Measuring Statistical Risk
Extremes, Joint Extremes and Copulae

Wolfgang Härdle
(with S. Borak, 陈颖, B. Choroś)

Institut für Statistik and Ökonometrie
CASE - Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin

Risk Management

Risikomanagement

Zarządzanie Ryzykiem

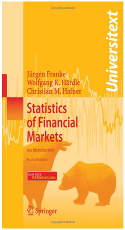
Řízení Rizika

风险管理

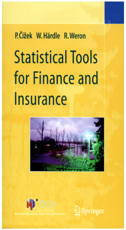
危険と管理

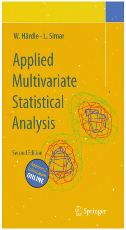
ادارة المخاطرة

위험관리


Universitext
Jürgen Franke
Wolfgang K. Härdle
Christian M. Hafner
Statistics of Financial Markets
An Introduction
Second Edition
Springer


Wolfgang Härdle
Nikolaus Hautsch
Ludger Overbeck
Editors
Applied Quantitative Finance
Second Edition
Springer


P. Čížek W. Härdle R. Weron
Statistical Tools for Finance and Insurance
Springer


W. Härdle · L. Simar
Applied Multivariate Statistical Analysis
Second Edition
ONLINE
Springer

- **Franke, J., Härdle, W., and Hafner, C.** (2008) Statistics of Financial Markets: an Introduction. 2nd extended ed., Springer Verlag, Heidelberg.
- **Härdle, W., Hautsch, N. and Overbeck, L.** (2009) Applied Quantitative Finance. 2nd extended ed., Springer Verlag, Heidelberg.
- **Cizek, P., Härdle, W. and Weron, R.** (2005) Statistical Tools in Finance and Insurance Springer Verlag, Heidelberg.
- **Härdle, W., Simar, L.** (2007) Applied Multivariate Statistical Analysis. 2nd extended ed., Springer Verlag, Heidelberg.

Fluctuat nec mergitur

AHW Investmentfonds

WKN: 930681 ISIN: LU 0108749473

Wertpapiergattungen
im AHW Top-Dividend Low-5 International

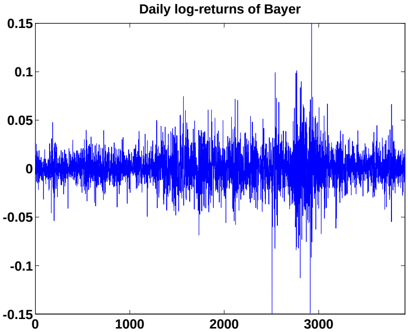
Aktie	Index		Dividendenrendite p.a.
ThyssenKrupp	DAX	3,86 %	3,82 %
TUI	DAX	4,53 %	3,76 %
Volkswagen AG	DAX	3,99 %	2,91 %
Deutsche Telekom	DAX	4,15 %	4,01 %
Daimler Chrysler AG	DAX	4,18 %	4,36 %
MAN AG	DAX	4,16 %	3,08 %
Credit Agricole	CAC-40	3,80 %	3,10 %
DEXIA SA	CAC-40	4,25 %	3,42 %
AGF	CAC-40	4,46 %	4,31 %
AXA	CAC-40	4,22 %	2,96 %
Arcelor	CAC-40	3,87 %	3,72 %
BNP	CAC-40	4,22 %	3,64 %
ABN Amro	AEX	4,09 %	5,19 %
Ing Groep NV	AEX	4,23 %	4,59 %
Fortis	AEX	4,37 %	7,11 %
Reed Elsevier NV	AEX	3,03 %	2,86 %
Aegon	AEX	4,00 %	4,06 %
Wolters Kluwer	AEX	3,51 %	3,90 %
KPN	AEX	4,17 %	5,07 %
General Electric	Dow-Jones	3,75 %	2,47 %
JP Morgan Chase	Dow-Jones	3,54 %	3,93 %
Verizon Communications	Dow-Jones	4,19 %	4,65 %
Merck & Co INC	Dow-Jones	4,30 %	4,75 %
Pfizer INC	Dow-Jones	4,06 %	2,95 %
SBC Communications INC	Dow-Jones	4,06 %	5,53 %
Barreserve		- 0,99 %	

Aktueller Zinssatz der Europäischen Zentralbank: **2,00 %**

Stand: 30.03.2005
Quellen: Dt. Bundesbank, Thomson Financial Datastream, Bloomberg.

Example

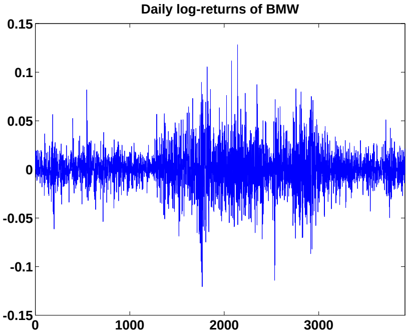
Daily log-returns of the German stock Bayer from 1992-01-01 to 2006-12-29. (+)



MSRbayer_log_returns

Example

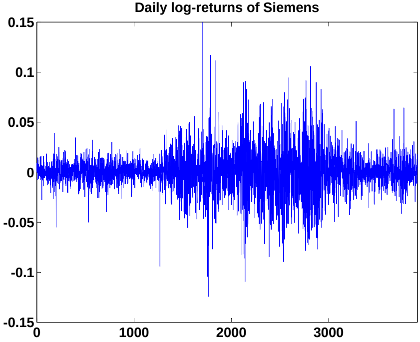
Daily log-returns of the German stock BMW from 1992-01-01 to 2006-12-29.



MSRbmw_log_returns

Example

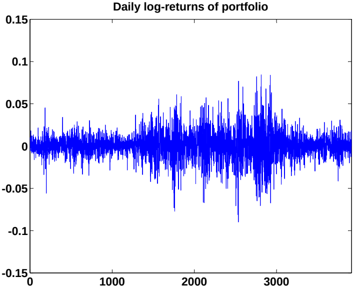
Daily log-returns of the German stock Siemens from 1992-01-01 to 2006-12-29.



MSRsiemens_log_returns

Example

Daily log-returns of the German stock portfolio (Bayer, BMW and Siemens) with trading strategy $b^\top = (1,1,1)$ from 1992-01-01 to 2006-12-29.



MSRportfolio_log_returns

Extreme Value

Extremwert

Wartość Ekstremalna

Krajní Hodnota

极值

極值

القيمة الحدية

극단값

Statistics of Extreme Risks

Stylized facts in financial markets

- Returns are heavy tailed distributed
- Volatility changes stochastically
- GARCH model yields fat tails but often underestimates for $q \geq 95\%$.

Model structure of the daily log-returns series of portfolio (Bayer, BMW, Siemens) from 1992-01-01 to 2006-12-29.

$$X_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{ iid } N(0,1)$$
$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$$

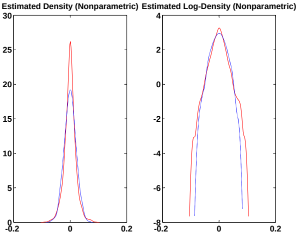


Figure 1: Kernel (normal kernel, bandwidth given by Silverman's rule-of-thumb) density (left) and log-density (right) estimate for portfolio's GARCH innovations (red) and normal variables (blue).

MSRportfolio_est

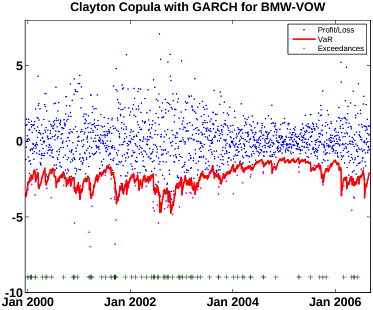


Figure 2: Backtesting results for Value-at-Risk estimation at 0.05 level for BMW and Volkswagen with Clayton copula. Time period: 1st January 1999 – 1st September 2006, size of moving window 250, Monte Carlo samples of 10.000 realizations of pseudo random variable. Margins modeled with GARCH with normal innovations. Exceedances ratio $\widehat{\alpha} = 0.0474$.

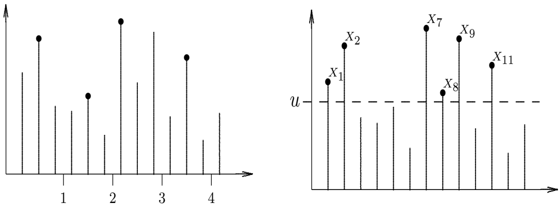
MSRvar_cop_GARCH_backtesting

Extreme value distributions

- yield more precise approximations in the tails
- probability of extreme events depends on the tail of $f(x)$ - pdf of ε_t
- apply methods of extreme value statistics to estimate "extreme" quantiles

Identifying extreme events

- Maxima (block maxima) taking in successive periods
- Peaks over threshold (POT): loss exceeds a given (high) threshold u.



The limits of maxima

$X_1, \ldots, X_n$ are i.i.d. with cdf $F(x)$

$$M_n = \max(X_1, \ldots, X_n)$$

Example

Let us consider a random variable which represent daily losses or returns: $X_t = -Z_t$ at day t where Z_t is the Profit & Loss random variable. Here we take a maximum for every consecutive block of $n = 3$ observations.

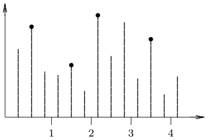


Figure 3: Block-maxima

One may easily compute the cdf of maxima: One may easily compute the cdf of maxima:

$$P(M_n \leq x) = P(X_1 \leq x, \ldots, X_n \leq x) = F^n(x). \qquad (1)$$

For unbounded random variables, (i.e. $F(x) < 1, \forall x < \infty$):

$$F^n(x) \to 0$$

hence

$$M_n \xrightarrow{P} \infty$$

Problem: The maximum of n unbounded random variables may become arbitrary large. For an analysis of asymptotics one needs the limit law for the block maxima M_n.

Definition (Maximum Domain of Attraction)

The random variables X_t belong to the maximum domain of attraction (MDA) of the nondegenerated distribution G, if there exist constants $c_n > 0$ and d_n such that:

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathcal{L}} G \quad \text{for} \quad n \to \infty,$$

i.e. $F^n(c_n x + d_n) \to G(x)$ for all points of continuity x of the cdf $G(x)$.

Extreme value distribution

Distribution G in the above Definition is called an *extreme value (EV) distribution*.

Three standard extreme value distributions:

$$\begin{aligned}
\textit{Fréchet:} \quad G_{1,\alpha}(x) &= \exp\{-x^{-\alpha}\}, \; x \geq 0 \;, \alpha > 0, & (2)\\
\textit{Gumbel:} \quad G_0(x) &= \exp\{-e^{-x}\}, \; x \in \mathbb{R}, & (3)\\
\textit{Weibull:} \quad G_{2,\alpha}(x) &= \exp\{-|x|^{-\alpha}\}, \; x \leq 0 \;, \alpha < 0. & (4)
\end{aligned}$$

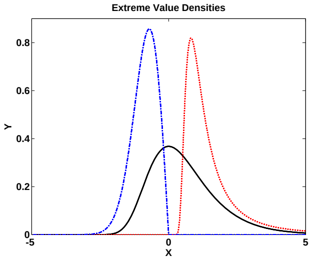


Figure 4: Fréchet (red), Gumbel (black) and Weibull (blue) probability density functions.

