X > \mathrm{VaR}_\alpha(X)\right)$ is the average value of X, when X is higher than $\mathrm{VaR}_\alpha(X)$.

Diagnostic plots

Let $X_1, \ldots, X_n$ be a series of observed data. Let E be the empirical distribution of this data, and $\hat{F}$ the estimated fitted distribution.

Probability plot $\left\{(E(X_i), \hat{F}(X_i)) \mid i = 1, \ldots, n\right\}$,

$$E(X_i) = \tfrac{i}{n+1}$$

Quantile plot $\left\{(E^{-1}(p), \hat{F}^{-1}(p)) \mid 0 < p < 1\right\}$,

$$E^{-1}(p) = X_{(i)}$$

Return level plot $\{(\log(y_p), R(y_p)) \mid 0 < p < 1\}$,

where $y_p = -\log(1-p)$ and $R(y_p) = \hat{F}^{-1}(y_p)$.

Packages in R

The following packages allow for extreme values analysis:

- `ismev`
- `extRemes`
- `POT`
- `evd`

See the documentation of these packages and their functions.

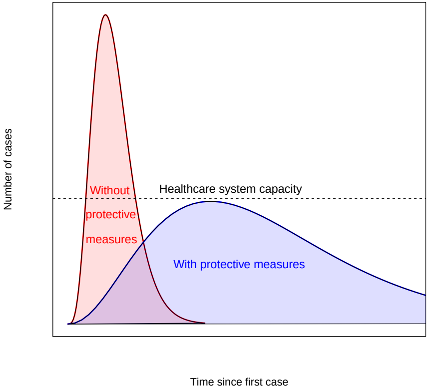
Risk Analytics

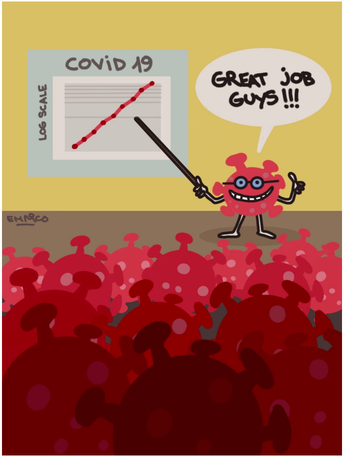
Introduction

Frédéric Aviolat & Juraj Bodik

2025

Risk Assessment Matters: Health Care Risk and Hospital Congestion





Risk Assessment Matters: Environmental Risk

Source: Cieau

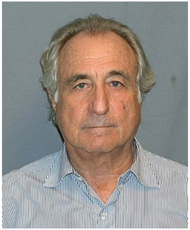
Risk Assessment Matters:
Environmental Risk → Enterprise/Insurance Risk

Source: 24 Heures; Rheinland-Pfalz government

Risk Assessment Matters: Cybersecurity/Terrorist Risk → Enterprise/Insurance Risk

Source: Wärtsilä; Business Insider

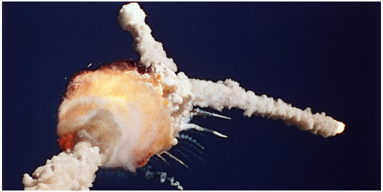
Risk Assessment Matters: Financial Risk



$$\Pr[T_A < 1, T_B < 1] = \Phi_2(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)), \gamma)$$

"Recipe for Disaster: The Formula That Killed Wall Street" Wired (2009)
"The World's Largest Hedge Fund is a Fraud" - 2005 (already)

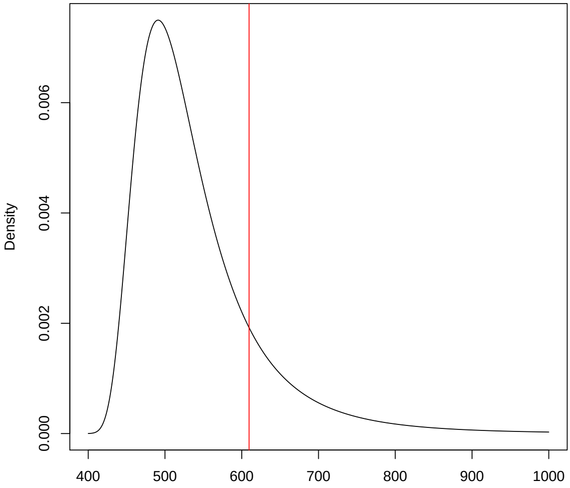
Risk Assessment Matters: Engineering/Technology Risk → Enterprise/Insurance Risk



Source: Welt; deepsense.ai

Analysis of extreme values

"The devil is in the tails"



Quantitative methods

Example: What is the distribution of the maximum?
(Fisher-Tippett Theorem - 1928)
The GEV has distribution function

$$H_{\mu,\sigma,\xi}(x) = \begin{cases} \exp\left(-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{-1/\xi}\right) & \text{if } \xi \neq 0\,, \\ \exp\left(-e^{-(x-\mu)/\sigma}\right) & \text{if } \xi = 0, \end{cases}$$

where $1+\xi(x-\mu)/\sigma > 0$ and ξ is the shape parameter. For

$$\left\{ \begin{array}{c} \xi > 0 \\ \xi = 0 \\ \xi < 0 \end{array} \right\} \text{ we also say that H is } \left\{ \begin{array}{c} \text{Fréchet} \\ \text{Gumbel} \\ \text{Weibull} \end{array} \right\}.$$

Course organisation

Week	Date	Lecture	Exercises
1	18.09	Workshop week (no RA lecture)	
2	25.09	Introduction + Time series	Practical 1 (TS)
3	02.10	Time series in practice	Practical 1 (TS)
4	09.10	EVT: Block Maxima + Practice (I)	Practical 1 (TS)
5	16.10	EVT: Peaks over threshold	Practical 2 (EVT)
6	23.10	Monitoring and reporting risk	Practical 2 (EVT)
7	30.10	Advanced topics in EVT	Practical 2 (EVT)
8	06.11	EVT in practice (II)	Practical 2 (EVT)
9	13.11	Q & A	Practical 3
10	20.11	Q & A	Practical 3
11	27.11	Q & A	Practical 3
12	04.12	Q & A	Practical 3
13	11.12	Q & A	Practical 3
14	18.12	Presentations	

Course organisation II

- Lectures and exercise sessions will be given in class. They will not be recorded.
- Practicals will be in groups of 4 to 5.
- Evaluation:
  - One report (for the 3 practicals together) per group, max 10 pages
  - Due **Monday 15.12 at 17.00**, with `R` code in appendix
  - Presentations of Practical 3 in the last week (15 min. speaking time + 5 min. questions)
  - Grade: report (1/2) + presentation (1/2)
  - No exam

Topics

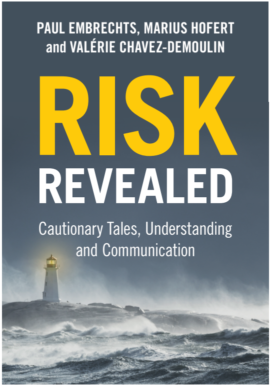
- Time series
  - ARMA processes
  - GARCH processes

- Extreme value theory
  - Block maxima
  - Peaks over threshold

- Risk measures
  - Return level
  - Value at risk
  - Expected shortfall

Vocabulary, appendices and references

In the upper section of the moodle page, you will find a Vocabulary and References file, in which we try to list the vocabulary used in the Course modules and in which we give further (non-mandatory) reading about each of the topic covered. Feel free to consult this file in case you are unsure about some prior notions of statistics and/or notations you encounter in the slides.



This course requires good notions of statistics and the R language.

Hope you enjoy this course !

Motivation

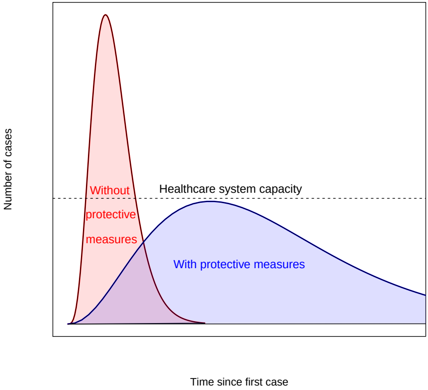
Why Risk Assessment matters?

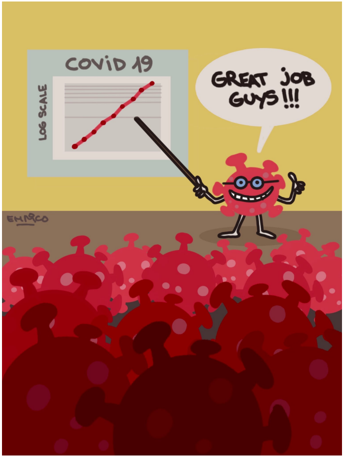
Risk Analytics

Frédéric Aviolat & Juraj Bodik

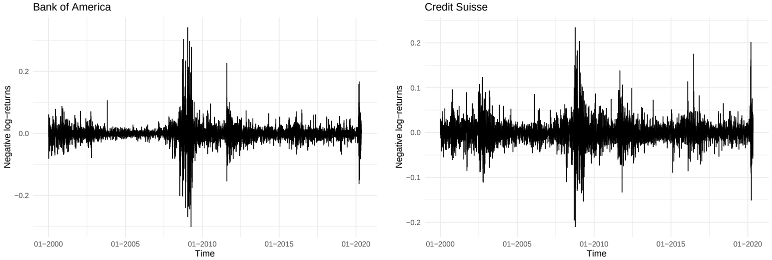
2025

Risk Assessment Matters: Health Care Risk and Hospital Congestion





Risk Assessment Matters: Financial Risk



Source: Yahoo Finance

Risk Assessment Matters: Environmental Risk

Source: Arolla.org(left) Cieau (right)

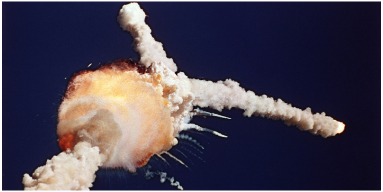
Risk Assessment Matters:
Environmental Risk $\rightarrow$ Enterprise/Insurance Risk

Source: 24 Heures (left) Good News Network (right)

Risk Assessment Matters:
Cybersecurity/Terrorist Risk → Enterprise/Insurance Risk

Source: Wärtsilä (top) Business Insider (bottom)

Risk Assessment Matters:
Engineering/Technology Risk $\rightarrow$ Enterprise/Insurance Risk



Source: Welt; deepsense.ai

Challenger Disaster Case I

On January 28, 1986, at 11:38 EST (16:38 UTC) at Kennedy Space Center, Cap Canaveral, the space shuttle Challenger takes off for the last time.
This 24^{th} launch could have been just a routine, but 73 seconds after the launch, at 14,000 meters above sea level, the shuttle suffered a tragic structural failure.

Source: earthsky.org (left) Britanica.com (right)

Challenger Disaster Case II

Source: Wikimedia Commons (left) and Wikipedia (right)

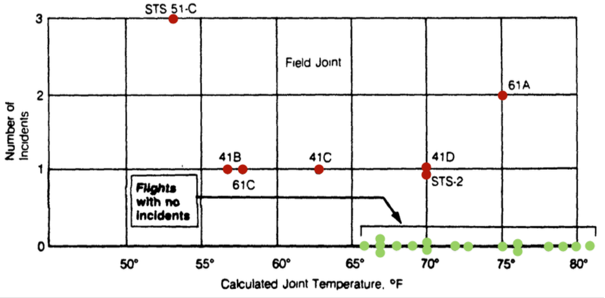
Challenger Disaster Case III: Technical and cultural/Management problems

That day, America lost for the first time, people in space.
A commission reported both technical and managerial or cultural causes:

- On the technical side, a failed o-ring seal in the right solid rocket booster caused by an exceptional cold weather (30.92° F) at launch time.
- On the cultural side, communication problems: failure to report all problems to the launch decision team. The problem of the effect of low temperature on the o-ring performance was known. However, anxious to keep their schedule, the NASA managers did not communicate this information to their superiors and decided to maintain the launch. Unfortunately.

