Extreme value statistics for financial risk

Lecture 1: Probability theory

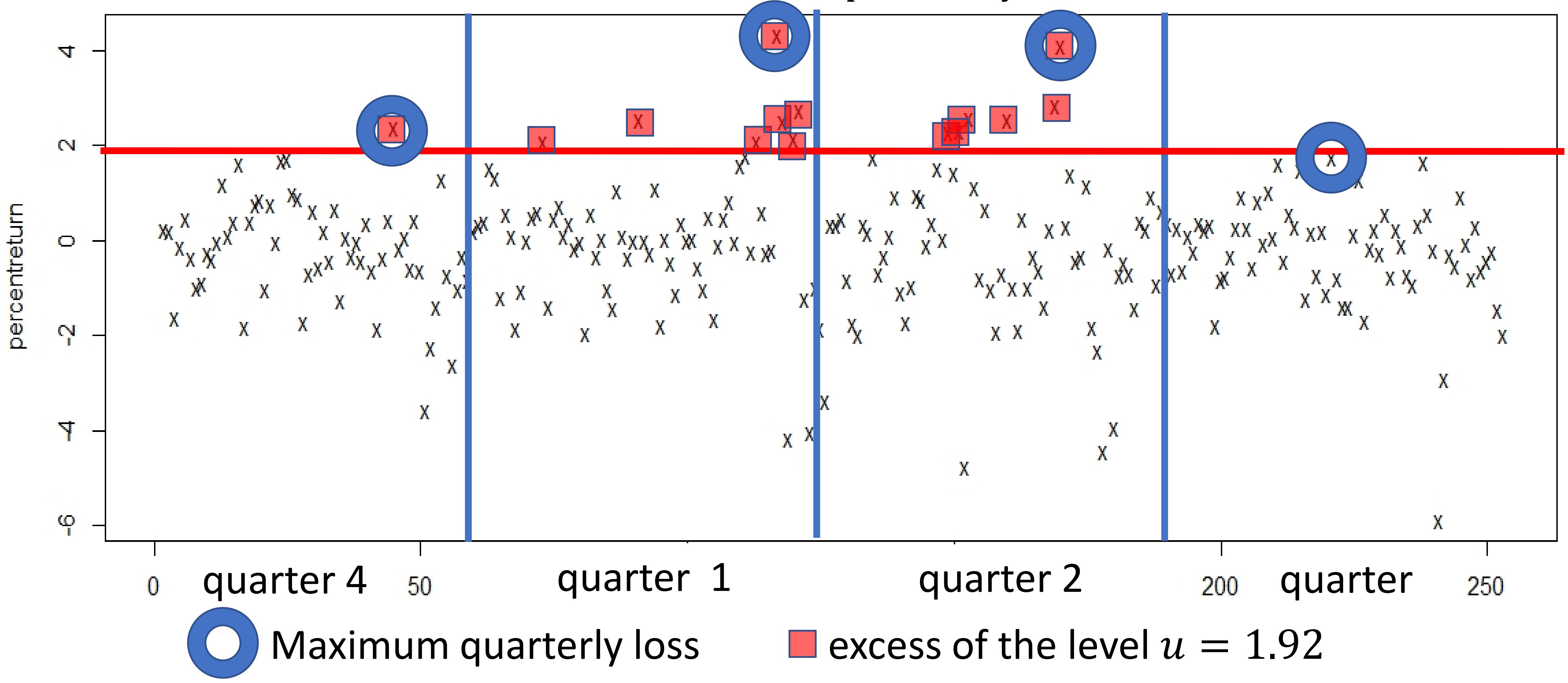
Lecture 2: Block maxima + PoT

Lecture 3: PoT + program packages

Lecture 4: Programming + multivariate block maxima

Lecture 5: Multivariate PoT

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

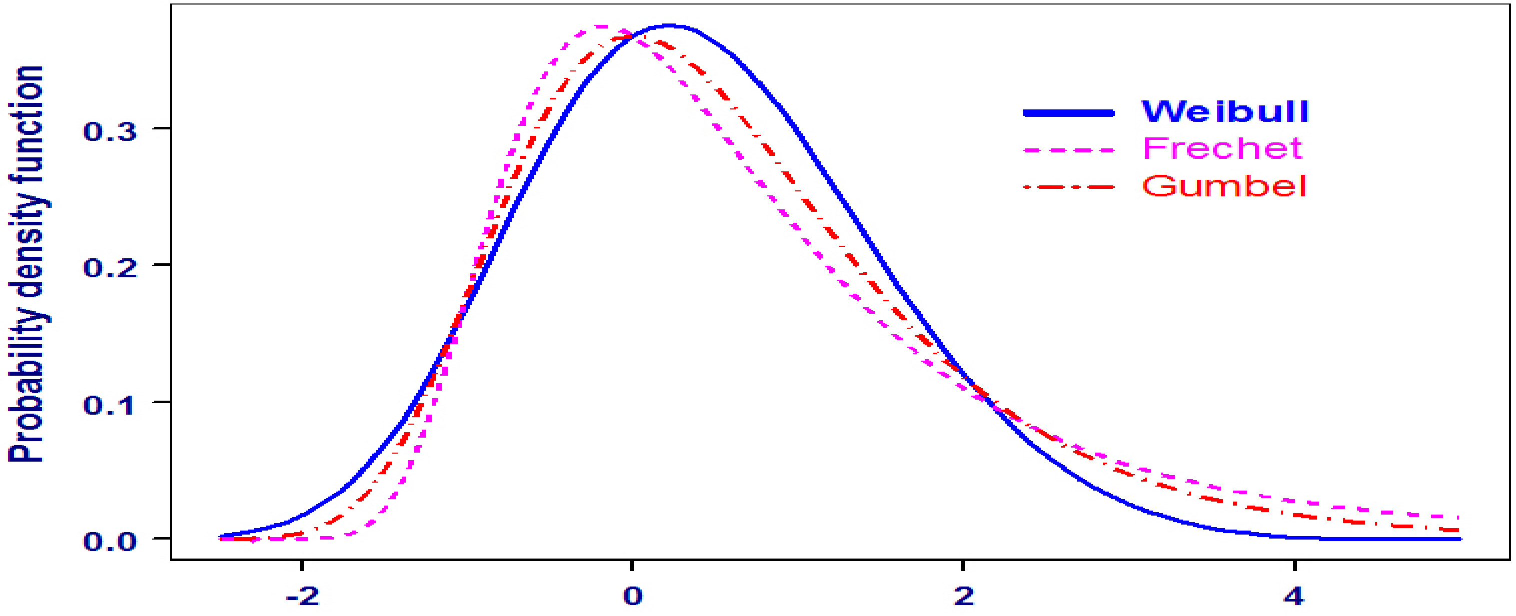


How large is the risk of a big quarterly loss?

How large is the risk of a big loss tomorrow?

Generalized extreme value (GEV) distributions

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



$\gamma > 0$ Frechet distribution, finite left endpoint $x > \mu + \sigma/\gamma$, heavytailed

$\gamma = 0$ Gumbel distribution, $G(x) = \exp\{-e^{-(x-\gamma)/\sigma}\}$, unbounded

$\gamma < 0$ Weibull distribution, finite right endpoint $x < \mu + \sigma/|\gamma|$

The block maxima method

Obtain observations $X_1, \dots, X_n$ of block maxima (e.g. weekly or yearly maxima)

- Assume the observations are i.i.d and follow a GEV distribution
- Use $X_1, \dots, X_n$ to estimate the parameters of the GEV distribution
- Use the fitted GEV to compute estimates and confidence intervals for quantities of interest, e.g. quantiles of the distribution of the 50-year maximum