MSRevt1

Jenkinson and von Mises suggested a parametric representation for the three standard distributions:

Definition (Generalized Extreme Value)

The generalized extreme value distribution (GEV) with the shape parameter $\gamma \in \mathbb{R}$ has the following cdf:

$$G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\},\ 1+\gamma x > 0 \ \text{ for } \gamma \neq 0$$
$$G_0(x) = \exp\{-e^{-x}\},\ x \in \mathbb{R}$$

Gumbel	G_0
Fréchet	$G_\gamma(\frac{x-1}{\gamma}) = G_{1,1/\gamma}(x)$ for $\gamma > 0$
Weibull	$G_\gamma(-\frac{x+1}{\gamma}) = G_{2,-1/\gamma}(x)$ for $\gamma < 0$.

R. von Mises on BBI: 

Theorem (Fisher and Tippett (1928) Theorem)

If there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ and some non-degenerated distribution function G such that

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathcal{L}} G \quad \text{for } n \to \infty,$$

then G is a GEV distribution.

Assume that we have a large enough block of n iid random variables and set $y = c_n x + d_n$, then $P(M_n \leq y) \approx G_\gamma(\frac{y-d_n}{c_n})$.

R. Fisher on BBI: 

Lemma (Convergence-Type Theorem)

Let $U_1, U_2, \ldots, V, W$ be random variables, $b_n, \beta_n > 0$, $a_n, \alpha_n \in \mathbb{R}$. If

$$\frac{U_n - a_n}{b_n} \xrightarrow{\mathcal{L}} V$$

in distribution for $n \to \infty$, then the following statement holds:

$$\frac{U_n - \alpha_n}{\beta_n} \xrightarrow{\mathcal{L}} W \qquad \textit{iff} \qquad \frac{b_n}{\beta_n} \longrightarrow b \geq 0, \; \frac{a_n - \alpha_n}{\beta_n} \longrightarrow a \in \mathbb{R}.$$

In this case W follows the same distribution as $bV + a$.

Notice that for all $n \geq 1$ the maximum M_n of n iid random variables $X_1, ..., X_n$ has the same distribution as $c_n X_1 + d_n$ given suitable constants $c_n > 0$ and d_n.

Properties of GEV

- In general we can change the center and the scale to obtain other GEV distributions:

$$G(x) = G_\gamma\left(\frac{x-\mu}{\sigma}\right)$$

with the shape parameter γ, the location parameter μ and the scale parameter $\sigma > 0$.

Properties of GEV

- GEV distributions are characterized by their max-stability. A probability density function F is max-stable if

$$F^n(d_n + c_n x) = F(x)$$

for a suitable choice of constants d_n and $c_n > 0$. For example, the maximum M_n of n iid random variables X_i has the same distribution as $c_n X_1 + d_n$ given suitable constants $c_n > 0$ and d_n.

For an exploratory data analysis one checks the graphs:

$$
\begin{array}{rl}
\text{PP-plot} & \left\{F(X^{(k)}),\ \frac{n-k+1}{n+1}\right\}_{k=1}^{n}, \\
\text{QQ-plot} & \left\{X^{(k)},\ F^{-1}\left(\frac{n-k+1}{n+1}\right)\right\}_{k=1}^{n}, \\
\text{mean excess-plot} & \left\{X^{(k)}, e_n(X^{(k)})\right\}_{k=1}^{n}.
\end{array}
$$

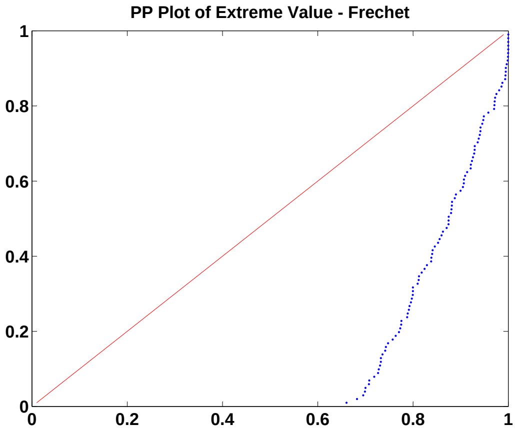


Figure 5: Normal PP plot of the pseudo random variables with Frechét distribution, see (2), with $\alpha = 2$. MSRevt2

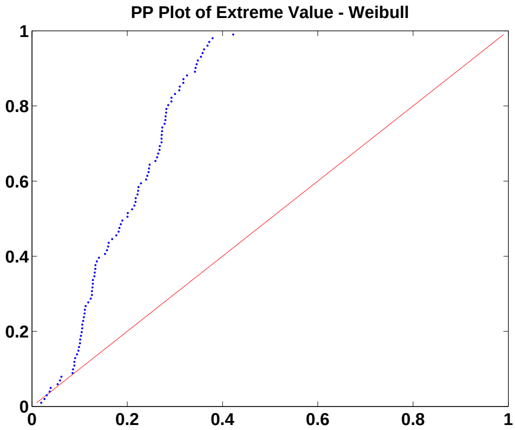


Figure 6: Normal PP plot of the pseudo random variables with Weibull distribution, see (4), with $\alpha = -2$. MSRevt2

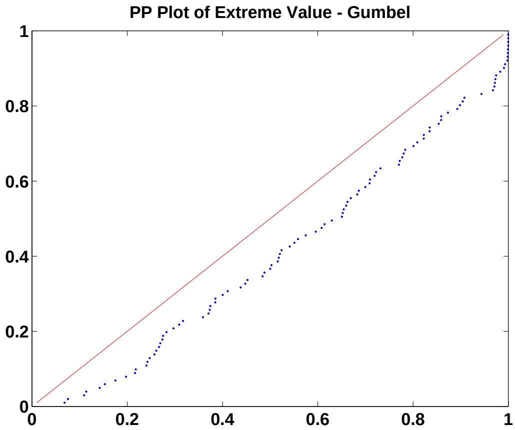


Figure 7: Normal PP plot of the pseudo random variables with Gumbel distribution, see (3). MSRevt2

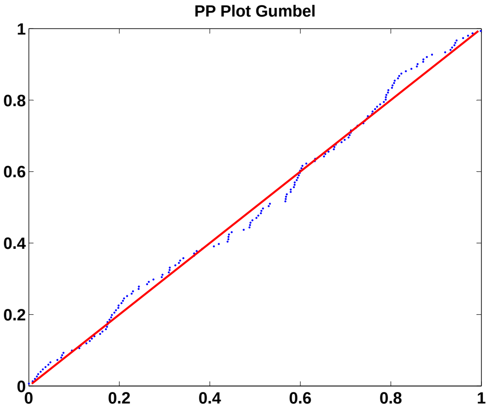


Figure 8: PP plot of the pseudo random variables with Gumbel distribution against theoretical Gumbel distribution, see (3). MSRevt3

Identifying the type of the limit (GEV) distributions

The deciding factor is how fast the probability for extremely large observations decreases beyond a threshold x, when x increases. It depends obviously on the decrease of the function:

$$\overline{F}(x) = P(X > x) = 1 - F(x)$$

for large x.

Theorem

a) For $0 \le \tau \le \infty$ *and every sequence of real numbers* $u_n, n \ge 1$*, it holds for* $n \to \infty$

$$n\overline{F}(u_n) \to \tau \quad \textit{iff} \quad P(M_n \le u_n) \to e^{-\tau}.$$

b) F *belongs to the MDA of the GEV distribution* G *with the standardized sequences* c_n, d_n*, exactly when* $n \to \infty$

$$n\overline{F}(c_n x + d_n) \to -\log G(x) \quad \textit{for all } x \in \mathbb{R}.$$

The excess probability of Fréchet $G_{1,\alpha}$ behaves as:

$$\overline{G}_{1,\alpha}(x) = \frac{1}{x^\alpha}\{1 + \mathcal{O}(1)\} \qquad \text{for} \quad x \to \infty.$$

All distributions that belong to the MDA of Fréchet $G_{1,\alpha}$ fulfill: $x^\alpha \overline{F}(x)$ is "almost constant" for $x \to \infty$ or more precisely $x^\alpha \overline{F}(x)$ is a slowly varying function.

Definition (Slowly Varying Functions)

A positive measurable function L in $(0, \infty)$ is called slowly varying, if for all $t > 0$:

$$\frac{L(tx)}{L(x)} \to 1 \qquad \text{for} \quad x \to \infty.$$

Example

$L(x) = \log(1 + x), x > 0$ is slowly varying (L'Hospital's rule).

Example

$L(x) = \log(1+x), x > 0$ is slowly varying?

$$
\begin{aligned}
\log(1+tu) &= \log t + \log(t^{-1}+u) \\
&= \log t + \log\{1+u+(t^{-1}-1)\} \\
\Delta_u &= \log\{1+u+(t^{-1}-1)\} - \log(1+u) \\
&\overset{\text{Taylor}}{=} -\frac{(t^{-1}-1)^2}{2} + \frac{(t^{-1}-1)^3}{3!} \pm \cdots \\
\text{i.e. } \{\Delta_u / \log(1+u)\} &= \mathcal{O}\{\log^{-1}(1+u)\} \\
\frac{\log(1+tu)}{\log(1+u)} &= 1 + \frac{\log t}{\log(1+u)} + \mathcal{O}\{\log^{-1}(1+u)\}
\end{aligned}
$$

Theorem (MDA of Frechét distribution)

F belongs to the MDA of the Frechét distribution $G_{1,\alpha}$, for $\alpha > 0$, if and only if $x^{\alpha}\overline{F}(x) = L(x)$ is a slowly varying function. The random variables X_t with the distribution function F are unbounded (i.e. $F(x) < 1$ for all $x < \infty$) and

$$\frac{M_n}{c_n} \xrightarrow{\mathcal{L}} G_{1,\alpha}$$

with $c_n = F^{-1}(1 - \frac{1}{n})$ or $\overline{F}(c_n) = P(X_t > c_n) = 1/n$.

M. Frechét on BBI: 

Theorem (MDA of Frechét distribution) states a criterion for obtaining the GEV Fréchet $G_{1,\alpha}$ as limit distribution.

The Weibull distribution can be obtained via the relationship $G_{2,\alpha}(-x^{-1}) = G_{1,\alpha}(x), \ x > 0$. However random variables, whose maxima are asymptotically Weibull distributed, are by all means bounded.

Therefore, in financial applications they are only interesting in special situations where using a type of hedging strategy, the loss, which results from an investment, is limited.

W. Weibull on BBI: 

Example

The Pareto distributions with cdf

$$W_{1,\alpha}(x) = 1 - \frac{1}{x^{\alpha}}, x \geq 1, \alpha > 0,$$

and all other cdfs with Pareto tails:

$$\overline{F}(x) = \frac{\kappa}{x^{\alpha}}\{1 + \mathcal{O}(1)\} \quad \text{for} \quad x \to \infty.$$

belong to the MDA of the Fréchet distribution.
In this case $\overline{F}^{-1}(q)$ for $q \approx 1$ behaves as $(\kappa/q)^{1/\alpha}$: Set $c_n = (\kappa n)^{1/\alpha}$:

$$\frac{M_n}{(\kappa n)^{1/\alpha}} \xrightarrow{\mathcal{L}} G_{1,\alpha} \quad \text{for } n \to \infty$$

Theorem (MDA of Gumbel distribution)

The cdf F of X_t belongs to the MDA of the Gumbel distribution iff there exist scaling functions $c(x), g(x) > 0$ and an absolute cts function $e(x) > 0$:
$c(x) \to c > 0,\ g(x) \to 1, e'(x) \to 0$ *for* $x \to \infty$ *s.t.* $z < \infty$*:*
$\overline{F}(x) = c(x) \exp\{-\int_z^x \frac{g(y)}{e(y)} dy\},\ z < x < \infty$. *In this case*

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathcal{L}} G_0$$

where $d_n = F^{-1}(1 - \frac{1}{n})$ *and* $c_n = e(d_n)$.