Challenger Disaster Case IV: NASA Data Analysis

Number of o-ring failures against temperature:



Source: Report of the Presidential Commission on the Space Shuttle Challenger Accident, 6 June 1986, Volume 1, page 145, color added by https://priceonomics.com/the-space-shuttle-challenger-explosion-and-the-o/

NASA managers excluded the flights where no failures happened. This data selection mistake leads to "bias selection".

Challenger Disaster Case V: Logistic Regression

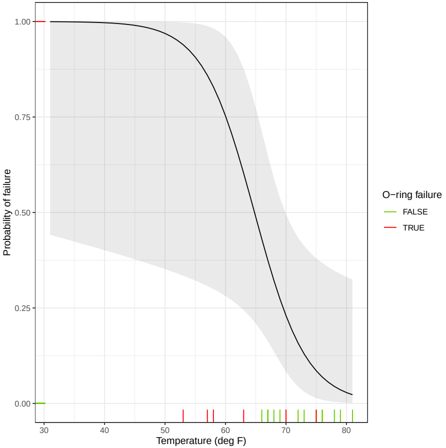
The information coming from the flights with no o-ring failure (especially happening at the lowest temperatures) should have been included !

- Logit model $\log\left(\frac{p(t)}{1-p(t)}\right) = \alpha + \beta t$ where t represents ground temperature at launch time and $p(t)$ represents the probability of O-ring failure at temperature t
- $\hat{\alpha} = 15.0429(7.3786)$ and $\hat{\beta} = -0.2322(0.1082)$
- Point estimate of a probability of failure given a certain temperature t:

$$p(t) = \frac{e^{15.0429-0.2322t}}{1 + e^{15.0429-0.2322t}}$$

- For $t = 31$, this value is bigger than 0.999 (and check uncertainty!)

Challenger Disaster Case VI: Logistic Regression Fit



Challenger Disaster Case VII: Societal Pressure

Despite the recommendation from the motor engineers, NASA ignored the risk of an O-ring failure.
The pressure to launch was the biggest enemy of this tragedy.

- Political pressure, shareholders pressure
- Public pressure: space discovery, “Teacher in Space” (Christa McAuliffe)

Along with the Space Shuttle Columbia accident on February 1, 2003, during the atmospheric re-entry phase, Challenger accident is one of the most significant in the American conquest of space.

Risk Analytics

Extreme Value Theory
More on EVT

Frédéric Aviolat & Juraj Bodik

2025

Modelling issues

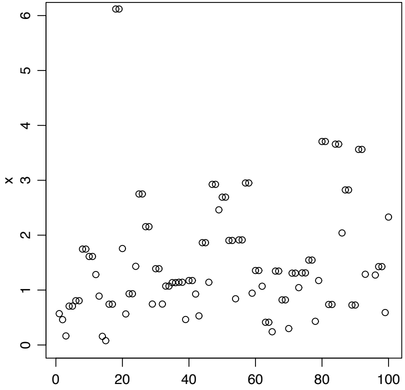
Extreme value data usually show:

- short term dependence (storms for example); clustering effect and extremal index;
- seasonality (due to annual cycles in meteorology);
- long–term trends (due to gradual climatic change);
- dependence on covariate effects;
- other forms of non–stationarity.

For temporal dependence there is a sufficiently wide-ranging theory which can be invoked. Other aspects have to be handled at the modelling stage.

Short-range dependence: example

Suppose $Y_1, Y_2, \ldots \overset{\text{iid}}{\sim} \exp(1)$, and $X_i = \max(Y_i, Y_{i+1})$:



Extremes tend to cluster in pairs.

More on Extremal index I

The previous example illustrates the following general result.

Theorem

Let (X_i) be a stationary process and (X_i^*) be independent variables with the same marginal distribution. Set $M_n = \max\{X_1, \ldots, X_n\}$ and $M_n^* = \max\{X_1^*, \ldots, X_n^*\}$. Under suitable regularity conditions,

$$\mathbb{P}\left((M_n^* - b_n)/a_n \leq z\right) \to H_1(z)$$

as $n \to \infty$ for normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$, where H_1 is a non-degenerate distribution function, if and only if

$$\mathbb{P}\left((M_n - b_n)/a_n \leq z\right) \to H_2(z),$$

where

$$H_2(z) = H_1^{\theta}(z)$$

for a constant θ called the extremal index that satisfies $0 < \theta \leq 1$.

Thus if H_1 is GEV, then so is H_2, with the same ξ.

This is a strong robustness result.

More on Extremal index II

- The extremal index can also be defined as

$$\theta = \lim_{n \to \infty} \mathbb{P}\left(\max(X_2, \ldots, X_{p_n}) \leq u_n \mid X_1 \geq u_n\right),$$

  where $p_n = o(n)$, and the sequence u_n is such that $\mathbb{P}(M_n \leq u_n)$ converges.
- Loosely, θ is the probability that a high threshold exceedance is the final element in a cluster of exceedances.
- Thus extremes occur in clusters whose (limiting) mean cluster size is $1/\theta$.

In fact the distribution of a cluster maximum is the same as the marginal distribution of an exceedance, so there is no bias in considering only cluster maxima, if we can identify clusters ...

Consequences

- When clustering occurs, the notion of return level is more complex:
  - if $\theta = 1$, then the '100-year-event' has probability $(\frac{99}{100})^{100} = 0.366$ of not appearing in the next 100 years;
  - if $\theta = 1/10$, then on average the event also occurs ten times in a millenium, but all together: it has probability $(\frac{99}{100})^{10} = 0.904$ of not appearing in the next 100 years.
- If we can estimate the tail of marginal distribution F (e.g. by fitting to block maxima), then

$$\mathbb{P}(M_n \le x) \approx F(x)^{n\theta} \approx H_{\mu,\sigma,\xi}(x),$$

  where H is GEV with parameters μ, σ, ξ. The marginal quantiles are approximately

$$F^{-1}(p) \approx H^{-1}(p^{n\theta}) > H^{-1}(p^{n}),$$

  so may be much larger than would be the case with $\theta = 1$.
- A similar argument shows that ignoring θ can lead to over-estimating a return level estimated using H.

Calculation of return levels

- The k-period return level is

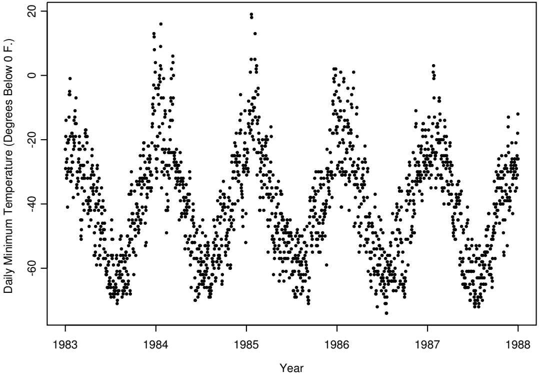
$$R_k = u + \frac{\beta}{\xi}\left((k\bar{F}(u)\theta)^{\xi} - 1\right),$$

  where β and ξ are the parameters of the threshold excess generalized Pareto distribution, $\bar{F}(u)$ is the probability of an exceedance of u, and θ is the extremal index.
- We can estimate

$$\hat{\bar{F}}(u) = \frac{n_u}{n} \text{ and } \hat{\theta} = \frac{n_c}{n_u}.$$

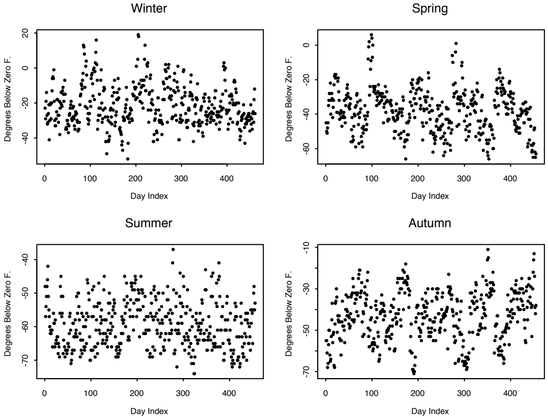
  where n_c is number of clusters and n_u is number of exceedances.
- So simply estimate the component $\bar{F}(u)\theta$ by n_c/n.

Example: Wooster temperatures



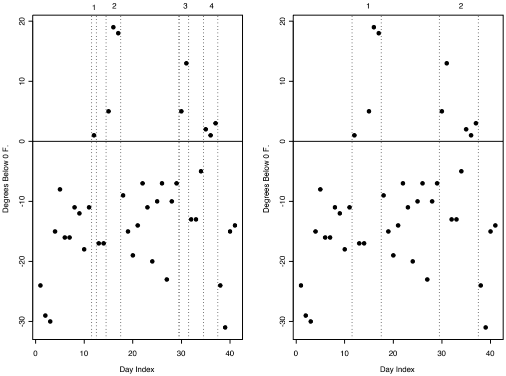
Daily minimum temperatures (degrees below 0°F) at Wooster.

Example: Wooster temperatures



Daily minimum temperatures (degrees F) at Wooster, split into seasons

Example: Wooster temperatures



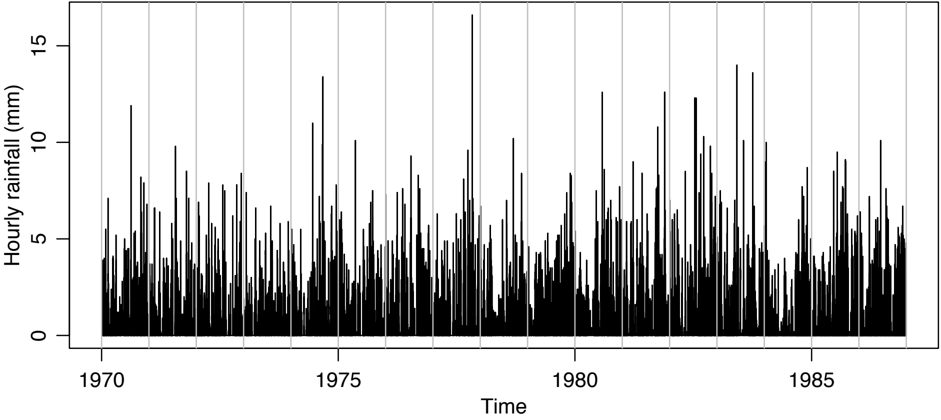
Simple estimates of the extremal index are based on empirical means of clusters. Here the runs method is used: a cluster is deemed to have terminated when there are r consecutive observations below the threshold. Left: $r = 1$ gives 4 clusters. Right: $r = 3$ gives 2 clusters.