- Many estimation methods: ML estimators, moment estimators, probability weighted moments estimators, biascorrected estimators, …
- **ML has the major advantage that it gives standardized ways for including trends in parameters and for testing of submodels – and for handling truncation and censoring**
- Profile likelihood or bootstrap confidence intervals often preferable

(A perhaps unnecessary explanation) **What does**

$$P(\frac{M_n - a_n}{b_n} \leq x) \to G(x), \quad \text{as} \quad n \to \infty \quad \text{for all} \quad x,$$

mean in practice? That $P(\frac{M_n - a_n}{b_n} \leq x) \approx G(x)$, for large n, or,

with $y = b_n x + a_n$ and $G(x) = \exp\{-(1 + \gamma \frac{x - \mu'}{\sigma'}))^{-1/\gamma}\}$, that

$$\begin{aligned} P(M_n \leq y) &\approx G(\frac{y - a_n}{b_n}) = \exp\{-(1 + \gamma \frac{y - (a_n + b_n \mu')}{b_n \sigma'})^{-1/\gamma}\} \\ &= \exp\{-(1 + \gamma \frac{y - \mu}{\sigma})^{-1/\gamma}\}, \text{ for } \mu = a_n + b_n \mu', \ \sigma = b_n \sigma. \end{aligned}$$

Since all the parameters are unknown anyway, we are left with the problem of estimating $\mu, \ \sigma$ from data, *i.e.* with the Block Maxima method.