As function $e(x)$ in Theorem (MDA of Gumbel distribution) one may choose the *mean excess function:*

$$e(x) = \frac{1}{\overline{F}(x)} \int_x^\infty \overline{F}(y)\, dy, \;\; x < \infty.$$

Example

The exponential distribution has the form: $F(x) = 1 - e^{-\lambda x}, x \geq 0$. Hence $\overline{F}(x) = e^{-\lambda x}$ fulfills the assumptions of Theorem (MDA of Gumbel distribution) with

$$c(x) = 1, \; g(x) = 1, \; z = 0 \text{ and } e(x) = 1/\lambda.$$

Example

The maximum of n iid exponentially distributed random variables with the parameter λ converges to the GEV Gumbel distribution:

$$\lambda(M_n - \frac{1}{\lambda}\ \log\ n) \xrightarrow{\mathcal{L}} G_0$$

for $\quad n \to \infty.$

Example

The maximum of n iid $N(0,1)$ random variables converges to the GEV Gumbel distribution:

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathcal{L}} G_0 \quad \text{for} \quad n \to \infty$$

where

$$\begin{aligned} c_n &= (2 \log n)^{-1/2} \\ d_n &= \sqrt{2 \log n} - \frac{\log \log n + \log(4\pi)}{2\sqrt{2 \log n}}. \end{aligned}$$

Peaks-over-threshold (POT) approach

Definition (Excess over threshold)

Let $K_n(u)$ and $N(u)$ be the index set and the number of observations over the threshold u. Denote the random variables $Y_j,\ j = 1, \ldots, N(u)$, as the excesses over the threshold value u with

$$\begin{aligned} \{Y_1, \ldots, Y_{N(u)}\} &= \{X_j - u;\ j \in K_n(u)\} \\ &= \{X^{(1)} - u, \ldots, X^{(N(u))} - u\} \end{aligned}$$

Let us consider a random variable which represent daily losses or returns: $X_t = -Z_t$ at day t where Z_t is the Profit & Loss random variable.

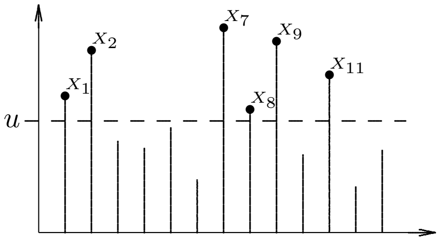


Figure 9: Excesses over a threshold u.

Definition

Let u be a threshold value and F a distribution function of some unbounded random variable X.

a) $F_u(x) = P\{X - u \leq x \mid X > u\} = \{F(u+x) - F(u)\}/\overline{F}(u),\ 0 \leq x < \infty$ is called *conditional excess distribution function* over the threshold u.

b) $e(u) = \mathbf{E}\{X - u \mid X > u\},\ \ 0 < u < \infty$ is the mean excess function.

With the integration by parts one obtains:

$$e(u) = \int_u^\infty \frac{\overline{F}(y)}{\overline{F}(u)} dy.$$

A random variable Δ_u with cdf $F_u(x)$ has expected value $\mathbf{E}\Delta_u = e(u)$.

Theorem (Pickands (1975), Balkema and de Haan (1974))

For a large class of underlying distribution functions F, the conditional excess distribution function $F_u(x)$ is well approximated by:

$$F_u(x) \approx W_{\gamma,\beta}(x) \quad u \to \infty.$$

where $W_{\gamma,\beta}(x)$ is the generalized Pareto distribution.

Definition (Generalized Pareto distribution)

The *generalized Pareto distribution* (GP) with the parameters $\beta > 0$, γ has the distribution function:

$$W_{\gamma,\beta}(x) = 1 - (1 + \frac{\gamma x}{\beta})^{-\frac{1}{\gamma}} \quad \text{for} \quad \begin{cases} x \geq 0 & \text{if} \quad \gamma > 0 \\ 0 \leq x \leq \frac{-\beta}{\gamma} & \text{if} \quad \gamma < 0, \end{cases}$$

and

$$W_{0,\beta}(x) = 1 - e^{-\frac{1}{\beta}x}, \; x \geq 0.$$

$W_\gamma(x) = W_{\gamma,1}(x)$ are called *generalized standard Pareto distributions* or *standardized GP* distributions.

Submodels of GP distribution

- Exponential (GP0): $W_0(x) = 1 - e^{-x}$, $x \geq 0$
- Pareto (GP1): $W_{1,\beta}(x) = 1 - x^{-\beta}$, $x \geq 1$ and $\beta > 0$
- Beta (GP2): $W_{2,\beta} = 1 - (-x)^{-\beta}, -1 \leq x \leq 0, \beta < 0$

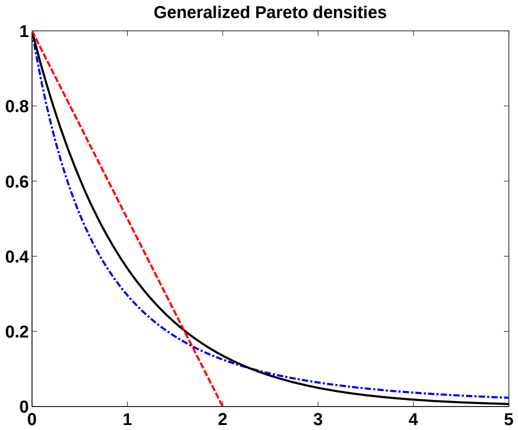


Figure 10: Generalized Pareto distributions ($\beta = 1$) with the parameters $\gamma = -0.5$ (Red), 0 (Black) and 0.5 (Blue). MSRgpdist

Theorem (Mean excess function)

Let X be a positive, unbounded rv with an absolute cts cdf F.

a) *The mean excess function $e(u)$ uniquely determines F:*

$$\overline{F}(x) = \frac{e(0)}{e(x)} \exp\left\{ -\int_0^x \frac{1}{e(u)} du \right\}, \quad x > 0.$$

b) *If F belongs to the MDA of the Fréchet distribution $G_{1,\alpha}$, then $e(u)$ is for $u \to \infty$ approximately linear i.e.: $e(u) = \frac{1}{\alpha - 1}\ u\{1 + \mathcal{O}(1)\}$.*

The generalized standard Pareto distribution is the adequate parametric distribution function for exceedances.

Theorem (MDA of GEV distribution)

The distribution F is contained in the MDA of the GEV distribution G_γ with the form parameter $\gamma \geq 0$, exactly when for a measurable function $\beta(u) > 0$ and the GP distribution $W_{\gamma,\beta}$ it holds that:

$$\sup_{x \geq 0} |F_u(x) - W_{\gamma,\beta(u)}(x)| \rightarrow 0 \textit{ for } u \rightarrow \infty.$$

A corresponding result also holds for the case when $\gamma < 0$, in which case the supremum of x must be taken for those $0 < W_{\gamma,\beta(u)}(x) < 1$.

For the generalized Pareto distribution $F = W_{\gamma,\beta}$ it holds for every finite threshold $u > 0$

$$F_u(x) = W_{\gamma,\beta+\gamma u}(x) \quad \text{for} \quad \begin{cases} x \geq 0 & \text{if} \quad \gamma \geq 0 \\ 0 \leq x < -\frac{\beta}{\gamma} - u & \text{if} \quad \gamma < 0, \end{cases}$$

In this case $\beta(u) = \beta + \gamma\, u$.

Estimation in extreme value models

Consider data $x_1, ..., x_m$ generated under a distribution function F^n. Thus each x_i is the maximum of n values that are governed by the distribution function F.

- Gumbel: $G_0(x) = exp\{-e^{-x}\}$. One may use the following two methods to estimate μ and σ of the Gumbel model: $G_{0,\mu,\sigma} = exp\{-e^{-(x-\mu)/\sigma}\}$.
  - MLE: $g_{0,\mu,\sigma} = \frac{1}{\sigma} e^{-(x-\mu)/\sigma} exp(-e^{-(x-\mu)/\sigma})$
  - Moment estimation: estimators of μ and σ are deduced from the sample mean $\overline{x}$ and variance s_n. For example, $\widehat{\sigma}_n = \sqrt{6} s_n / \pi$.

- Fréchet model: $G_{1,\alpha}(x) = exp(-x^{-\alpha})$ for $\alpha > 0$ and $x > 0$. MLE can be used. Keep in mind that the left endpoint of $G_{1,\alpha,0,\sigma}$ is always equal to 0.
- Weibull model: $G_{2,\alpha,0,\sigma}$ for $x \leq 0$, $\alpha < 0$ and $\sigma > 0$.

Estimation in Generalized Pareto Models

Let $X_i, i = 1, \ldots, n$ be the original data which are governed by a cdf F.

Notation:

$X_{(1)} \leq \ldots \leq X_{(n)}$ (increasing) order statistics
$X^{(1)} \geq \ldots \geq X^{(n)}$ (decreasing) order statistics
i.e. $X_{(1)} = X^{(n)},\ X_{(n)} = X^{(1)}$.
We deal with upper extremes which are either

- the exceedances $y_1, ..., y_m$ over a fixed threshold u, or
- the m upper ordered values $y_1, ..., y_m = X^{(1)}, ..., X^{(m)}$.