Example: Wooster temperatures

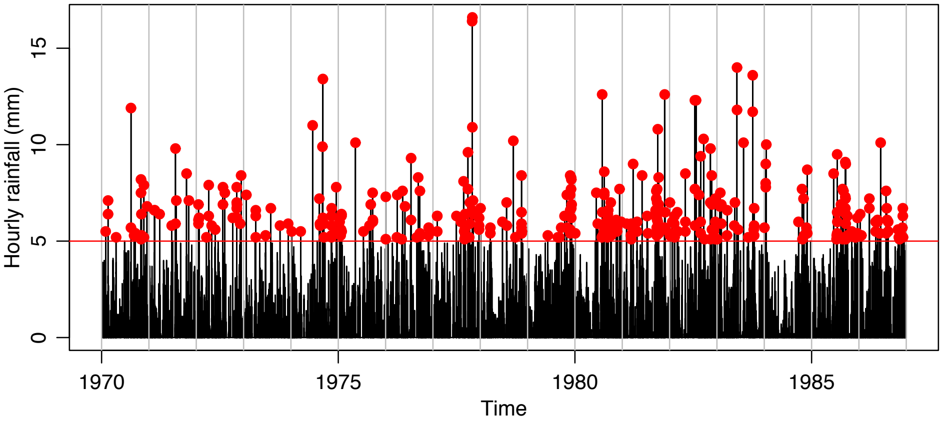
	$u = -10$		$u = -20$	
	$r = 2$	$r = 4$	$r = 2$	$r = 4$
n_c	31	20	43	29
$\hat{\beta}$	11.8 (3.0)	14.2 (5.2)	17.4 (3.6)	19.0 (4.9)
$\hat{\xi}$	−0.29 (0.19)	−0.38 (0.30)	−0.36 (0.15)	−0.41 (0.19)
$\hat{R}_{100}$	27.7 (12.0)	26.6 (14.4)	26.2 (9.3)	25.7 (9.9)
$\hat{\theta}$	0.42	0.27	0.24	0.16

Results may be sensitive to choice of threshold u, and run length r.
Return level is relatively stable towards u and r.

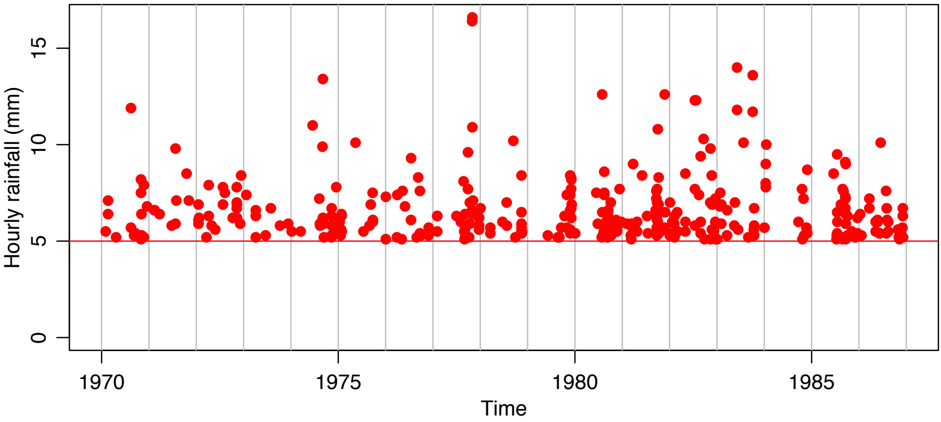
Example: Eskdalemuir rainfall



Example: Eskdalemuir rainfall



Example: Eskdalemuir rainfall



Example: Eskdalemuir rainfall

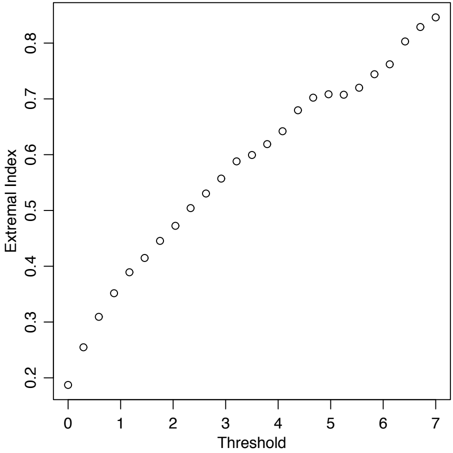
```
rain.fit <- fpot(esk.rain, threshold=5, npp=365.25*24)

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Estimates
     loc      scale      shape
10.13628   1.86637   0.06696

Standard Errors
    loc     scale     shape
0.35380  0.23673  0.05379
```

Example: Eskdalemuir rainfall

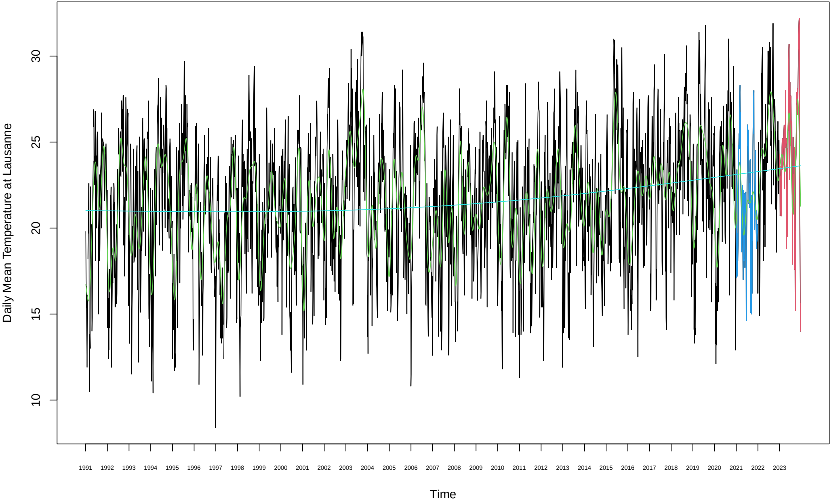
```
rain.fit <- fpot(esk.rain, threshold=5, cmax=T, npp=365.25*24)

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Number of Clusters: 272
Extremal Index: 0.764

Estimates
    loc     scale     shape
9.81557  1.83937  0.04178

Standard Errors
   loc    scale   shape
0.3519  0.2406  0.0632
```

Example: Eskdalemuir rainfall

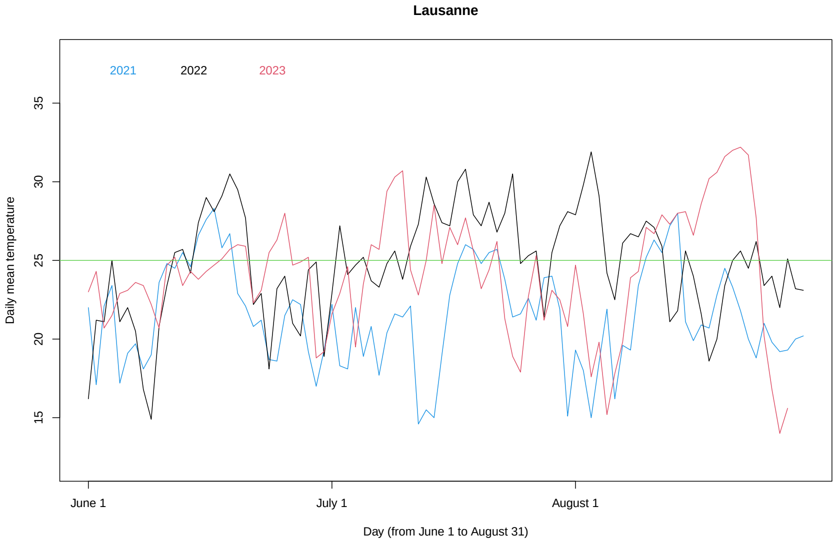
For finite threshold u, θ increases with u, suggesting that very extreme hourly rainfall totals occur singly.



Non-stationarity example: Daily mean temperature in Lausanne during summer



Non-stationarity example: Daily mean temperature during summer



Non-stationarity

General results not broad enough for application — hence, model trends, seasonality and covariate effects by parametric or nonparametric models for the usual extreme value model parameters.

Some possibilities for parametric modelling:

$$
\begin{aligned}
\mu(t) &= \alpha + \beta t; \\
\sigma(t) &= \exp(\alpha + \beta t); \\
\xi(t) &= \begin{cases} \xi_1, & t \le t_0, \\ \xi_2, & t > t_0; \end{cases} \\
\mu(t) &= \alpha + \beta y(t).
\end{aligned}
$$

Parameter estimation

- Model specification (example)

$$Z_t \sim \mathrm{GEV}(\mu(t), \sigma(t), \xi(t)),$$

- Likelihood (for complete parameter set β),

$$L(\beta) = \prod_{t=1}^{m} h(z_t; \mu(t), \sigma(t), \xi(t)),$$

  where h is GEV model density.

- Maximization of L yields maximum likelihood estimates.
- Standard likelihood techniques also yield standard errors, confidence intervals, etc.

Model reduction

- For nested models $\mathcal{M}_0 \subset \mathcal{M}_1$, the deviance statistic is

$$D = 2\left(\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\right),$$

  where ℓ_i is the log-likelihood of model i.

- Based on asymptotic likelihood theory, $\mathcal{M}_0$ is rejected by a test at the α-level of significance if $D > c_\alpha$, where c_α is the $(1-\alpha)$ quantile of the χ^2_k distribution, and k is the difference in the dimensionality of $\mathcal{M}_1$ and $\mathcal{M}_0$.

Model diagnostics

- Assuming a fitted model

$$Z_t \sim \text{GEV}\left(\hat{\mu}(t), \hat{\sigma}(t), \hat{\xi}(t)\right),$$

the standardized variables

$$\tilde{Z}_t = \frac{1}{\hat{\xi}(t)} \log\left(1 + \hat{\xi}(t)\frac{Z_t - \hat{\mu}(t)}{\hat{\sigma}(t)}\right),$$

each have the standard Gumbel distribution, with probability distribution function

$$\mathbb{P}(\tilde{Z}_t \leq z) = \exp(-e^{-z}), \quad z \in \mathbb{R}.$$

Possible diagnostics:

- probability plot: $\left\{i/(m+1), \exp(-\exp(-\tilde{z}_{(i)}));\ i = 1, \ldots, m\right\}$
- quantile plot: $\left\{\left(-\log\left[-\log\{i/(m+1)\}\right], \tilde{z}_{(i)}\right);\ i = 1, \ldots, m\right\}$

Asymptotic model for minima

- Let $X_1, X_2, \ldots$ be iid from F and let $\tilde{M}_n = \min(X_1, \ldots, X_n)$.

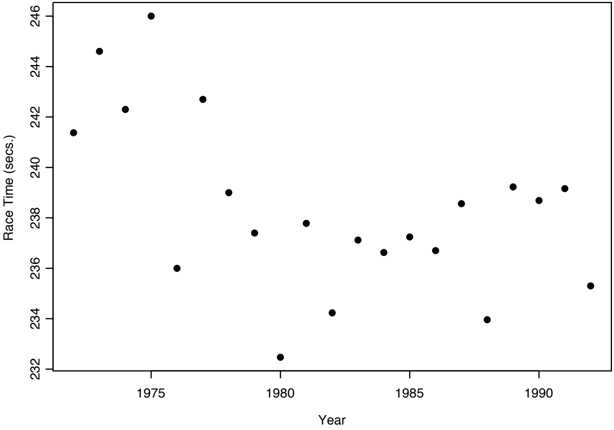
- **Theorem** If we can find sequences of real numbers $a_n > 0$ and b_n such that $\left(\tilde{M}_n - b_n\right)/a_n$, the sequence of normalized minima, converges in distribution to a non-degenerate distribution $\tilde{H}$, then $\tilde{H}$ is a GEV distribution for minima defined by

$$\tilde{H}_{\tilde{\mu},\sigma,\xi}(x) = 1 - \exp\left(-\left(1 - \xi\left(\frac{x-\tilde{\mu}}{\sigma}\right)\right)^{-1/\xi}\right)$$

  where $1 - \xi(x - \tilde{\mu})/\sigma > 0$.