Maximum Likelihood (ML) inference

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

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

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

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

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

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

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

ML inference: asymptotic properties

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

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

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

$k_{a/2}$ = the $\alpha/2$-th quantile from the top of the standard normal distribution

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

asymptotic 100(1- α) % confidence interval for θ_i

ML inference: the delta method

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

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

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

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

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

Works well if g is approximatly linear, not so well otherwise. Alternative: simulate from limiting normal distribution.

ML inference: Likelihood Ratio (LR) tests

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

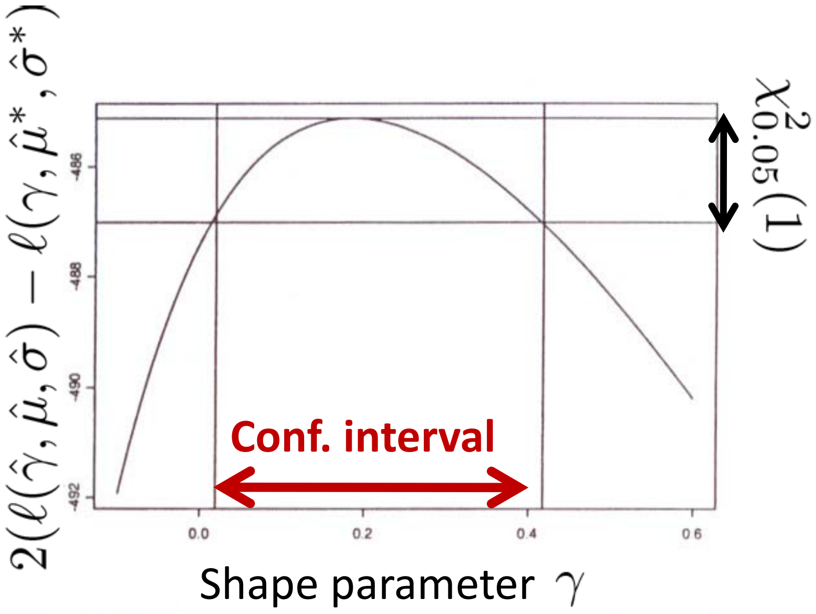
$2(\ell(\hat{\theta}) - \ell(\theta_1, \hat{\theta}_2^*))$ asymptotically has a χ^2 distribution with *d*-*p* degrees of freedom if θ_1 is the true value → LR test:

Reject $\mathrm{H}_0 : \theta_1 = \theta_1^0$ at the significance level 100 α% if

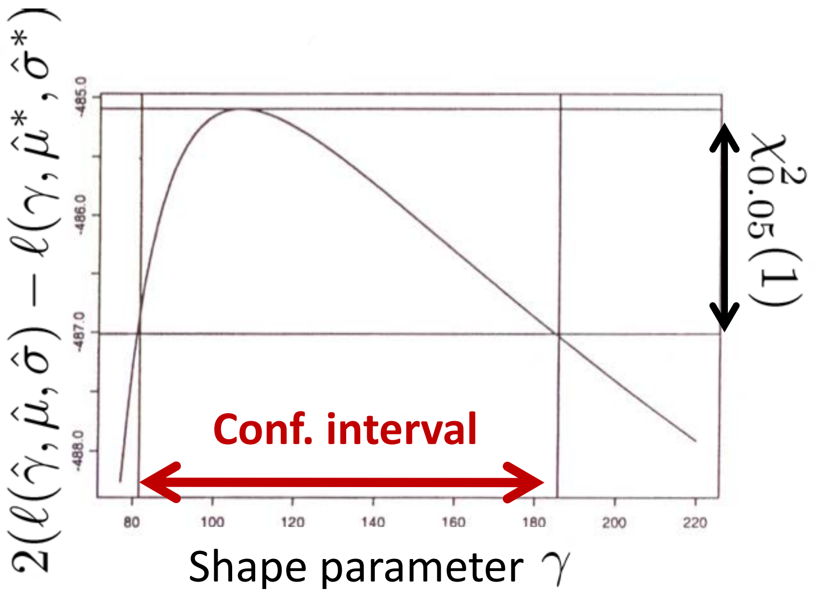
$2\left(l(\hat{\theta}) - l(\theta_1^0, \hat{\theta}_2^*)\right) > \chi_\alpha^2(d-p)$, where $\chi_\alpha^2(d-p)$ is the α-th quantile from the top of the χ^2 distribution with *d*-*p* degrees of freedom