Definition (Empirical Mean Excess Function)

Let $K_n(u) = \{j \leq n;\ X_j > u\}$, $N(u) = \#K_n(u)$ and define the edf as:

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} I(X_j \leq x)$$

Define $\overline{\widehat{F}}_n = 1 - \widehat{F}_n$ and the empirical mean excess distribution:

$$\begin{aligned} e_n(u) &= \int_u^{\infty} \overline{\widehat{F}}_n(y)dy / \overline{\widehat{F}}_n(u) \\ &= \frac{1}{N(u)} \sum_{j \in K_n(u)} (X_j - u) = \frac{1}{N(u)} \sum_{j=1}^{n} \max\{(X_j - u), 0\} \end{aligned}$$

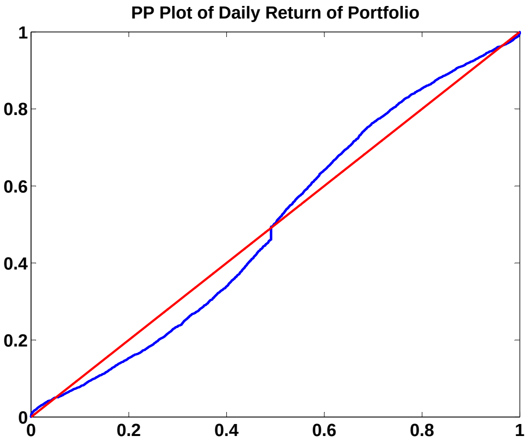


Figure 11: Normal PP plot of daily log-returns of portfolio (Bayer, BMW, Siemens) from 1992-01-01 to 2006-12-29. MSRportfolio

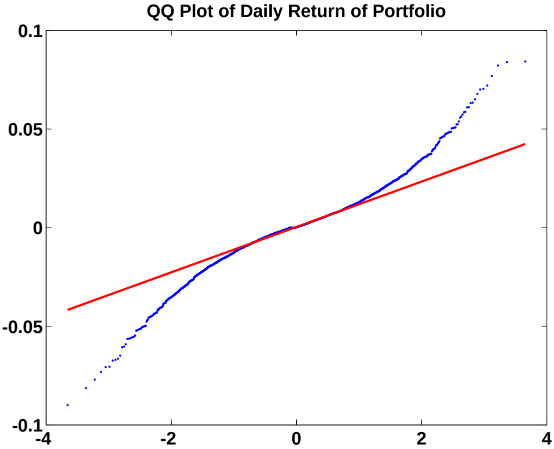


Figure 12: Normal QQ plot of daily log-returns of portfolio (Bayer, BMW, Siemens) from 1992-01-01 to 2006-12-29. MSRportfolio

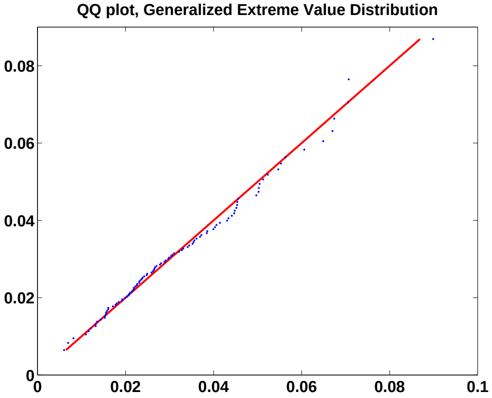


Figure 13: QQ plot of 100 tail values of daily log-returns of portfolio (Bayer, BMW, Siemens) from 1992-01-01 to 2006-09-01 against Generalized Extreme Value Distribution with parameter $\gamma = 0.0498$ estimated globally with block maxima method. MSRtailGEV

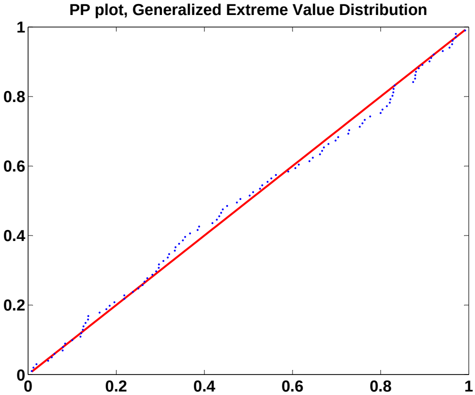


Figure 14: PP plot of 100 tail values of daily log-returns of portfolio (Bayer, BMW, Siemens) from 1992-01-01 to 2006-09-01 against Generalized Extreme Value Distribution with parameter $\gamma = 0.0498$ estimated globally with block maxima method. MSRtailGEV

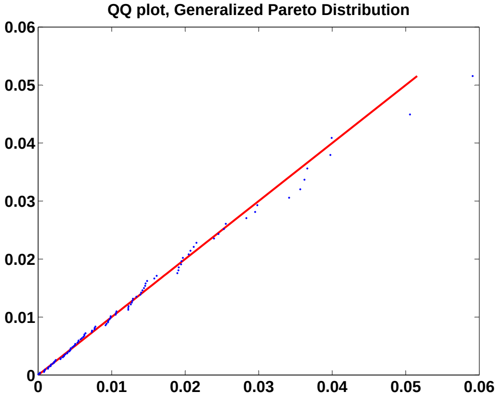


Figure 15: QQ plot of 100 tail values of daily log-returns of portfolio (Bayer, BMW, Siemens) from 1992-01-01 to 2006-09-01 against Generalized Pareto Distribution with parameter $\gamma = -0.0768$ estimated globally with POT method. MSRtailGPareto

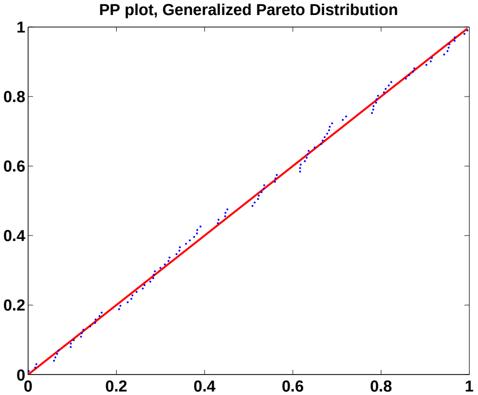


Figure 16: PP plot of 100 tail values of daily log-returns of portfolio (Bayer, BMW, Siemens) from 1992-01-01 to 2006-09-01 against Generalized Pareto Distribution with parameter $\gamma = -0.0768$ estimated globally with POT method. MSRtailGPareto

Nonparametric method

Let $y_1, \ldots, y_m$ be the exceedances over u which are assumed to be iid with cdf F_u.

$$\overline{F}_u(y) = P(X - u > y \mid X > u) = \overline{F}(y+u)/\overline{F}(u), \quad \text{i.e.}$$
$$\overline{F}(x) = \overline{F}(u) \cdot \overline{F_u}(x-u), \;\; u < x < \infty. \tag{5}$$

For large u and using Theorem (MDA of GEV distribution) we can approximate F_u with $W_{\gamma,\beta}$ by choosing γ and β approximately. $\widehat{F}_n(u)$ is replaced by

$$\widehat{F}_n(u) = \frac{n - N(u)}{n} = 1 - \frac{N(u)}{n}.$$

Definition (POT Estimator)

The POT estimator for $\overline{F}(x), x$ large is defined by

$$\overline{F}^{\wedge}(x) = \frac{N(u)}{n}\,\overline{W}_{\widehat{\gamma},\widehat{\beta}}(x-u) = \frac{N(u)}{n}\left\{1 + \frac{\widehat{\gamma}(x-u)}{\widehat{\beta}}\right\}^{-1/\widehat{\gamma}},\ u < x < \infty,$$

where $\widehat{\gamma}, \widehat{\beta}$ are appropriate estimators for γ, β.

$\widehat{\gamma}$ and $\widehat{\beta}$ may be computed via the ML method on the basis of the excesses $Y_1, \ldots, Y_{N(u)}$.

MLE of $\widehat{\gamma}$ and $\widehat{\beta}$

Fix $N(u) = m$ for the moment. $Y_1, \ldots, Y_m$ iid Pareto $W_{\gamma,\beta}, \gamma > 0$, with pdf:

$$p(y) = \frac{1}{\beta}(1 + \frac{\gamma y}{\beta})^{-\frac{1}{\gamma}-1}, \;\; x \geq 0.$$

Log-likelihood:

$$\ell(\gamma, \beta \mid Y_1, \ldots, Y_m) = -m \log \beta - \left(\frac{1}{\gamma} + 1\right) \sum_{j=1}^{m} \log(1 + \frac{\gamma}{\beta} Y_j).$$

Theorem

For all $\gamma > -\frac{1}{2}$, it holds for $m \to \infty$:

$$\sqrt{m}(\widehat{\gamma} - \gamma,\ \frac{\widehat{\beta}}{\beta} - 1) \xrightarrow{\mathcal{L}} N_2(0, D^{-1}),$$

where $D = (1+\gamma)\begin{pmatrix} 1+\gamma & -1 \\ -1 & 2 \end{pmatrix}$, i.e. $(\widehat{\gamma}, \widehat{\beta})$ are asymptotically normal distributed. The estimators are also asymptotically efficient.

Definition (POT Quantile estimator)

The POT quantile estimator $\widehat{x}_q$ for the q-quantile $x_q = F^{-1}(q)$ is the solution of $\overline{F}^{\wedge}(\widehat{x}_q) = 1 - q$, i.e.

$$\widehat{x}_q = u + \frac{\widehat{\beta}}{\widehat{\gamma}} \left[\left\{ \frac{n}{N(u)}(1-q) \right\}^{-\widehat{\gamma}} - 1 \right].$$

Comparison to the empirical quantile

Choose u such that $N(u) = m > n(1-q)$, i.e. $u = X^{(m+1)}$.

POT quantile estimator:

$$\widehat{x}_{q,m} = X^{(m+1)} + \frac{\widehat{\beta}_m}{\widehat{\gamma}_m}\left[\left\{\frac{n}{m}(1-q)\right\}^{-\widehat{\gamma}_m} - 1\right],$$

Empirical quantile: $\widehat{x}_q^s = X^{([n(1-q)]+1)}$.
Simulation studies show that

$$m_0 = \text{argmin}_m \mathbf{E}(\widehat{x}_{q,m} - x_q)^2$$

is bigger than $[n(1-q)] + 1$. This means that the POT estimator is better than $\widehat{x}_q^s$ in MSE terms.

Mean Square Error Dilemma

- u too big: there are not enough exceedances Y and thus the variance is too high.
- u too small: the approximation by Pareto is not good enough and thus a bias occurs.

Theorem

Let Z be a $W_{\gamma,\beta}$ distributed random variable with $0 \leq \gamma < 1$, then the mean excess function of Z is linear:

$$e(u) = \mathbf{E}\{Z - u | Z > u\} = \frac{\beta + \gamma u}{1 + \gamma}, \quad u \geq 0, \quad \textit{for } 0 \leq \gamma < 1.$$

Motivation: Choose u of the POT estimator such that the empirical mean excess function is approximately linear.

Models for distributions with ‘fat tails’

Pareto distribution:

$$\overline{F}(x) = P(X > x) \sim kx^{-\alpha}$$

where $\alpha > 0$. If $\alpha > c$, then $\mathbf{E}[|X_t|^c] < \infty$.

Estimation of α:

$$\log \overline{F}(x) \approx \log k - \alpha \log x. \tag{6}$$

Take the sample $X_1, \ldots, X_n$ and form the order statistic $X^{(1)} \geq \ldots \geq X^{(n)}$.

Estimate the probability $\overline{F}(x)$ for $x = X_{(j)}$ by the relative frequency

$$\frac{\#\{t; X_t \geq X_{(j)}\}}{n} = \frac{j}{n}$$

Replace $\overline{F}(X_{(j)})$ in (6) with the estimator $\frac{j}{n}$:

$$\log \frac{j}{n} \approx \log k - \alpha \log X_{(j)} \ . \tag{7}$$

$\widehat{\alpha}$ is the slope of the linear regression (7) obtained e.g. by least squares.

The linear approximation of $\log \overline{F}(x)$ will only be good in the tails. Thus we estimate (7) using the m biggest order statistics, only.