- If $Y_i = -X_i$, then the normalized $\max\{Y_i\}$ converge to $H_{\mu,\sigma,\xi}$. It is easy to show that the normalized $\min\{X_i\}$ converge to $\tilde{H}_{\tilde{\mu},\sigma,\xi}$, where $\tilde{\mu} = -\mu$.

Example: Race times



Annual fastest race times for women's 1500m event, with an obvious time trend.

Example: Race times

- We model the race time Z_t in year t as

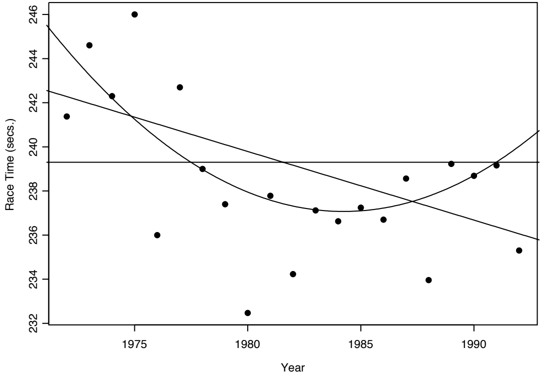
$$Z_t \sim \tilde{H}(\tilde{\mu}(t), \sigma, \xi).$$

- The motivation for this model is that fastest race times in each year are the minima of many such race times. But because of overall improvements in athletic performance, the distribution is non-homogeneous across years.

- Models:

$$\begin{aligned} \text{Constant: } & \tilde{\mu}(t) = \beta_0 \\ \text{Linear: } & \tilde{\mu}(t) = \beta_0 + \beta_1 t \\ \text{Quadratic: } & \tilde{\mu}(t) = \beta_0 + \beta_1 t + \beta_2 t^2 \end{aligned}$$

Example: Race times



Fitted models for location parameter in womens 1500 metre race times.

Example: Race times

Model	Log-likelihood	$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$	$\hat{\sigma}$	$\hat{\xi}$
Constant	-54.5	239.3	3.63	-0.469
		(0.9)	(0.64)	(0.141)
Linear	-51.8	$242.9, -0.311$	2.72	-0.201
		(1.4, 0.101)	(0.49)	(0.172)
Quadratic	-48.4	$247.0, -1.395, 0.049$	2.28	-0.182
		(2.3, 0.420, 0.018)	(0.45)	(0.232)

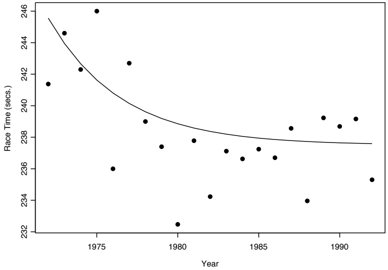
Quadratic model apparently preferable. But it would lead to slower races in recent and future events.

Example: Race times

Alternative exponential model

$$\tilde{\mu}(t) = \beta_0 + \beta_1 e^{-\beta_2 t}.$$

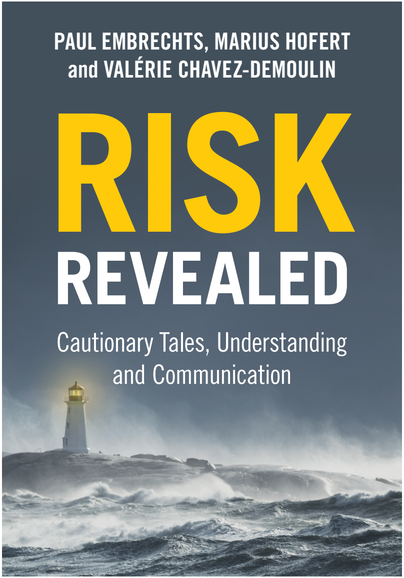
has log-likelihood -49.5. Not so good as quadratic model, though comparison via likelihood ratio test is invalid as models are not nested. Better behaviour for large t suggests a preferable model though.



Other extreme value models

- Similar techniques are applicable for the threshold exceedance model, but threshold selection is likely to be a more sensitive issue.
- Time-varying thresholds may also be appropriate, though there is little guidance on how to make such a choice.
- Use of covariates can sometimes be helpful.

To go further



Risk Analytics

Monitoring and reporting risk

Frédéric Aviolat & Juraj Bodik

2025

Talking about risk

Risk is notoriously hard to quantify:

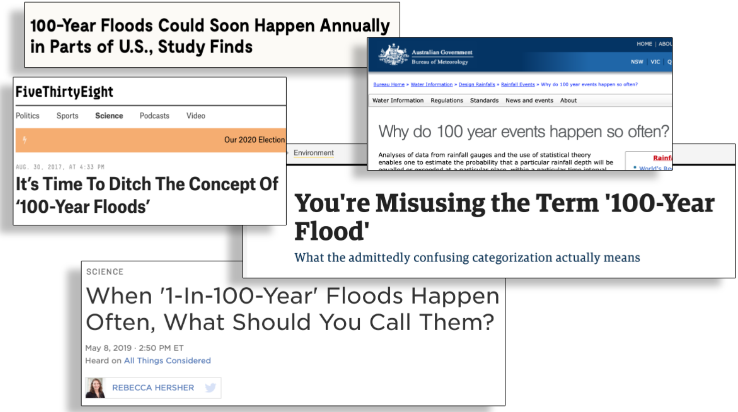


Figure: Even an event with 1% chance can happen regularly (!).

Risk management

- Many management decisions are taken under risk.
- Understanding and communicating risk effectively is important.
- It is also important to check if your risk model is well-calibrated.

- In this module:
  1. Modelling Risks
  2. Risk measures
  3. Uncertainty
  4. Backtesting

Modelling Risks

- Risk = possibility of loss or of an unfavourable outcome happening associated with an action.
- To model risk we use language of probability theory. Risks are represented by random variables mapping unforeseen future states of the world.
- The risks which interest us include:
  - number of patients in the hospital
  - stocks and bonds rate
  - sea-level
  - heatwave
  - wildfire
  - a demand too big to be met
  - ...

Variable of interest or risk factor

- Denote X_t the variable value at time t or the risk factor at time t. We assume this random variable is observable at time t.
- Suppose we look at risk from perspective of time t and we consider the time period $[t, t+1]$. The value X_{t+1} at the end of the time period is unknown to us.

Conditional or Unconditional Distribution?

- This issue is related to the time series properties of $(X_t)_{t\in\mathbb{N}}$, the series of risk factor changes. If we assume that $X_t, X_{t-1}, \ldots$ are iid random variables, the issue does not arise. But, if we assume that they form a strictly stationary time series then we must differentiate between conditional and unconditional.
- Many standard accounts of risk management fail to make the distinction between the two.
- If we cannot assume stationarity of the time series at risk for at least some window of time extending from the present back into intermediate past, then any statistical analysis is difficult.

The Conditional Problem

- Let $\mathcal{H}_t$ represent the history of the risk factor up to the present.
- In the conditional problem we are interested in the distribution of X_{t+1} given $\mathcal{H}_t$, i.e. the conditional (or predictive) distribution of the risk factor for the next time interval given the history of risk factor developments up to time t.
- This problem forces us to model the dynamics of the risk factor time series

The Unconditional Problem

- In the unconditional problem we are interested in the distribution of X_{t+1} when X is a generic variable of risk factor change with the same distribution F_X as $X_t, X_{t-1}, \ldots$.
- When we neglect the modelling of dynamics we inevitably take this view. Particularly when the time interval is large, it may make sense to do this.

More Formally

Conditional loss distribution: $F_{[X_{t+1}|\mathcal{H}_t]}$
Unconditional loss distribution: F_X

Risk measure: Value-at-Risk

- In our usual context, X is an indicator with higher values being worse.
- The Value-at-Risk of X at level α, denoted $\mathrm{VaR}_\alpha(X)$, is the α-quantile of F_X:

$$\mathrm{VaR}_\alpha(X) = F_X^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\},$$

with α typically high (close to 1).

Interpretation 1: With probability α, the variable X will be smaller than $\mathrm{VaR}_\alpha(X)$.
Interpretation 2: $\mathrm{VaR}_\alpha(X)$ is the value that the variable can exceed with low probability $1-\alpha$.

Return levels vs Value-at-Risk

- Value-at-Risk is typically used in finance and insurance.
- In other fields, the notion of return level is preferred.
- The *n*-period return level is $R_n = F^{-1}(1 - \frac{1}{n})$, the $(1 - 1/n)$-quantile. The name refers to the following idea:

$$
\begin{aligned}
F(y_n) = 1 - 1/n &\Longleftrightarrow \mathbb{P}(Y \le y_n) = 1 - 1/n \\
&\Longleftrightarrow \mathbb{P}(Y > y_n) = 1/n,
\end{aligned}
$$

  so that, on average, among n draws of Y, one exceeds the value y_n.
- But this doesn't mean only one excess happens!
  - Under independence, the number of excesses among n is Binomial$(n, 1/n)$, so the probability of having at least two excesses among $n = 100$ is 26%.

Risk measure: Expected Shortfall

- Value-at-Risk gives an upper bound on risk.
- It doesn't quantify what happens if $X > x$, only the probability of this happening.
- But what if in fact $X > x$?

The Expected Shortfall at level α is the average value of X, when it is higher than $\mathrm{VaR}_\alpha(X)$:

$$\mathrm{ES}_\alpha(X) = \mathbb{E}\left(X \mid X > \mathrm{VaR}_\alpha(X)\right)$$

Interpretation: If I have an extreme event, how bad does it get?

For a continuous distribution,

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\beta(X)\,\mathrm{d}\beta$$

VaR and ES in Visual Terms

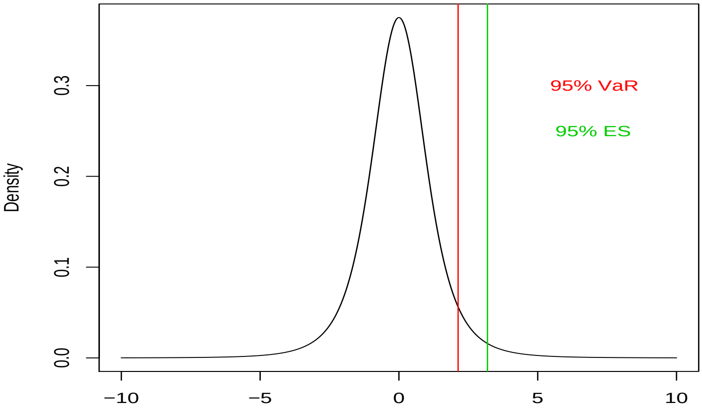


Figure: Risk measures at the 95% level. The area right of the red line has probability 0.05. $\mathrm{VaR}_\alpha(X) \simeq 2$, $\mathrm{ES}_\alpha(X) \simeq 3.5$.

Ratio of risk measures

The expected shortfall can be much larger than the Value-at-Risk.