ML inference: profile likelihood confidence intervals

(often more accurate than delta method intervals)



Profile likelihood confidence interval for γ

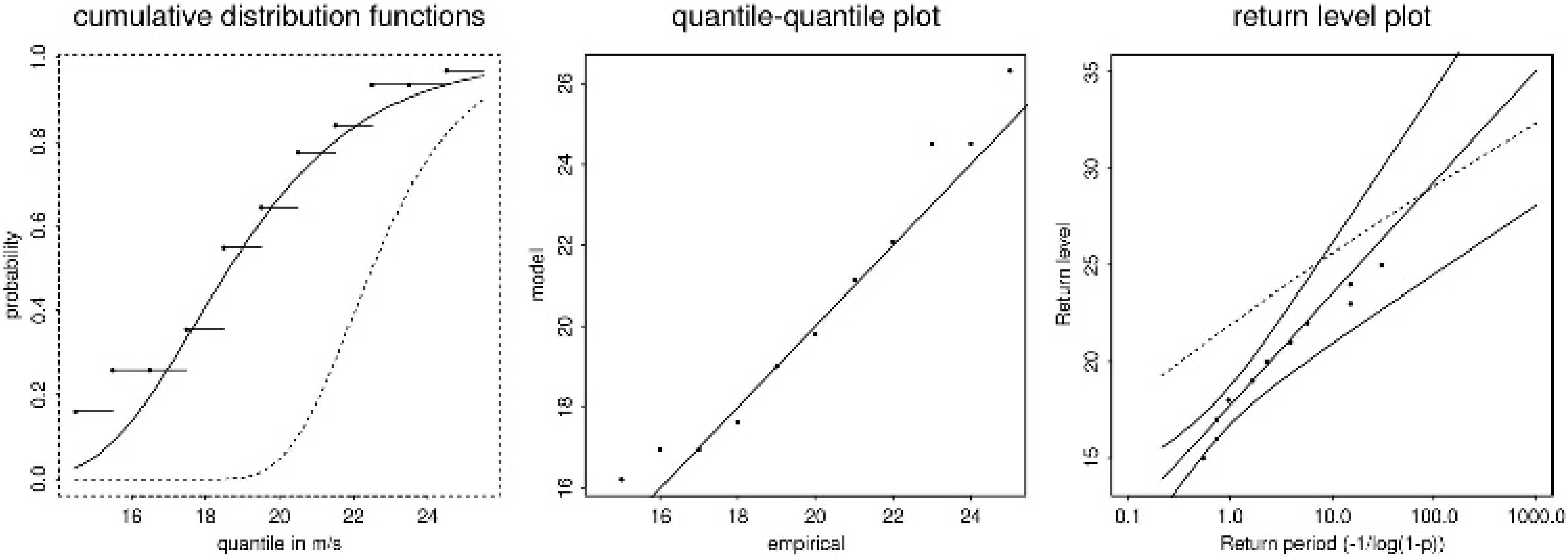


Profile likelihood confidence interval for γ

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

Sometimes standard distributions, Gaussian, Weibull, ... are fitted to all available data (not just to block maxima or to peaks over thresholds) and the results then is used to derive distributions of extremes.

It is a very bad idea to do this without checking that the extrapolation to extremes is valid. (Led to serious underestimation of risk in the early years of VaR)



Yearly maximum 10 minute average windspeeds 1961–1990 at Barkåkra, Sweden. Solid lines: estimated GEV distribution of yearly maxima. Dotted lines: estimated Weibull distribution obtained by using all measurements (the Weibull fit was very good in the center of data)

- *Data:* 10 minute average wind speed at the start of each hour during 1961–1990, from 12 synoptic meteorological stations in Sweden
- *Aim:* Investigate methods to estimate the $1/50$ or $1/100$ upper quantiles of yearly maximum windspeed, as contribution to development of Swedish wind standards
- *Analysis:* Block maxima method with block $=$ year. **Maximum likelihood estimation which took rounding of wind speeds to whole knots into account.** Separate analysis for each station
- *Conclusions: U*se Block maxima method, not Weibull fitting; if rounding is neglected then estimation uncertainty is underestimated

Perrin et al, Theor. Appl. Climatol. 2006