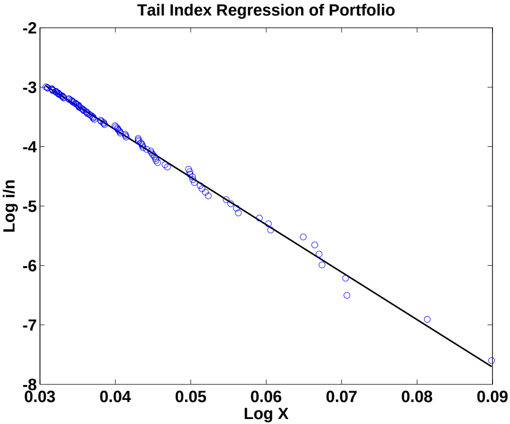


Figure 17: Right tail of the logarithmic empirical distribution of the portfolio (Bayer, BMW, Siemens) negative log-returns from 1992-01-01 to 2006-06-01. MSRtailport

ANOVA TABLE

Source	SS	df	MS	F	Prob $>$ F
Columns	802.644	1	802.644	1863.54	0
Error	85.28	198	0.431		
Total	887.924	199			

Table 1: Analysis of variance performed on the negative log-returns of the portfolio (Bayer, BMW, Siemens).

Hill estimator

The Hill estimator is based on maximum likelihood:

$$\widehat{\alpha} = \left(\frac{1}{m} \sum_{i=1}^{m} \log X^{(i)} - \log X^{(m)} \right)^{-1}, \tag{8}$$

m denotes the number of observations from the tails used in the estimation.

How to choose m?

If m is large, there is a large bias,
if m is small, there is a large variance.

'Rule of thumb'

$$0.005 \leq \frac{m}{n} \leq 0.01$$

Motivation of the Hill estimator

$\overline{F}(x) = x^{-\alpha} L(x) \quad L(x) \approx c^{\alpha} \quad V_j \stackrel{\text{def}}{=} \log(X_j/c).$

Then

$P(V_j > v) = P(X_j > ce^v) = \overline{F}(ce^v) = \frac{c^{\alpha}}{(ce^v)^{\alpha}} = e^{-\alpha v},\ y \geq 0,$

$V_1, \ldots, V_n$ are i.i.d. exponentially distributed random variables with parameter α.

$(\mathbf{E} V_j)^{-1} = \alpha$. The MLE for α: $\alpha = 1/\overline{V}_n$.

$$\widehat{\alpha}_H = \left(\frac{1}{m} \sum_{j=1}^{m} \log X^{(j)} - \log X^{(m)} \right)^{-1}.$$

Hill estimator is consistent:

$$\widehat{\alpha}_H \xrightarrow{P} \alpha$$

$n, m \to \infty,\ m/n \to 0.$

$$\sqrt{m}(\widehat{\alpha}_H - \alpha) \xrightarrow{\mathcal{L}} N(0, \alpha^2)$$

Also in this situation one has the **bias -variance tradeoff:**

$$\frac{\overline{F}(x)}{\overline{F}(X^{(m)})} = \frac{L(x)}{L(X^{(m)})} \left(\frac{X^{(m)}}{x} \right)^{\alpha} \approx \left(\frac{X^{(m)}}{x} \right)^{\alpha}, \tag{9}$$

$\overline{F}(X^{(m)}) \approx m/n.$

B. Hill on BBI: 

Hill Estimator of $\overline{\mathrm{F}}(\mathrm{x})$

$$\overline{F}_H^{\wedge}(x) = \frac{m}{n}\left(\frac{X^{(m)}}{x}\right)^{\widehat{\alpha}_H}$$

Hill Quantile Estimator

$$\begin{aligned}
\widehat{x}_{q,H} &= X^{(m)}\left\{\frac{n}{m}(1-q)\right\}^{-1/\widehat{\alpha}_H} \\
&= X^{(m)} + X^{(m)}\left[\left\{\frac{n}{m}(1-q)\right\}^{-\widehat{\gamma}_H} - 1\right]
\end{aligned}$$

where $\widehat{\gamma}_H = 1/\widehat{\alpha}_H$ SFEhillquantile

Shape Parameter Estimates

Mathod	$\widehat{\gamma}$
Block Max	0.0498
POT	-0.0768
Regression	0.0125
Hill	0.3058

Table 2: Values of shape Parameter estimated with different methods for the 100 tail observations of the portfolio (Bayer, BMW, Siemens) negative log-returns from 1992-01-01 to 2006-09-01.

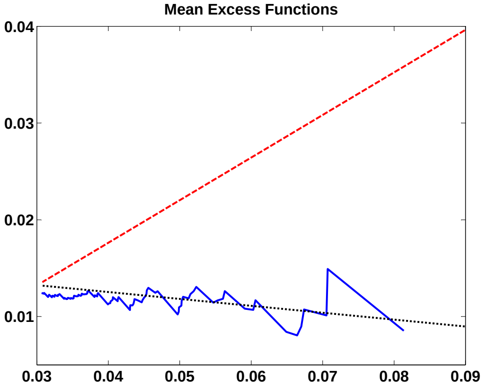


Figure 18: Empirical mean excess plot (blue line), mean excess plot of generalized Pareto distribution (black line) and mean excess plot of Pareto distribution with parameter estimated with Hill estimator (red line) for portfolio (Bayer, BMW, Siemens) negative log-returns from 1992-01-01 to 2006-09-01. MSRmeanExcessFun

Value-at-Risk with Block Maxima Model

For a sample of negative returns $\{X_t\}_{t=1}^{T}$

1. decompose the time period T into k non-overlapping time periods of length n
2. select maximal returns $\{Z_j\}_{j=1}^{k}$ where $Z_j = \min\{X_{(j-1)n+1}, \ldots, X_{jn}\}$
3. estimate the parameters of generalized extreme value distribution for the maximal returns $\{Z_j\}_{j=1}^{k}$
4. compute the VaR of the position with given α ($\alpha = 0.95$)

$$\begin{aligned} \alpha^n = 1 - F(VaR) &= exp\left[-\left\{1 + \gamma\left(\frac{VaR - \mu}{\sigma}\right)^{-1/\gamma}\right\}\right] \\ VaR &= \mu + \frac{\mu}{\gamma}[\{(1 - \alpha^n)\}^{\gamma} - 1]. \end{aligned}$$

Moving Window

- use static windows of size $w = 250$ scrolling in time t for VaR estimation:

$$\{X_t\}_{t=s-w+1}^{s}$$

  for $s = w, \ldots, T$

- the VaR estimation procedure generates a time series $\{\widehat{VaR}^{t}_{1-\alpha}\}_{t=w}^{T}$ and $\{\widehat{\mu}_t\}_{t=w}^{T}$, $\{\widehat{\sigma}_t\}_{t=w}^{T}$, $\{\widehat{\gamma}_t\}_{t=w}^{T}$ of parameters estimates.

Backtesting

The estimated VaR values are compared with true realizations $\{l_t\}$ of the Profit and Loss function. An *exceedance* occurs when l_t is smaller than $\widehat{VaR}^t_{1-\alpha}$.
The ratio of the number of exceedances to the number of observations gives the *exceedances ratio*:

$$\widehat{\alpha} = \frac{1}{T-h} \sum_{t=h+1}^{T} I(l_t < \widehat{VaR}^t_{1-\alpha})$$

Value-at-Risk with Block Maxima Model

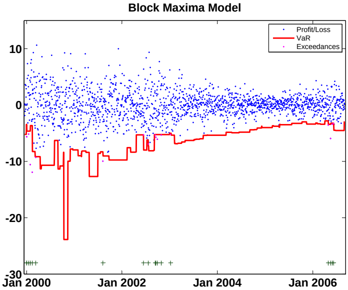


Figure 19: Value-at-Risk estimation at 0.01 level for portfolio: Bayer, BMW, Siemens. Time period: from 1992-01-01 to 2006-09-01. Size of moving window 250. Backtesting result $\widehat{\alpha} = 0.0103$. MSRvar_block_max_backtesting

Value-at-Risk with Block Maxima Model

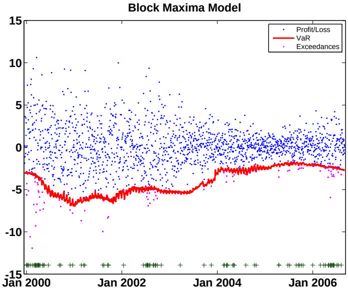


Figure 20: Value-at-Risk estimation at 0.05 level for portfolio: Bayer, BMW, Siemens. Time period: from 1992-01-01 to 2006-09-01. Size of moving window 250. Backtesting result $\widehat{\alpha} = 0.0514$. MSRvar_block_max_backtesting

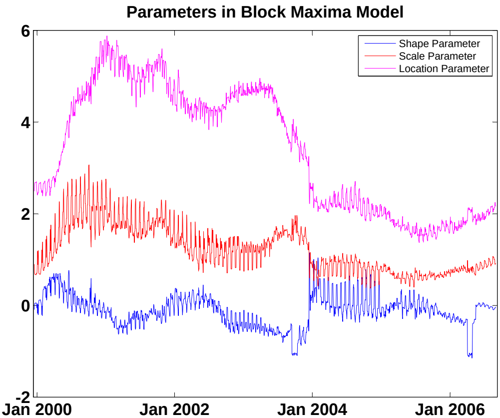


Figure 21: Parameters estimated in Block Maxima Model for portfolio: Bayer, BMW, Siemens. Time period: from 1992-01-01 to 2006-09-01.
MSRvar_block_max_params

Value-at-Risk with POT Model

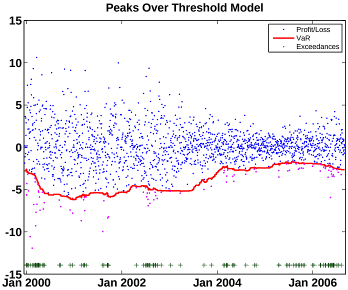


Figure 22: Value-at-Risk estimation at 0.05 level for portfolio: Bayer, BMW, Siemens. Time period: from 1992-01-01 to 2006-09-01. Size of moving window 250. Backtesting result $\widehat{\alpha} = 0.0571$.
MSRvar_pot_backtesting

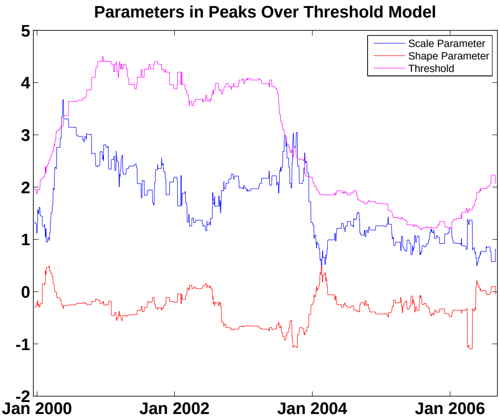


Figure 23: Parameters estimated in POT Model for portfolio: Bayer, BMW, Siemens. Time period: from 1992-01-01 to 2006-09-01. MSRvar_pot_params

Joint Extreme Values

Gemeinsamer Extremwert

Łączne Wartości Ekstremalne

Sdružené Krajní Hodonoty

联合极值

同時極値

القيمة الحدية
المشتركة

극단값

Measures of dependence

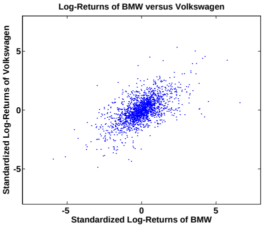


Figure 24: Scatter plot of daily standardized log-returns of BMW versus Volkswagen. MSRsca_bmw_vw

- Pearson's correlation coefficient ρ
- Kendall's τ
- Spearman's rank correlation coefficient ρ_S

Pearson's ρ

The linear correlation coefficient between X and Y

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{\sigma^2(X)\sigma^2(Y)}},$$

where $Cov(X, Y) = E(XY) - E(X)E(Y)$.