Examples

$$X \sim N(\mu, \sigma^2) \implies \lim_{\alpha \to 1} \frac{\mathrm{ES}_\alpha(X)}{\mathrm{VaR}_\alpha(X)} = 1$$

$$X \sim t_\nu \implies \lim_{\alpha \to 1} \frac{\mathrm{ES}_\alpha(X)}{\mathrm{VaR}_\alpha(X)} = \frac{\nu}{\nu - 1} > 1$$

Return Levels: block maxima approach

We address the following question:

- What is the k–period return level $R_{n,k}$?

In this question we define and estimate a rare stress or scenario loss. $R_{n,k}$, the k n–block return level, is defined by

$$\mathbb{P}\left(M_n > R_{n,k}\right) = 1/k\,;$$

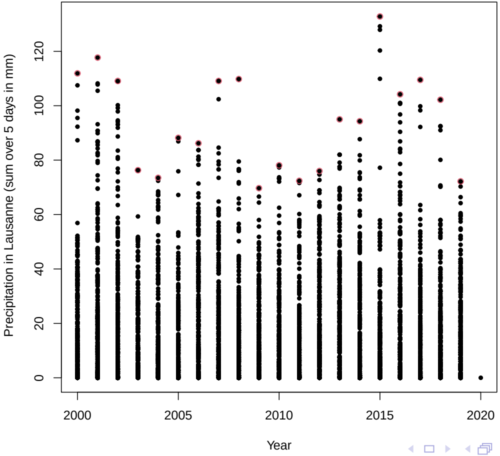
i.e. it is that level which is exceeded in one out of every k n–blocks, on average.
We use the approximation by the $(1-1/k)$-quantile of a GEV

$$R_{n,k} \approx H^{-1}_{\xi,\mu,\sigma}(1-1/k) \approx \mu + \sigma \left(\left(-\log(1-1/k)\right)^{-\xi} - 1 \right)/\xi\,.$$

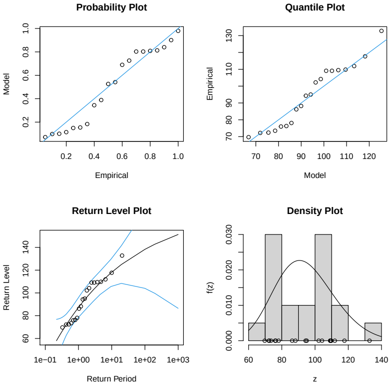
We wish to estimate this functional of the unknown parameters of our GEV model for maxima of n–blocks.

Return Levels: block maxima approach
Example: Rainfall yearly maxima

- What is the 10–period return level $R_{365,10}$? i.e. what level is exceeded in one out of every 10 years, on average?



Return Levels: block maxima approach
Example: Rainfall yearly maxima



$$\hat{R}_{365,10} \approx \hat{H}^{-1}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(1 - 1/k) \approx 87.2 + 16.5\frac{(-\log(1 - 1/10))^{0.18} - 1}{-0.18}$$

≈ 117 mm is the estimated value of daily rainfall (sum over the last 5 days) that can be exceeded once every 10 years.

Quantifying uncertainty

- Whenever reporting or assessing a measure of risk, one should always take into account uncertainty.
  - EVT model is an approximation for finite data.
  - Parameters are estimated with some uncertainty (e.g. profile likelihood).
- More flexible: simulation from the fitted model (accounts for uncertainty in formulas).
- One can even simulate from estimator distribution (accounts for parameter uncertainty).

Quantifying uncertainty: simulation example

- Concrete example
- $X_i \sim \mathrm{GEV}(\hat{\mu}, \hat{\sigma}, \hat{\xi})$, with the usual maximum likelihood distributions

$$\hat{\mu} \sim N(397, 5), \quad \hat{\sigma} \sim N(48, 2.9), \quad \hat{\xi} \sim N(-0.25, 0.03)$$

- Algorithm to estimate $\mathrm{VaR}_\alpha(X_i)$:
  1. For $b = 1, \ldots, B$
     - 1.1 Draw parameters μ_b, σ_b, ξ_b from the above distributions
     - 1.2 Set

$$y_b = \mu_b + \frac{\sigma_b}{\xi_b}\left((-\log(1-\alpha))^{-\xi_b} - 1\right)$$

  2. The $(y_b)_{b=1}^B$ are an approximate distribution for $\mathrm{VaR}_\alpha(X_i)$.
- Parametric, because we assume the ML distributions are accurate. Could do a full bootstrap by re-fitting the model on resampled data in step 1.1.

The POT Model: summary

When the data form a stationary time series then the timing and magnitude of threshold exceedances are both of interest. The POT (peaks-over-thresholds) model is a limit model for threshold exceedances in iid processes. The limit is derived by considering datasets $X_1, \ldots, X_n$ and thresholds u_n that increase with n and letting $n \to \infty$. The limit model says that:

- Exceedances occur according to a homogenous Poisson process in time.
- Excess amounts above the threshold are iid and independent of exceedance times.
- Distribution of excesses is generalised Pareto (GPD).

GPD Model: Calculation of VaR and Expected Shortfall

Let $X \sim F$ with $F \in \mathrm{MDA}(\mathrm{H}_\xi)$ and consider the excess $W = X - u$ where u is a high threshold. We have $W \sim \mathrm{GPD}(\xi, \beta)$.

- For $\alpha \geq F(u)$, the Value at Risk is:

$$\mathrm{VaR}_\alpha(X) = u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\xi} - 1\right).$$

  where $\bar{F}(u) = \mathbb{P}(X > u)$ can be approximated by N_u/n.

- Assuming that $\xi < 1$ the associated expected shortfall can be calculated easily. We obtain

$$\mathrm{ES}_\alpha(X) = \frac{\mathrm{VaR}_\alpha(X)}{1-\xi} + \frac{\beta - \xi u}{1-\xi}.$$

Ratios of Risk Measures

It is interesting to look at how the ratio of the two risk measures behaves for large values of the quantile probability α. We can show that

$$\lim_{\alpha \to 1} \frac{\mathrm{ES}_\alpha(X)}{\mathrm{VaR}_\alpha(X)} = \begin{cases} (1-\xi)^{-1}, & \xi \geq 0, \\ 1, & \xi < 0, \end{cases}$$

so that the shape parameter ξ of the GPD effectively determines the ratio when we go far enough out into the tail.

Estimating Tails and Risk Measures

- Recall that for $x > u$ in the GPD model,

$$\bar{F}(x) = \bar{F}(u)\left(1 + \xi\frac{x-u}{\beta}\right)^{-1/\xi}$$

- Tail probabilities, VaRs and ESs are all given by formulas of the form $g(\xi, \beta, \bar{F}(u))$. We estimate these quantities by replacing ξ and β by their estimates and replacing $\bar{F}(u)$ by the simple empirical estimator N_u/n (fraction of excesses).
- For tail probabilities we use the estimator

$$\hat{\bar{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)_+^{-1/\hat{\xi}},$$

which is valid for $x \geq u$.
- Asymmetric confidence intervals can be constructed using profile likelihood method.
  - A high u reduces bias in estimating excess function.
  - A low u reduces variance in estimating excess function and $F(u)$.

Purposes of Risk Measurement

Risk measures are used for the following purposes:

- Determination of risk capital. Risk measure gives amount of capital needed as a buffer against (unexpected) future losses to satisfy a regulator. *Basel committee. Solvency*.
- Management tool. Risk measures are used in internal limit systems. *High demands*.
- Protect against climate/meteorological extremes. *Floods, droughts, heatwaves*.
- Protect against technological problems. *Cyber attacks*.

Backtesting (I)

Suppose we observe a time series of T iid data $x_1, \ldots, x_T$ on which we fit a model and want to backtest the VaR/risk measure.

- We split the data into two parts: one $(x_1, \ldots, x_m)$ is the historical data we will use to estimate the VaR and $x_{m+1}, \ldots, x_T$ is a window of $n = T - m$ observations on which we want to backtest the VaR_α.
- At each time $t = 1, \ldots, n$, we use a window of time of size w of the historical data $x_{m+t-w}, \ldots, x_{m+t-1}$ to estimate the VaR at time $m + t$.
- We compare the estimated VaR_α with the observed value at time $m + t$, that is x_{m+t}.

Backtesting (II)

- A violation occurs whenever $x_{m+t} > \widehat{\mathrm{VaR}}_\alpha(x_{m+t})$.
- If the model is correct, for a level α, a violation occurs according to a Bernoulli law with probability $1 - \alpha$.
- Considering that the violations are independent, the total number of violations over the n tested data follows a Binomial distribution with size n and probability $1 - \alpha$.
- We can perform a Binomial test to formally test the null hypothesis that the observed number of violations (number of times the observation is larger than the estimated VaR) is equal to the expected number of violations, that is $n(1 - \alpha)$.

Risk Analytics

As human society becomes more complex and interconnected, it is more vulnerable to rare but catastrophic events, such as the coronavirus pandemic, the extremes of climate that regularly affect all regions of the world or the continuing turbulence in the financial markets. For risk managers at major companies, banks and public bodies, an accurate quantitative assessment of the risks linked to such events plays an increasingly crucial role in decision making processes. Risk assessment involves using past observations to forecast the future as well as possible, often extrapolating beyond existing data, and assessing the uncertainties surrounding these forecasts. A critical awareness of the statistical/stochastic ideas behind such calculations is essential in understanding both their limitations and their sensitivity to failure of the underlying assumptions, and, thus, in appreciating when and why such extrapolation may be particularly dangerous.

Objectives

Upon completion of that course the students will be able to

- Apply techniques from a general methodological toolkit for measuring risk in most fields of risk management
- Use risk management tools available in the statistical language R
- Analyse and estimate risk measures and their uncertainty assessment through concrete datasets from fields of operations, insurance, finance and environment.

Course

This course on methodologies for risk assessing includes the following topics:

- Why risk analytics matters?
- Time series analysis
- Extreme value theory (EVT)
- Monitoring and reporting risk

Prerequisites

Basic statistics and knowledge of statistical software R.

References

- McNeil, A. J., Rüdiger, F. and Embrechts, P., 2015, *Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition*, Princeton Series in Finance.
- Stuart Coles, 2001, *An Introduction to Statistical Modeling of Extreme Values*, Springer Verlag.
- Embrechts, P., Hofert, M. & Chavez-Demoulin, V., 2024, *Risk Revealed (Cautionary Tales, Understanding and Communication).* Cambridge University Press.

Exam

Without exam.
Evaluation: At the end of the project, students will have to provide a detailed written report and give a presentation.

Final grade = 0.5 * Report grade + 0.5 * Presentation grade

Retake

Without exam
Evaluation: A complement to the report will be asked. The final report is 100% of the grade.

Risk Analytics

Time Series: summary and practice

Frédéric Aviolat & Juraj Bodik

2025

ARMA processes

- AR(p):

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z};$$

- MA(q):

$$X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad t \in \mathbb{Z};$$

- ARMA(p, q):

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad t \in \mathbb{Z},$$

where (ε_t) is a white noise process.

ARIMA

- ARIMA is an extension to ARMA, with differencing. (I stands for Integrated.)
- Instead of using series (y_t), one can use
  - first-order differencing: $y'_t = y_t - y_{t-1}$;
  - second-order differencing: $y''_t = y'_t - y'_{t-1}$;
  - ...
  - seasonal differencing: $y^m_t = y_t - y_{t-m}$, where m is the seasonality period.
- ARIMA(p, d, q) is ARMA(p, q) on the d^{th}-order differences.
- ARMA(p, q) $\equiv$ ARIMA(p, 0, q)

ARMA/ARIMA in `R`

In library `fpp2` (see Hyndman and Athanasopoulos [2018]):

- function `Arima`(p, d, q) specifies the orders of the AR, differencing and MA components.
- function `auto.arima` fits best ARIMA model.

In library `fpp3` (see Hyndman and Athanasopoulos [2021]):

- use function `ARIMA`, with parameters (p, d, q) as above and (P, D, Q) for seasonal components (see doc).

ARCH/GARCH

- General form:

$$X_t = \sigma_t Z_t, \quad t \in \mathbb{Z};$$

- ARCH(p):

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2, \quad \alpha_j > 0.$$

- GARCH(p, q) process:

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2, \quad \alpha_j, \beta_k > 0.$$

GARCH in R

- Use library fGarch
- See code

References

Rob J Hyndman and George Athanasopoulos. Forecasting: Principles and Practice, 2nd edition. OTexts, https://otexts.com/fpp2/, 2018.

Rob J Hyndman and George Athanasopoulos. Forecasting: Principles and Practice, 3rd edition. OTexts, https://otexts.com/fpp3/, 2021.

Risk Analytics

Time Series

Frédéric Aviolat & Juraj Bodik

2025

Basics of Time Series Analysis

Definition A time series is a collection of random elements indexed by an index set.

Mathematically, we write such a collection $(X_t)_{t \in I}$, where I is an index set.

Usually, the random elements are random variables with values in $\mathbb{R}$ or $\mathbb{R}^d$.

If the index set $I \subset \mathbb{R}$ (real numbers), it is a continuous time series.
If the index set $I \subset \mathbb{N}$ or $\mathbb{Z}$ (integers), it is a discrete time series.

In this course, we will work with real-valued discrete time series.

(In what follows, we take $I = \mathbb{Z}$ for ease of presentation.)

Moments of a Random Variable

The moments of a random variable X are indicators of the shape of its distribution.

- First moment is the expectation or the mean of the variable:

$$\mu = \mathbb{E}[X]$$

- Second moment is the variance of the variable:

$$\sigma^2 = \mathbb{E}[(X - \mu)^2]$$

- Third moment is the skewness of the variable:

$$\gamma = \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^3\right]$$

Stationarity of a Time Series

Definition

A time series $(X_t)_{t \in \mathbb{Z}}$ is stationary if

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{\mathrm{d}}{=} (X_{t_1+h}, \ldots, X_{t_n+h})$$

for all $t_1, \ldots, t_n, h \in \mathbb{Z}$.

In particular this means that X_t has the same distribution for all $t \in \mathbb{Z}$, and this distribution is known as the stationary distribution (or marginal distribution).

Unlike an independent series, stationarity does not forbid X_t being dependent on previous values, although X_{t+h} must have the same dependence on its previous values as X_t.

A stationary time series cannot have trend, seasonality or other deterministic cycles.

How to make a time-series stationary ?

Motivation

- Stationary time-series are easier to model.
- Many theoretical results are only valid under stationarity conditions.

Tranformation techniques

- **Differencing** $\tilde{X}_t = X_t - X_{t-1}$
  - Helps to stabilize the mean.
  - Eliminates trends.
  - Can also remove seasonality, if differences are taken appropriately (e.g. differencing observations 1 year apart to remove a yearly seasonality).
- **Logarithm** $\tilde{X}_t = \log(X_t)$
  - Helps to stabilize the variance.
  - Removes multiplicative dependency.

Autocovariance

- Let $(X_t)_{t\in\mathbb{Z}}$ be a time series.
- The autocovariance function of (X_t) is defined by

$$\gamma(t,s) = \mathrm{cov}(X_t, X_s) = \mathbb{E}\left[(X_t - \mathbb{E}(X_t))(X_s - \mathbb{E}(X_s))\right].$$

- If (X_t) is a stationary time series, then

$$\gamma(t,s) = \gamma(t+h, s+h), \quad \text{for all } t,s,h \in \mathbb{Z}$$

- This implies that covariance only depends on the separation in time of the observations $|t-s|$, also known as the lag.
- A time series for which the first two moments are constant over time (and finite) and for which this condition holds, is known as covariance stationary, or second-order stationary.

The Autocorrelation Function

- Rewrite the autocovariance function of a stationary time series as

$$\gamma(h) = \gamma(h, 0) = \mathrm{cov}(X_h, X_0), \quad h \in \mathbb{Z}.$$

- Note that $\gamma(0) = \mathrm{var}(X_0) = \mathrm{var}(X_t)$, $t \in \mathbb{Z}$.
- The autocorrelation function is given by

$$\rho(h) = \gamma(h)/\gamma(0), \quad h \in \mathbb{Z}.$$

  Observe that $\rho(0) = 1$.
- We refer to $\rho(h), h = 1, 2, 3, \ldots$ as autocorrelations or serial correlations.

Time Domain

- If we study dependence structure of a time series by analysing the autocorrelations we analyse (X_t) in the time domain.
- An important instrument in the time domain is the correlogram, which gives empirical estimates of serial correlations.

Correlogram

- Given time series data $X_1, \ldots, X_n$ we calculate the sample autocovariances

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \overline{X})(X_{t+h} - \overline{X}) \quad \text{where} \quad \overline{X} = \sum_{t=1}^{n} X_t / n.$$

  and the sample autocorrelations

$$\widehat{\rho}(h) = \widehat{\gamma}(h) / \widehat{\gamma}(0), \;\; h = 0, 1, \ldots.$$

- The correlogram (ACF) is the plot

$$\{(h, \widehat{\rho}(h)), \;\; h = 0, 1, 2, \ldots\}.$$

- For many standard underlying processes, it can be shown that the $\widehat{\rho}(h)$ are consistent, and asymptotically normal estimators of the autocorrelations $\rho(h)$.

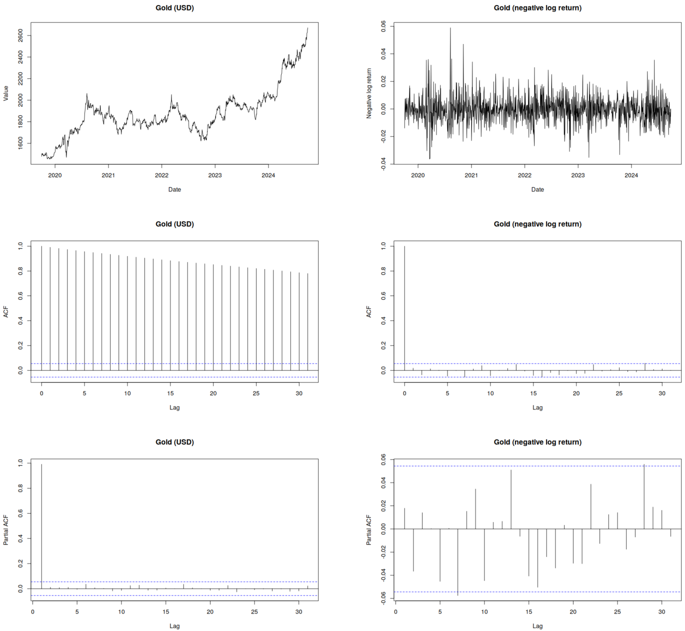
Correlogram (variant)

- A variant to the AutoCorrelation Function (ACF) is the Partial AutoCorrelation Function (PACF).
- The partial autocorrelation π_k at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.
- Formally, it is calculated recursively as

$$\pi_n = \pi_{n,n} = \frac{\rho(n) - \displaystyle\sum_{k=1}^{n-1} \pi_{n-1,k}\,\rho(n-k)}{1 - \displaystyle\sum_{k=1}^{n-1} \pi_{n-1,k}\,\rho(k)}$$

where $\pi_{n,k} = \pi_{n-1,k} - \pi_n\,\pi_{n-1,n-k}$ for $1 \leq k \leq n-1$.

Example: Gold rate



White Noise Processes

- Processes with no appreciable dependence structure in the time domain.
- A white noise process is a covariance stationary time series process whose autocorrelation function is given by

$$\rho(0) = 1, \quad \rho(h) = 0, \ h \neq 0.$$

  That is, a process showing no serial correlation.
- A strict white noise process is a process of independent and identically distributed (i.i.d.) random variables.

Classical ARMA Processes

- Classical ARMA (AutoRegressive Moving Average) processes are constructed from white noise.
- Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise process with mean zero and finite variance σ_ε^2.
- The (ε_t) form the innovations that drive the ARMA process.

Moving Average Process

- These are defined as linear sums of the noise (ε_t).
- (X_t) follows a MA(q) process if

$$X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t.$$

Autoregressive Process

- These are defined by stochastic difference equations, or recurrence relations.
- $(X_t)_{t\in\mathbb{Z}}$ follows a AR(p) process if for every t

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t,$$

  where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise.
- In order for these equations to define a covariance stationary causal process (depending only on past innovations) the coefficients ϕ_j must obey certain conditions.

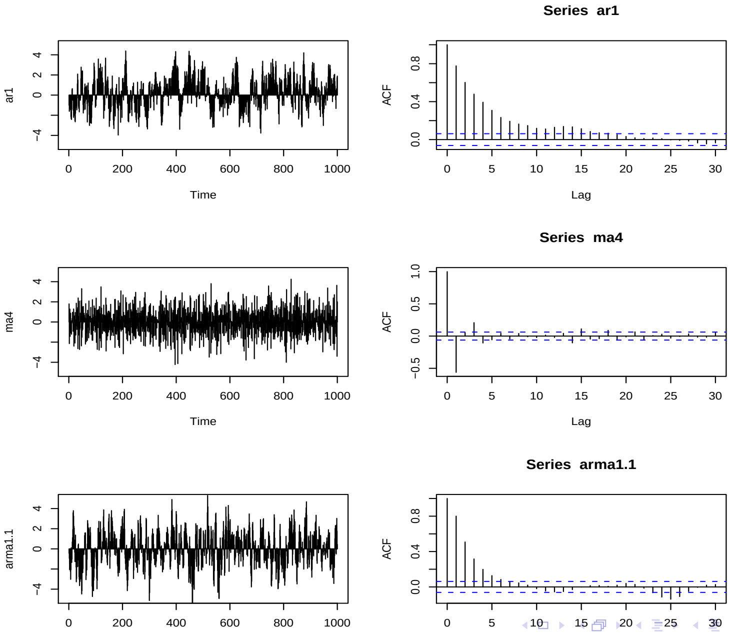
ARMA Process

- Autoregressive and moving average features can be combined to form ARMA processes.
- $(X_t)_{t\in\mathbb{Z}}$ follows an ARMA(p,q) process if for every t

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \qquad (1)$$

  where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise.
- Again, there are conditions on the coefficients ϕ_j and θ_i for these equations to define a covariance stationary causal process.
- The autocorrelation functions of ARMA processes show a number of typical patterns, including exponential decay and damped sinusoidal decay.

ARMA Examples



Example: AR(1) Process

Consider an AR(1) process:

$$X_t = \varepsilon_t + \phi X_{t-1} \tag{2}$$

We can calculate recursively that

$$\begin{aligned} X_t &= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2 X_{t-2} \\ &= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3 X_{t-3} \\ &= \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \end{aligned} \tag{3}$$

Thus this process is an infinite order MA process (or MA(∞)). The condition for a stationary solution to (2) is that the sum (3) should converge. We require $|\phi| < 1$.