- Need finite variance (extreme value distributions e.g. Fréchet).
- Linear correlation is a measure of linear dependence.
- Linear correlation is invariant only under strictly increasing linear transformations: $\rho(X, Y) \neq \rho(\log X, \log Y)$.

Kendall's τ

Kendall's tau for the random vector (X, Y)

$$\tau(X, Y) = P\{(X - \widetilde{X})(Y - \widetilde{Y}) > 0\} - P\{(X - \widetilde{X})(Y - \widetilde{Y}) < 0\},$$

where $(\widetilde{X}, \widetilde{Y})$ is an independent copy of (X, Y).

Spearman's ρ_s

Spearman's rho for the random vector (X, Y)

$$\rho_S(X,Y) = 3[P\{(X-\widetilde{X})(Y-Y') > 0\} - P\{(X-\widetilde{X})(Y-Y') < 0\}],$$

where (X, Y), $(\widetilde{X}, \widetilde{Y})$ and (X', Y') are independent copies.

- Kendall's τ and Spearman's ρ_s are invariant under strictly increasing componentwise transformations.

Portfolios	τ	ρ_s
BAY - SIE	0.3483	0.4822
BMW - VOW	0.4257	0.5822
SIE - VOW	0.3693	0.5106

Table 3: Kendall's τ and Spearman's ρ_s of daily log-returns of three different portfolios. Time period: 1st January 1999 – 1st September 2006.

Tail Dependence

- Risk behavior is determined by tails large losses that can occur jointly.
- Pearson's correlation can not capture joint large loss events.
- Tail dependence describes the limiting proportion that one margin exceeds a certain threshold given that the other margin has already exceeded that threshold.

Upper tail dependence

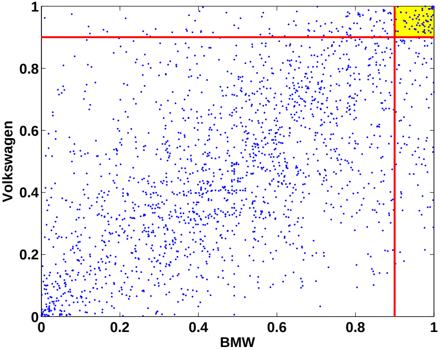


Figure 25: UTD for standardized log-returns of BMW vs Volkswagen transformed by t-Student cdf. MSRutd_bmw_vw

Upper tail dependence

Let $(X, Y) \sim F$ with margins F_1 and F_2.
Coefficient of upper tail dependence (UTD):

$$\lambda_U = \lim_{u \nearrow 1} P\{Y > F_2^{-1}(u) | X > F_1^{-1}(u)\}.$$

Asymptotical upper tail dependence: $\lambda_U \in (0, 1]$.
Asymptotical upper tail independence: $\lambda_U = 0$.

Lower tail dependence

Let $(X, Y) \sim F$ with margins F_1 and F_2.
Coefficient of lower tail dependence (LTD):

$$\lambda_L = \lim_{u \searrow 0} P\{Y \leq F_2^{-1}(u) | X \leq F_1^{-1}(u)\}.$$

Asymptotical lower tail dependence: $\lambda_L \in (0, 1]$.
Asymptotical lower tail independence: $\lambda_L = 0$.

Elliptical distribution

An elliptical distribution is obtained by a transformation

$$X = \mu + AY$$

of a spherical distribution Y with $\mu \in \mathbb{R}^d$, $A \in \mathbb{R}^{d \times d}$.
Y has a spherical distribution, if and only if the characteristic function can be represented as

$$E\{\exp(it^\top Y)\} = \phi(t_1^2 + \ldots + t_d^2)$$

with some function $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ called the characteristic generator.

Elliptical distribution

Let $X \sim N_n(0, I_n)$. Since the components $X_i \sim N(0,1)$, $i = 1, \ldots,$ n, are independent and the characteristic function of X_i is $exp(-t_i^2/2)$, the characteristic function of X is

$$exp\left\{-\frac{1}{2}(t_1^2 + \ldots + t_n^2)\right\} = exp\left\{-\frac{1}{2}t^\top t\right\}.$$

Normal distribution is elliptical distribution, where the characteristic generator $\phi(u) = exp(-u/2)$.

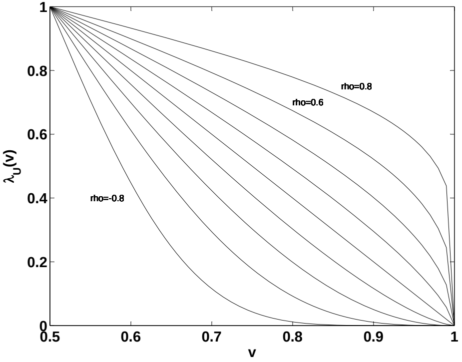


Figure 26: The function $\lambda_U(v) = 2 \cdot P\{X_1 > F_1^{-1}(v) \mid X_2 = F_2^{-1}(v)\}$ for a bivariate normal distribution with correlation coefficients $\rho = -0.8, -0.6, \ldots, 0.6, 0.8$. Note that $\lambda_U = 0$ for all $\rho \in (-1, 1)$.
MSRtail_dep_normal

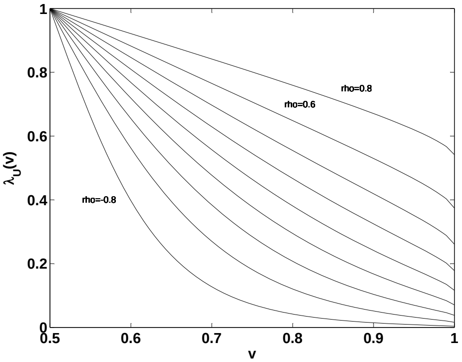


Figure 27: The function $\lambda_U(v) = 2 \cdot P\{X_1 > F_1^{-1}(v) \mid X_2 = F_2^{-1}(v)\}$ for a bivariate t-distribution with 3 degrees of freedom and with correlation coefficients $\rho = -0.8, -0.6, \ldots, 0.6, 0.8$.

MSRtail_dep_tStudent

Elliptical distribution

The density function, if it exists, of an elliptically-contoured distribution has the following form:

$$f(x) = |\Sigma|^{-1/2} g\big\{(x-\mu)^\top \Sigma^{-1}(x-\mu)\big\}, \quad x \in \mathbb{R}^n,$$

for some function $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$, called the density generator.

Regular variation index

Elliptical distributions are upper and lower tail-dependent if the tail of their density generator is regularly varying.
A measurable function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is called *regularly varying* (at ∞) with index $\alpha \in \mathbb{R}$ if for any $t > 0$

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\alpha}.$$

When $\alpha = 0$, f is called slowly varying.

Tail dependence for elliptical distributions

Let (X_1, X_2) be elliptically distributed with a regularly-varying density generator with an index $\alpha > 0$ and a correlation coefficient:

$$\rho = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}}.$$

Tail dependence coefficient is given by

$$\lambda(\alpha, \rho) = \lambda_U = \lambda_L = \frac{\int_0^{h(\rho)} \frac{u^\alpha}{\sqrt{1-u^2}} du}{\int_0^1 \frac{u^\alpha}{\sqrt{1-u^2}} du},$$

with

$$h(\rho) = \left\{1 + \frac{(1-\rho)^2}{1-\rho^2}\right\}^{-1/2}.$$

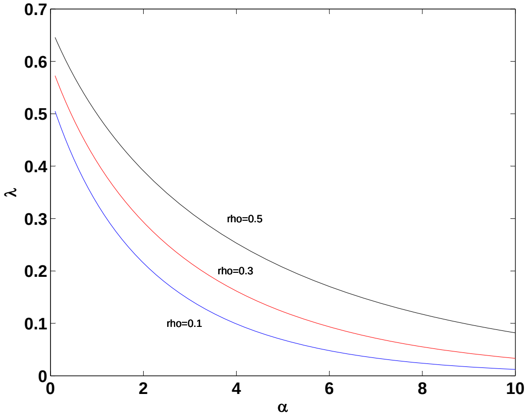


Figure 28: Tail-dependence coefficient λ versus regular variation index α for "correlation" coefficients $\rho = 0.5,\ 0.3,\ 0.1$.

MSRtaildepVsVarIndex

Type	Density generator g or characteristic generator ϕ	Parameters	α for $n=2$
Normal	$g(u) = c_n \exp(-u/2)$	—	∞
t	$g(u) = c_n \left(1 + \frac{t}{\theta}\right)^{-(n+\theta)/2}$	$\theta > 0$	θ
Symmetric general. hyperbolic	$g(u) = c_n \frac{K_{\lambda - \frac{n}{2}}\{\sqrt{\varsigma(\chi+u)}\}}{(\sqrt{\chi+u})^{\frac{n}{2}-\lambda}}$	$\varsigma, \chi > 0$ $\lambda \in \mathbb{R}$	∞
Symmetric θ-stable	$\phi(u) = \exp\left\{ -\left(\frac{1}{2}u\right)^{\theta/2}\right\}$	$\theta \in (0, 2]$	θ
logistic	$g(u) = c_n \frac{\exp(-u)}{\{1+\exp(-u)\}^2}$	—	∞

Table 4: Regular variation index α for various density generators g of multivariate elliptical distributions. K_ν is the modified Bessel function of the third kind.

Copulae

A copula is a multivariate distribution function defined on the unit cube $[0,1]^d$, with uniformly distributed margins.

$$
\begin{aligned}
P(X_1 \leq x_1, \ldots, X_d \leq x_d) &= C\left\{P(X_1 \leq x_1), \ldots, P(X_d \leq x_d)\right\} \\
&= C\left\{F_1(x_1), \ldots, F_d(x_d)\right\}
\end{aligned}
$$

W. Hoeffding on BBI:

Bivariate copulae

A *2-dimensional copula* is a function $C : [0,1]^2 \rightarrow [0,1]$ with the following properties:

1. For every $u \in [0,1]$, $C(0,u) = C(u,0) = 0$ (**grounded**)
2. For every $u \in [0,1]$, $C(u,1) = u$ and $C(1,u) = u$
3. For every $(u_1,u_2),(v_1,v_2) \in [0,1] \times [0,1]$ with $u_1 \leq v_1$ and $u_2 \leq v_2$: $C(v_1,v_2) - C(v_1,u_2) - C(u_1,v_2) + C(u_1,u_2) \geq 0$ (**2-increasing**)

Sklar's Theorem

For a distribution function F with marginals $F_{X_1} \dots, F_{X_d}$. There exists a copula $C : [0,1]^d \to [0,1]$, such that

$$F(x_1, \dots, x_d) = C\{F_{X_1}(x_1), \dots, F_{X_d}(x_d)\} \tag{10}$$

for all $x_i \in \overline{\mathbb{R}}$, $i = 1, \dots, d$. If $F_{X_1}, \dots, F_{X_d}$ are cts, then C is unique. If C is a copula and $F_{X_1}, \dots, F_{X_d}$ are cdfs, then the function F defined in (10) is a joint cdf with marginals $F_{X_1}, \dots, F_{X_d}$.