Example: AR(1) Process, Continued

- The autocovariances of this process can be calculated:

$$\begin{aligned}\gamma(s) &= \operatorname{cov}(X_t, X_{t+s}) \\ &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right)\left(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t+s-j}\right)\right] \\ &= \frac{\phi^{|s|}\sigma_\varepsilon^2}{1-\phi^2}\end{aligned}$$

- Therefore $\rho(s) = \gamma(s)/\gamma(0) = \phi^{|s|}, \quad s \in \mathbb{Z}$.
- The serial correlations decay exponentially with possibly alternating sign.

Stationarity conditions

AR(p) process $$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t$$

- For an AR(1) model: $-1 < \phi_1 < 1$.
- For an AR(2) model: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.
- When $p \geq 3$, the conditions are much more complicated.

MA(q) process $$X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

- For an MA(1) model: $-1 < \theta_1 < 1$.
- For an MA(2) model: $-1 < \theta_2 < 1$, $\quad \theta_2 + \theta_1 > -1$, $\theta_1 - \theta_2 < 1$.
- When $q \geq 3$, the conditions are much more complicated.

ARMA(p, q) process $$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

- Conditions on ϕ_j and θ_i are also quite complicated and are usually taken care of by the software function.

ARIMA process

- ARIMA stands for AutoRegressive Integrated Moving Average.
- An ARIMA(p, d, q) process is an ARMA(p, q) process over the time series differenced d times.
- The I stands for integrated (i.e. opposite of differenced)

- ARIMA processes allow to model a very wide variety of time series patterns. They provide statistical models that fit the historical data well, taking into account the time based variation, and therefore they allow to produce meaningful and relevant forecasts.

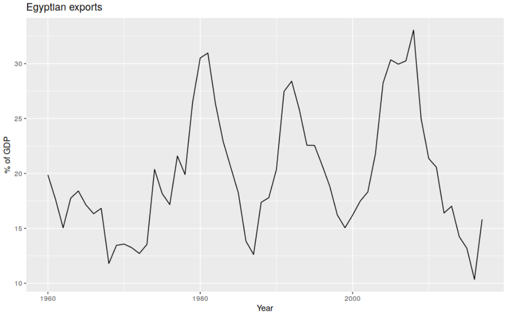
How to choose p, q and d ?

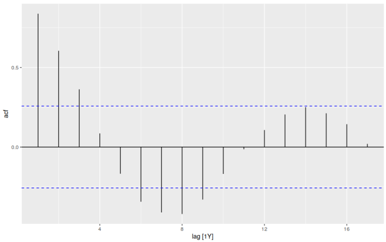
- Each partial autocorrelation can be estimated as the last coefficient in an autoregressive model. Specifically, π_k, the k^{th} partial autocorrelation coefficient, is equal to the estimate of ϕ_k in an AR(k) model.
- The data may follow an ARIMA(p,d,0) model if the ACF and PACF plots of the differenced data show the following patterns:
  - the ACF is exponentially decaying or sinusoidal;
  - there is a significant spike at lag p in the PACF, but none beyond lag p.
- The data may follow an ARIMA(0,d,q) model if the ACF and PACF plots of the differenced data show the following patterns:
  - the PACF is exponentially decaying or sinusoidal;
  - there is a significant spike at lag q in the ACF, but none beyond lag q.
- The time series should be differenced as many times as needed so that it becomes stationary. Thus d is the number of differencing needed.

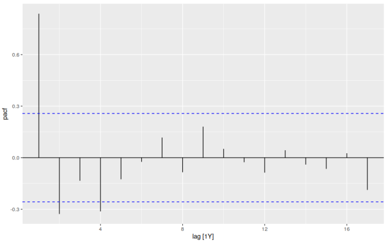
Estimating the parameters

- We need to estimate the parameters of the model:
  - autoregressive coefficients ϕ_j
  - moving average coefficients θ_i
  - variance σ_ε of the innovation process (ε_i)
- Let $\mathbf{\Phi} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_\varepsilon)$. The goal is to optimize a function $Q(\mathbf{\Phi})$ subject to some linear and non-linear constraints.
- The following methods can be used:
  - Least squares method:
    minimize the sum of squared residuals
  - Maximum likelihood estimation:
    maximize the (log-)likelihood function
  - Generalized method of moments:
    minimize the distance between the theoretical moments and 0
- High quality software programs, such as R are available to do it.

ARIMA example







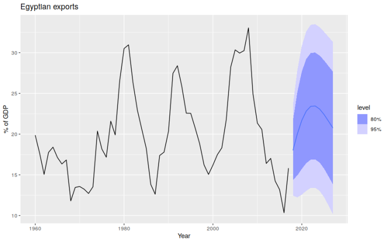
ARIMA example (cont'd)

- Model 1: ARIMA(4,0,0)

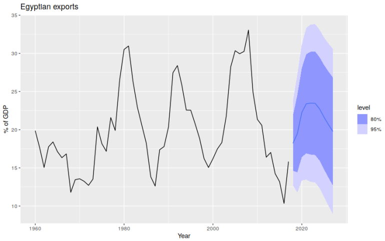
$$X_t = 0.9861X_{t-1} - 0.1715X_{t-2} + 0.1807X_{t-3} - 0.3283X_{t-4} + 6.6922$$

- Model 2: ARIMA(2,0,1)

$$X_t = 1.6764X_{t-1} - 0.8034X_{t-2} - 0.6896\varepsilon_{t-1} + 2.5623$$



ARIMA(4,0,0)
log likelihood = -140.53



ARIMA(2,0,1)
log likelihood = -141.57

The Conditional Mean

- All ARMA processes can be written in the form

$$X_t = \mu_t + \varepsilon_t, \quad t \in \mathbb{Z},$$

where, for example, $\mu_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$ in the general ARMA process (1).

- Let $\mathcal{F}_t$ denote the history (or information content) of the process up to and including time t — in mathematical language the filtration generated by $(X_s)_{s \le t}$.
- Then $\mathbb{E}(X_t \mid \mathcal{F}_{t-1}) = \mathbb{E}(\mu_t + \varepsilon_t \mid \mathcal{F}_{t-1}) = \mu_t$, since $\mathbb{E}(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$.
- ARMA models are models for the conditional mean, the expected value of tomorrow's obervation, given the history until today.
- ARCH and GARCH are models for the conditional standard deviation.

Modelling Return Series with ARCH/GARCH

- Let $(Z_t)_{t \in \mathbb{Z}}$ follow a strict white noise process with mean zero and variance one.
- ARCH and GARCH processes $(X_t)_{t \in \mathbb{Z}}$ take general form

$$X_t = \sigma_t Z_t, \quad t \in \mathbb{Z}, \tag{4}$$

  where σ_t, the (random) volatility, is a function of the history up to time $t-1$ represented by $\mathcal{F}_{t-1}$.
- Z_t is assumed independent of $\mathcal{F}_{t-1}$.
- Mathematically, σ_t is $\mathcal{F}_{t-1}$-measurable, where $\mathcal{F}_{t-1}$ is the filtration generated by $(X_s)_{s \le t-1}$, and therefore

$$\mathrm{var}(X_t \mid \mathcal{F}_{t-1}) = \mathrm{var}(\sigma_t Z_t \mid \mathcal{F}_{t-1}) = \sigma_t^2 \mathrm{var}(Z_t \mid \mathcal{F}_{t-1}) = \sigma_t^2,$$

  using the measurability of σ_t w.r.t. to $\mathcal{F}_{t-1}$.
- Hence, volatility is the conditional standard deviation of the process.

ARCH and GARCH Processes

ARCH = AutoRegressive Conditional Heteroskedasticity

- (X_t) follows an ARCH(p) process if, for all t,

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2, \quad \alpha_j > 0.$$

  Intuition: volatility is influenced by large observations in recent past.

- (X_t) follows a GARCH(p,q) process (generalised ARCH) if, for all t,

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2, \quad \alpha_j, \beta_k > 0. \tag{5}$$

  Intuition: more persistence is built into the volatility.

Stationarity and Autocorrelations

- The condition for the GARCH equations to define a covariance stationary process with finite variance is that

$$\sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k < 1.$$

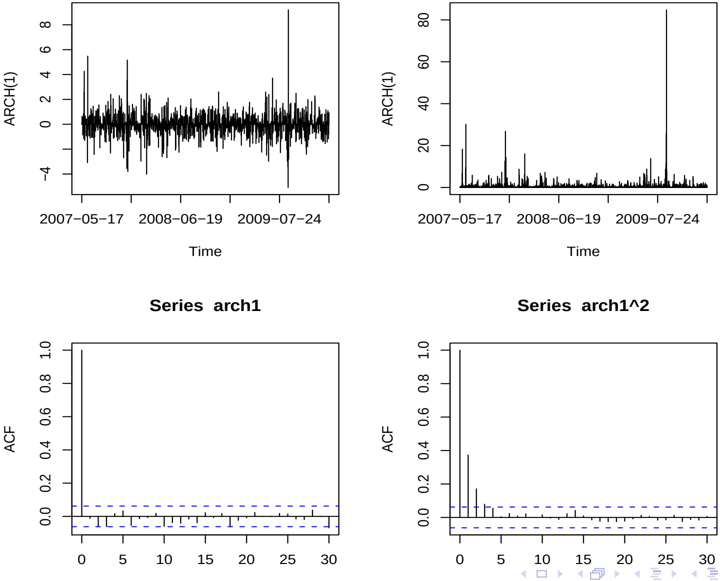
- ARCH and GARCH are technically white noise processes since

$$\begin{aligned} \gamma(h) &= \mathrm{cov}(X_t, X_{t+h}) \\ &= \mathbb{E}(\sigma_{t+h} Z_{t+h} \sigma_t Z_t) - \mathbb{E}(\sigma_{t+h} Z_{t+h}) \mathbb{E}(\sigma_t Z_t) \\ &= \mathbb{E}(Z_{t+h}) \mathbb{E}(\sigma_{t+h} \sigma_t Z_t) = 0. \end{aligned}$$

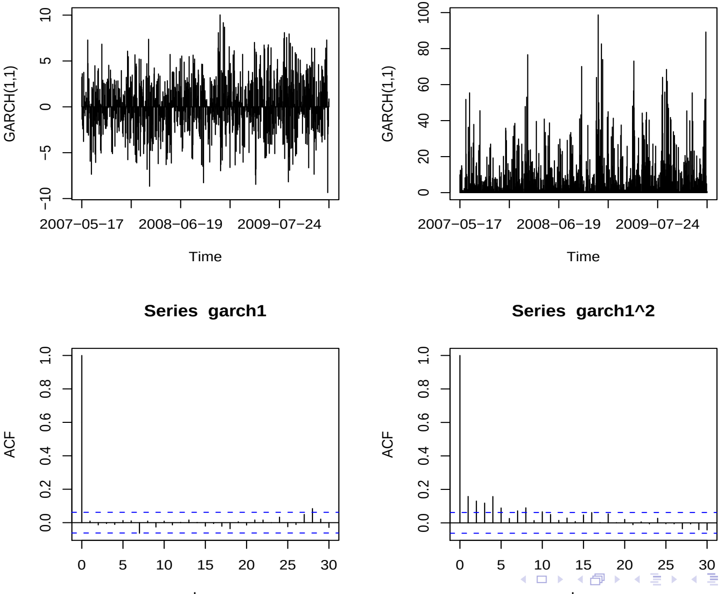
Absolute and Squared GARCH Processes

- Although (X_t) is an uncorrelated process, it can be shown that the processes (X_t^2) and $(|X_i|)$ possess profound serial dependence.
- In fact (X_t^2) can be shown to have a kind of ARMA-like structure.
- A GARCH(1,1) model is like an ARMA(1,1) model for (X_t^2).

GARCH Simulated Example I



GARCH Simulated Example II



Hybrid ARMA/GARCH Processes

- Although changes in volatility are the most obvious feature of financial return series, there is sometimes some evidence of serial correlation at small lags. This can be modelled by

$$\begin{aligned} X_t &= \mu_t + \varepsilon_t, \\ \varepsilon_t &= \sigma_t Z_t, \end{aligned} \tag{6}$$

  where μ_t follows an ARMA specification, σ_t follows a GARCH specification, and (Z_t) is a zero mean, variance 1 strict white noise.

- μ_t and σ_t are respectively the conditional mean and standard deviation of X_t given history to time $t-1$; they satisfy

$$\mathbb{E}(X_t \mid \mathcal{F}_{t-1}) = \mu_t, \quad \mathrm{var}(X_t \mid \mathcal{F}_{t-1}) = \sigma_t^2.$$

A Simple Effective Model: AR(1)+GARCH(1,1)

- The following model often suffices in practice:

$$\begin{aligned} \mu_t &= c + \phi(X_{t-1} - c)\,, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 (X_{t-1} - \mu_{t-1})^2 + \beta\sigma_{t-1}^2, \end{aligned} \qquad (7)$$

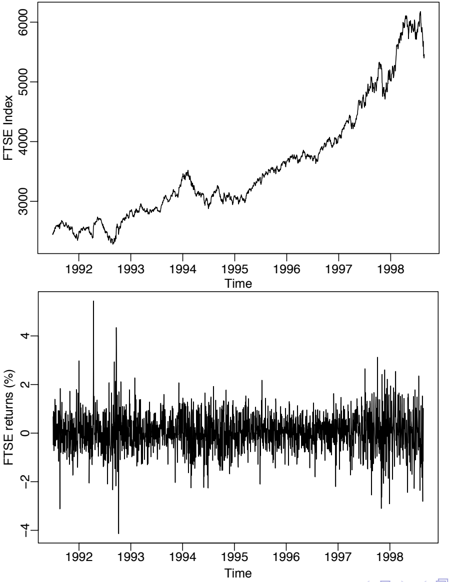
  with α_0, α_1, $\beta > 0$, $\alpha_1 + \beta < 1$ and $|\phi| < 1$ for a stationary model with finite variance.
- This model is a reasonable fit for many daily financial return series, particularly under the assumption that the driving innovations are heavier-tailed than normal.

Fitting GARCH Models to Financial Data

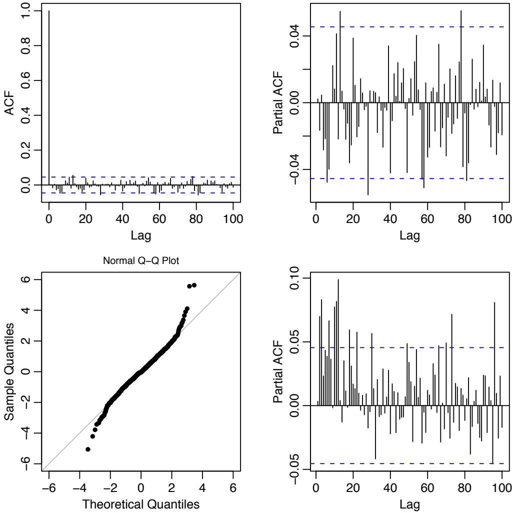
- There are a number of possible fitting methods, but the most common is maximum likelihood (ML), which is also a standard method of fitting ARMA processes to data.
- **Possibilities:**
  - Assume (Z_t) are standard iid normal innovations and estimate GARCH parameters (α_j and β_k) by ML.
  - Assume (Z_t) are (scaled) Student t_ν innovations and estimate GARCH parameters plus ν by ML.
  - Make no distributional assumptions and estimate GARCH parameters by quasi maximum likelihood (QML). (Effectively uses Gaussian ML but calculates standard errors differently.)

Example: FTSE returns

The Financial Times Stock Exchange Index, 1991–1998

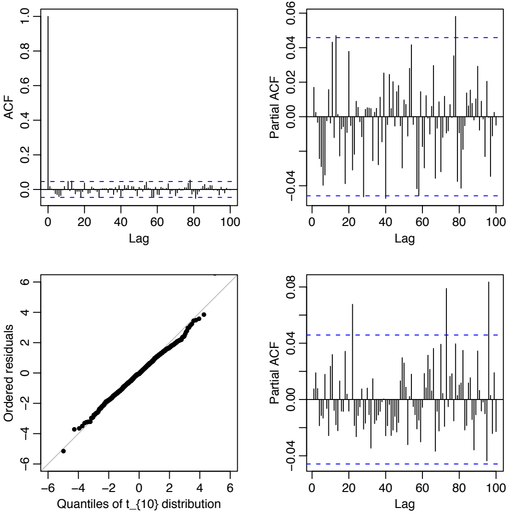


FTSE data: Fit of AR(1)-ARCH(1) model



Residuals show signs of structure, including non-normality, so try fitting an AR(1)-GARCH(1,1) with scaled t_ν residuals.

FTSE data: Fit of AR(1)-GARCH(1,1), t_ν model



Residuals show some possible signs of structure, but overall fit is much better, and the data seem to be white noise (next).

FTSE data: Fitted model

- The model is fitted by maximum likelihood and gives

$$\begin{aligned} Y_t - 0.051_{0.018} &= 0.070_{0.024}(Y_{t-1} - \mu) + \sigma_t \varepsilon_t, \\ \sigma_t^2 &= 0.006_{0.004} + 0.036_{0.011}(Y_{t-1} - \mu)^2 + 0.955_{0.016}\sigma_{t-1}^2, \\ \varepsilon_t &\overset{\text{iid}}{\sim} \sqrt{7.7/9.7}\, t_{9.7_{1.9}} \end{aligned}$$

- Comments:
  - the mean and autocorrelation are significantly positive but small, so not very useful for prediction;
  - dependence between successive σ_t^2 accounts for most of the changes in variance, with some being added by the $(Y_{t-1} - \mu)^2$;
  - the usual t_ν density has variance $\nu/(\nu - 2)$, but here is scaled to have unit variance.
- More about the innovation distribution:
  - other heavy-tailed innovation distributions could be used;
  - often the innovations are asymmetric.

Some extensions

- (G)ARCH models have the following weaknesses:
  - σ_t reacts identically to positive and negative shocks
  - Strong restrictions on the parameters are needed for stationarity and finite variance
  - They give no insight into why a series behaves as it does
  - They react slowly to large shocks in the returns, so may over-predict volatility
- They have been extended to integrated GARCH (IGARCH) models, for example

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) Y_{t-1}^2,$$

which is analogous to an ARIMA model, in that past volatility shocks persist

- The exponential GARCH (EGARCH) model allows the variance to depend on the sign of the series, for example giving

$$\sigma_t^2 = \sigma_{t-1}^{2\alpha} \exp\left(\alpha_* + (\gamma + \theta \mathrm{sign}(Y_{t-1})) \frac{|Y_{t-1}|}{\sigma_{t-1}}\right)$$

for suitable constants α, α_*, γ, θ; we expect that $\theta < 0$ if negative shocks have higher impacts

Time series dependence

Given two time series data $(X_t)_{t=1}^n$ and $(Y_t)_{t=1}^n$ we calculate the sample cross-covariances

$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_t - \overline{X})(Y_{t+h} - \overline{Y}), \quad h = 0, 1, 2, \dots$$

$$\widehat{\gamma}(-h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h} - \overline{X})(Y_t - \overline{Y}), \quad h \geq 0 \quad \text{(symetrically)}$$

$$\text{where} \quad \overline{X} = \sum_{t=1}^{n}\frac{X_t}{n} \quad \text{and} \quad \overline{Y} = \sum_{t=1}^{n}\frac{Y_t}{n}.$$

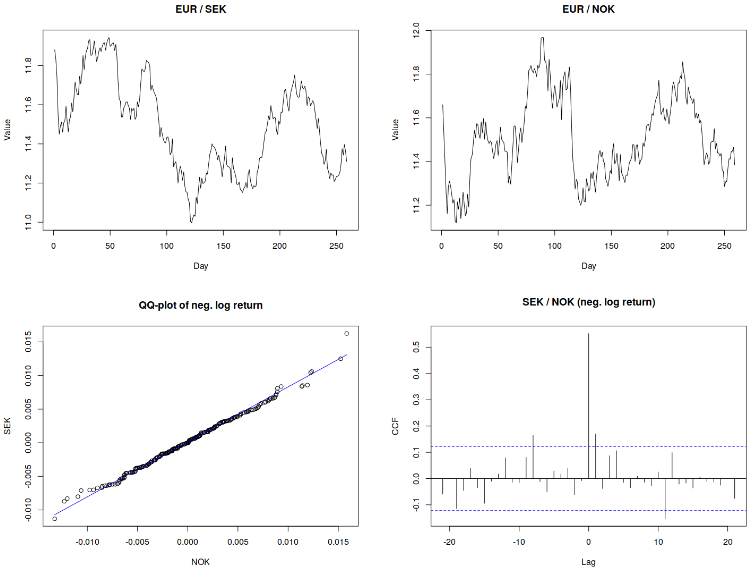
and sample cross-correlations

$$\widehat{\rho}(h) = \widehat{\gamma}(h)/\widehat{\gamma}(0), \quad h \in \mathbb{Z},\ -n < h < n$$

The cross-correlogram (CCF) is the plot

$$\{(h, \widehat{\rho}(h)),\ h \in \mathbb{Z},\ -n < h < n\}$$

Example



Exchange rate of SEK and NOK against EUR (June 2023 - June 2024)

Granger Causality Test

The Granger Causality test is used to examine if one time series may be used to forecast another.

- Consider two time series $(X_t)_{t=1}^{n}$ and $(Y_t)_{t=1}^{n}$.
- Knowing the value of a time series (X_t) at a given lag is valuable for forecasting the value of a time series (Y_t) at a later time period is referred to as Granger-causes.
- The Granger Causality Test tests the following H_0 hypothesis:

  Time series (X_t) does not Granger-cause time series (Y_t).
- Alternative Hypothesis (H_1):

  Time series (X_t) Granger-causes time series (Y_t).

Granger Causality Test

Mathematically

$$Y_t = c + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_t \tag{8}$$

$$Y_t = c + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j X_{t-j} + \varepsilon_t \tag{9}$$

H_0 : no β_j is significant in the model
The null hypothesis that (X_t) does not Granger-cause (Y_t) is rejected if and only if any lagged values of (X_t) is retained in the regression.

Intuition

A time series (X_t) Granger-causes another time series (Y_t) if predictions of the value of (Y_t) based on its own past values and on the past values of (X_t) are better than predictions of (Y_t) based only on (Y_t) own past values.

Example

```
Granger causality test

Model 1: sek_lr ~ Lags(sek_lr, 1:1) + Lags(nok_lr, 1:1)
Model 2: sek_lr ~ Lags(sek_lr, 1:1)
  Res.Df Df      F   Pr(>F)
1    254
2    255 -1 8.4557 0.003961 **
```

Model $\quad \texttt{sek\_lr}_t = \mu + \alpha_1 \, \texttt{sek\_lr}_{t-1} + \beta_1 \, \texttt{nok\_lr}_{t-1}$

$\mu \cong 0$, $\alpha_1 = -0.0747_{0.0631}$, $\beta_1 = 0.4513_{0.0420}$