Copulae

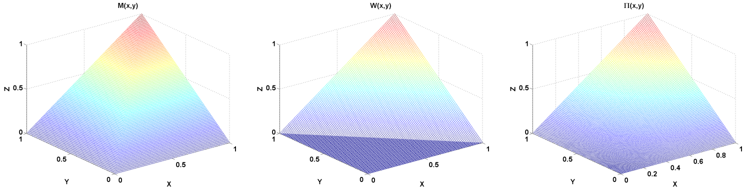


Figure 29: $M(u,v) = \min(u,v)$, $W(u,v) = \max(u+v-1,0)$ and $\Pi(u,v) = uv$ MSR_Frechet_bounds

M. Fréchet on BBI: 

Gauss Copula

$$
\begin{aligned}
C(u_1, u_2) &= \Phi_\rho\{\Phi^{-1}(u_1), \Phi^{-1}(u_2)\} \\
&= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} \, dx\, dy
\end{aligned}
$$

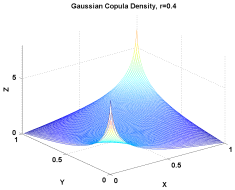


Figure 30: Gauss copula density, parameter $\rho = 0.4$. MSRpdf_cop_Gauss

C. Gauss on BBI: 

t-Student Copula

$$
\begin{aligned}
C(u_1, u_2) &= t_{\rho,\nu}\{t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)\} \\
&= \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right\}^{-(\nu+2)/2} dx\, dy
\end{aligned}
$$

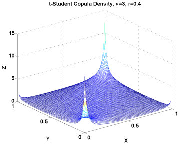


Figure 31: t-Student copula density, parameters $\nu = 3$ and $\rho = 0.4$.
MSRpdf_cop_tStudent

W. Gosset on BBI: 

Archimedean Copulae

Archimedean copula:

$$C(u, v) = \psi^{[-1]}\{\psi(u) + \psi(v)\}$$

for a continuous, decreasing and convex ψ, $\psi(1) = 0$.

$$\psi^{[-1]}(t) = \begin{cases} \psi^{-1}(t), & 0 \le t \le \psi(0), \\ 0, & \psi(0) < t \le \infty. \end{cases}$$

For $\psi(0) = \infty$: $\psi^{[-1]} = \psi^{-1}$.

Gumbel Copula

$$C(u,v) = \exp\left[-\left\{(-\log u)^{\theta} + (-\log v)^{\theta}\right\}^{\frac{1}{\theta}}\right]$$

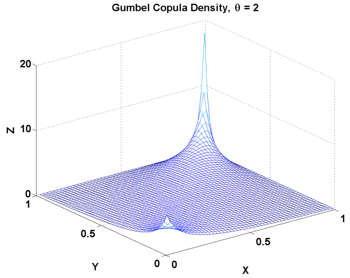


Figure 32: Gumbel copula density, $\theta = 2$. MSRpdf_cop_Gumbel

E. Gumbel on BBI: 

Clayton Copula

$$C(u,v) = \max\left\{(u^{-\theta} + v^{-\theta} - 1)^{\frac{1}{\theta}}, 0\right\}$$

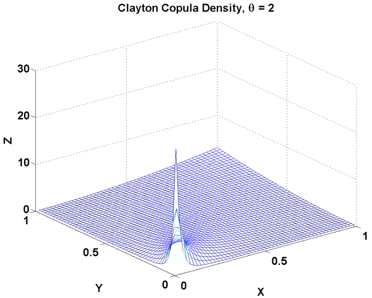


Figure 33: Clayton copula density, $\theta = 2$. MSRpdf_cop_Clayton

Frank Copula

$$C(u,v) = -\frac{1}{\theta}\log\left\{1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right\}$$

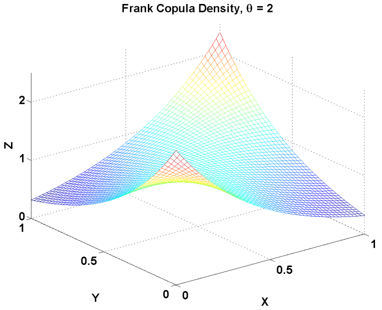


Figure 34: Frank copula density, $\theta = 2$. MSRpdf_cop_Frank

Tail Dependence and Copulae

Tail dependence is a copula property:

$$\begin{aligned}\lambda_U &= \lim_{v \nearrow 1} \frac{1 - 2v + C(v,v)}{1-v}, \\ \lambda_L &= \lim_{v \searrow 0} \frac{C(v,v)}{v}. \end{aligned} \tag{11}$$

Copula	τ	λ_U	λ_L
Gauss	$\frac{2}{\pi}\arcsin\rho$	0 for $\rho < 1$	0 for $\rho < 1$
	$-1 \leq \tau \leq 1$	1 for $\rho = 1$	1 for $\rho = 1$
t_ν	$\frac{2}{\pi}\arcsin\rho$	$2\bar{t}_{\nu+1}\left(\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$	λ_U
	$-1 \leq \tau \leq 1$	$-1 \leq \rho_S \leq 1$	
Gumbel	$1 - \frac{1}{\theta}$	$2 - 2^{\frac{1}{\theta}}$	0
	$0 \leq \tau \leq 1$		
Clayton	$\frac{\theta}{\theta+2}$	0	$2^{-\frac{1}{\theta}}$
	$0 \leq \tau \leq 1$		
Frank	$1 - \frac{4}{\theta}\{1 - D_1(\theta)\}$	0	0
	$D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k}{e^t-1}dt$		
	$-1 \leq \tau \leq 1$		

Table 5: Kendall's τ and TDCs for various selected copulae.

Tail Dependence for t-copula

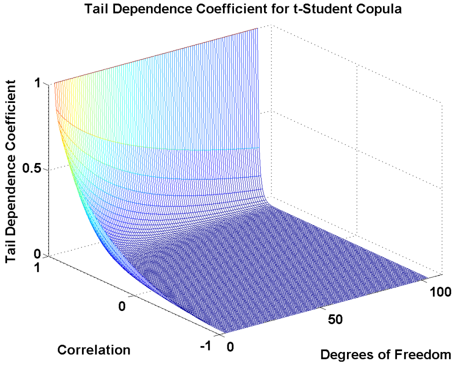


Figure 35: Value of upper (lower) tail dependence coefficient for the t-copula. MSR_TDC_tStudent

Empirical Copula

Let $\{x^i_{(1)}, \ldots, x^i_{(T)}\}$ be the order statistics of i-th stock and $\{r^i_1, \ldots, r^i_T\}$ corresponding rank statistics such that $x^i_{(r^i_t)} = x^i_t$ for all $i = 1, \ldots, d$. Any function

$$\widehat{C}\left(\frac{t_1}{T}, \ldots, \frac{t_d}{T}\right) = \frac{1}{T}\sum_{t=1}^{T}\prod_{i=1}^{d} I\left(r^i_t \le t_i\right) \tag{12}$$

is an empirical copula.

Estimation of the upper TDC

$\{X_j\}_{j=1}^n \in \mathbb{R}^2$ i.i.d. the empirical copula is

$$C_n(u,v) = F_n\{F_{1n}^{-1}(u), F_{2n}^{-1}(v)\},$$

F_{in} empirical cdfs of X_{ij}, $j = 1, ..., n$.

$$\begin{aligned}\widehat{\lambda}_{U,n}^{(1)} &= \frac{n}{k} C_n\Big(\Big(1-\frac{k}{n},1\Big] \times \Big(1-\frac{k}{n},1\Big]\Big) \\ &= \frac{1}{k}\sum_{j=1}^{n} I(R_{n1}^{(j)} > n-k, R_{n2}^{(j)} > n-k)\end{aligned}$$

Here $R_{n1}^{(j)}$ and $R_{n2}^{(j)}$ is the rank of $X_1^{(j)}$ and $X_2^{(j)}$ respectively.

Estimation of the lower TDC

Similarly

$$\widehat{\lambda}^{(1)}_{L,n} = \frac{n}{k} C_n\Big(\frac{k}{n}, \frac{k}{n}\Big) = \frac{1}{k} \sum_{j=1}^{n} I(R^{(j)}_{n1} \le k, R^{(j)}_{n2} \le k), \qquad (13)$$

where $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$, From EVT:

$$\begin{aligned} \widehat{\lambda}^{(2)}_{U,n} &= 2 - \frac{n}{k}\Big\{1 - C_n\Big(1 - \frac{k}{n}, 1 - \frac{k}{n}\Big)\Big\} \\ &= 2 - \frac{1}{k} \sum_{j=1}^{n} I(R^{(j)}_{n1} > n - k \text{ or } R^{(j)}_{n2} > n - k), \qquad (14) \end{aligned}$$

obtains the usual nonparametric bias-variance problem.

Application

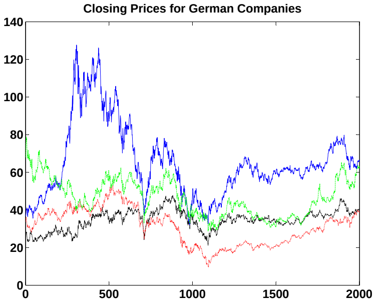


Figure 36: Closing prices of stocks: BMW, Bayer, Siemens, Volkswagen. Time period: 1st January 1999 – 1st September 2006, 2000 data points.

MSRclose

Returns

Let $P_1, \ldots, P_n$ be a time series of stock's prices.

- Simple return is defined as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

- Logarithmic return (log-return) is defined as

$$r_t = \log \frac{P_t}{P_{t-1}}.$$

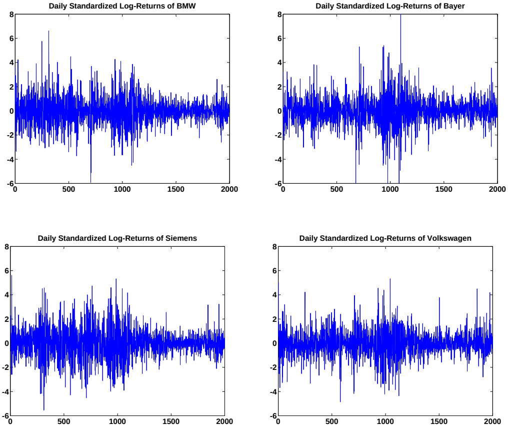


Figure 37: Daily stock standardized log-returns: BMW, Bayer, Siemens, Volkswagen. MSRstdlogret

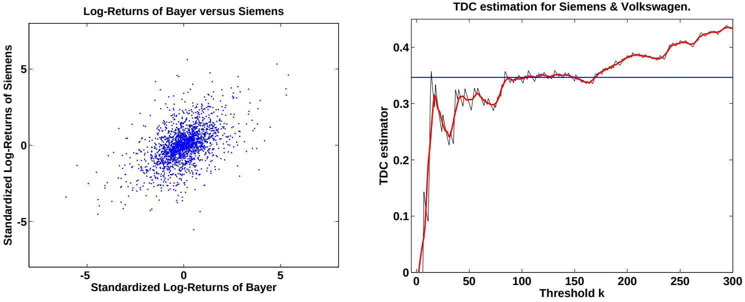


Figure 38: Scatter plot of Bayer versus Siemens daily stock standardized log-returns and the corresponding TDC estimate $\widehat{\lambda}_U$ for various thresholds k. Chosen $k \approx 135$, TDC $\widehat{\lambda}_U = 0.3465$. MSRnonp_utd

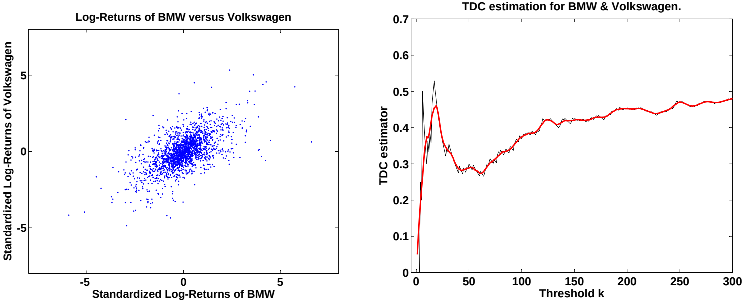


Figure 39: Scatter plot of BMW versus Volkswagen daily stock standardized log-returns and the corresponding TDC estimate $\widehat{\lambda}_U$ for various thresholds k. Chosen $k \approx 128$, TDC $\widehat{\lambda}_U = 0.4184$. MSRnonp_utd

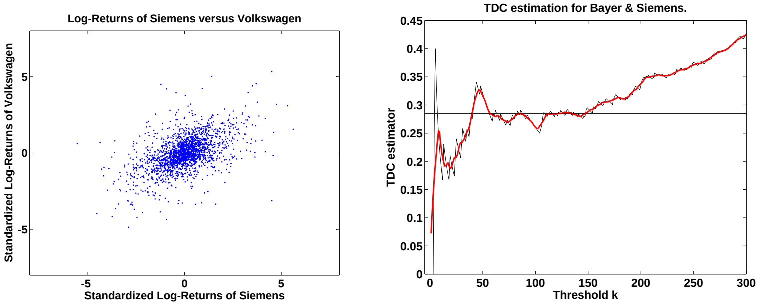


Figure 40: Scatter plot of Siemens versus Volkswagen daily stock standardized log-returns and the corresponding TDC estimate $\widehat{\lambda}_U$ for various thresholds k. Chosen $k \approx 103$, TDC $\widehat{\lambda}_U = 0.2848$. MSRnonp_utd

Fitting copulae to data

Copula	Parameters	Upper TDC	Lower TDC
Gauss	0.5202	0	0
$t(4)$	0.5202	0.2774	0.2774
Gumbel	1.5344	0.4290	0
Clayton	1.0688	0	0.5228
Frank	3.4876	0	0
Nonparametric	—	0.2848	0.4057
bivariate t-cdf	5	0.1778	0.1778

Table 6: Tail dependence coefficients of different copulae for BAY and SIE. Standardized margins modeled with t-distribution.

Fitting copulae to data

Copula	Parameters	Upper TDC	Lower TDC
Gauss	0.6199	0	0
$t(4)$	0.6199	0.3397	0.3397
Gumbel	1.7412	0.51100	0
Clayton	1.4824	0	0.6265
Frank	4.5260	0	0
Nonparametric	—	0.4184	0.5272
bivariate t-cdf	8	0.1802	0.1802

Table 7: Tail dependence coefficients of different copulae for BMW and VOW. Standardized margins modeled with t-distribution.

Fitting copulae to data

Copula	Parameters	Upper TDC	Lower TDC
Gauss	0.5481	0	0
$t(4)$	0.5481	0.2936	0.2936
Gumbel	1.5856	0.4517	0
Clayton	1.1712	0	0.5533
Frank	3.7528	0	0
Nonparametric	—	0.3465	0.3470
bivariate t-cdf	6	0.1621	0.1621

Table 8: Tail dependence coefficients of different copulae for SIE and VOW. Standardized margins modeled with t-distribution.

Value-at-Risk with Copulae

For a sample $\{X_t\}_{t=1}^T$

1. specification of marginal distributions $F_{X_j}(x_j;\delta_j)$
2. specification of copula $C(u_1,\dots,u_d;\theta)$
3. fit of the copula C
4. generation of Monte Carlo data $X_{T+1}{\sim}C\{F_1(x_1),\dots,F_d(x_d);\widehat{\theta}\}$
5. generation of a sample of portfolio losses $L_{T+1}(X_{T+1})$
6. estimation of $\widehat{VaR}_{1-\alpha}$, the empirical quantile at level α from $L_{T+1}(X)$.

Moving window

For a sample $\{X_t\}_{t=1}^{T}$

1. specification of marginal distributions $F_{X_j}(x_j; \delta_j)$
2. specification of returns' subsets of size h: $\{y_{j,t}\}_{t=s-h+1}^{s}$ for $s = h, \ldots, T-1$
3. specification of copulae $C_s(u_1, \ldots, u_d; \theta)$ for every subset $\{y_{j,t}\}_{t=s-h+1}^{s}$
4. fit of the copulae C_s, $s = h, \ldots, T-1$
5. generation of Monte Carlo data $X_{s+1} \sim C_s\{F_1(x_1), \ldots, F_d(x_d); \widehat{\theta}\}$ for $s = h, \ldots, T-1$
6. generation of a samples of portfolio losses $L_{s+1}(X_{s+1})$
7. estimation of $\{\widehat{VaR}^{j}_{1-\alpha}\}_{j=1}^{T-h}$.

Backtesting

The estimated VaR values are compared with true realizations $\{l_t\}$ of the Profit and Loss function. An *exceedance* occurs when l_t is smaller than $\widehat{VaR}^t_{1-\alpha}$.
The ratio of the number of exceedances to the number of observations gives the *exceedances ratio*:

$$\widehat{\alpha} = \frac{1}{T-h} \sum_{t=h+1}^{T} I(l_t < \widehat{VaR}^t_{1-\alpha})$$

Value-at-Risk estimation

Copula	BAY - SIE	BMW - VOW	SIE - VOW
Gauss	0.0320	0.0394	0.0371
t-Student	0.0309	0.0400	0.0360
Gumbel	0.0337	0.0411	0.0389
Clayton	0.0297	0.0343	0.0349
Frank	0.0320	0.0400	0.0377
Normal distribution	0.1217	0.1006	0.1217

Table 9: Backtesting results for Value-at-Risk estimation at 0.05 level for 3 portfolios, $w = (1,1)^T$, size of moving window 250, Monte Carlo samples of 10.000 realizations of pseudo random variable. Standardized margins modeled with t-distribution.

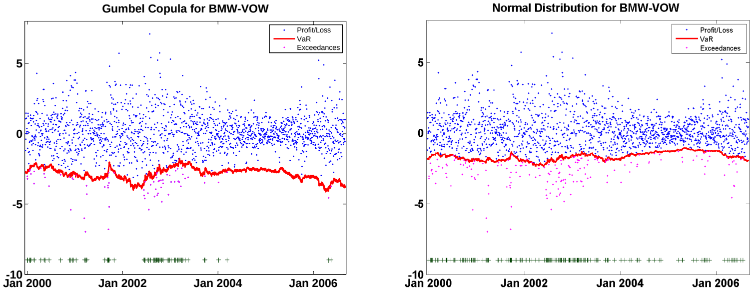


Figure 41: VaR, P&L and exceedances estimated with Gumbel copula ($\widehat{\alpha} = 0.0411$) and bivariate normal distribution ($\widehat{\alpha} = 0.1006$) for BMW and Volkswagen.

MSRvar_cop_backtesting_gumbel MSRvar_cop_backtesting_normal

Copula

Copula

Kopuła

Kopula

关联结构

連辞

الارتباط-الصلة

코플러

Applications of Copulae for the Calculation of Value-at-Risk

Value-at-Risk (VaR) computation: most VaR methods assume a multivariate normal distribution of the risk factors.

Several pitfalls!

Copulae can be used to describe the dependence between two or more random variables with arbitrary marginal distributions. Backtesting often shows that copula produce more accurate results than "correlation dependence".

Copula, ae [latin]

1.
    a) Band, Leine, Koppel;
    b) Enterhaken
2. Verbindung

关联结构

連辞

الارتباط. الصلة

코플러

What is a copula?

A function that links a multidimensional distribution to its one-dimensional margins.

The **joint cumulative distribution functions (cdf)** of d random variables $X_1, \ldots, X_d$ with cdf $F_1, \ldots, F_d$ is:

$$\begin{aligned} P(X_1 \le x_1, \ldots, X_n \le x_d) &= C\{P(X_1 \le x_1), \ldots, P(X_d \le x_d)\} \\ &= C\{F_1(x_1), \ldots, F_d(x_d)\} \end{aligned}$$

Value-at-Risk

1. *value of linear portfolio* $w = (w_1, \ldots, w_d)^\top$ of assets $S_t = (S_{1,t}, \ldots, S_{d,t})^\top$:

$$V_t = \sum_{j=1}^{d} w_j S_{j,t}$$

2. *profit and loss* (P&L) function:

$$\begin{aligned} L_{t+1} &= V_{t+1} - V_t = \sum_{j=1}^{d} w_j S_{j,t}(e^{X_{j,t+1}} - 1) \\ X_{t+1} &= \log S_{t+1} - \log S_t \end{aligned}$$

3. *Value-at-Risk* at level α:

$$VaR(\alpha) = F_L^{-1}(\alpha)$$

Log returns DCX & VW at 20030408

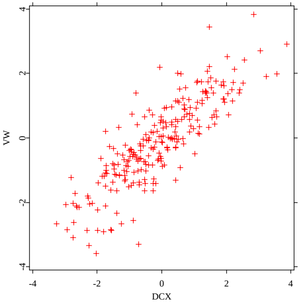


Figure 42: Standardized log returns, DaimlerChrysler (DCX) and Volkswagen (VW), 20020415-20030408

Log returns DCX & VW at 20041027

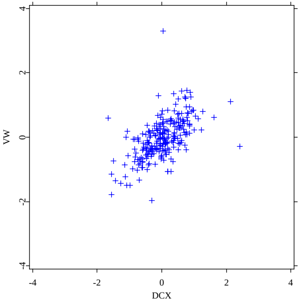


Figure 43: Standardized log returns, DaimlerChrysler (DCX) and Volkswagen (VW), 20031103-20041027

VaR depends on the distribution F_X of log returns $X = (X_1, \ldots, X_d)^\top$.

1. How to model F_X and the dependency among $X_1, \ldots, X_d$?
2. How does F_X and the dependency among $X_1, \ldots, X_d$ vary over time ?

Traditional approach (RiskMetrics)

Log returns conditionally normal

$$X_t \sim N(0, \Sigma_t)$$

Drawbacks from multivariate normal distribution:

1. no heavy-tails
2. joint extreme values relatively infrequent
3. symmetry (elliptical distribution)

Copula based approach

Log returns conditionally distributed with copula C:

$$X_t \sim C\{F_{X_1}(x_1), \ldots, F_{X_d}(x_d); \theta_t\}$$

where $F_{X_1}, \ldots, F_{X_d}$ are marginal distributions and θ_t is the copula (dependence) parameter.

Adaptive Copulae

Global parameter $\theta_t = \theta$, too optimistic.

1. local parametric assumption: θ_t nearly constant on *homogeneity intervals*
2. find largest interval where homogeneity is acceptable ($\omega\rho\alpha\kappa\lambda\epsilon$)
3. for each t, *adaptively* find homogeneity interval

Estimate dependence parameter θ_t in a time varying interval

Local Parametric Assumption

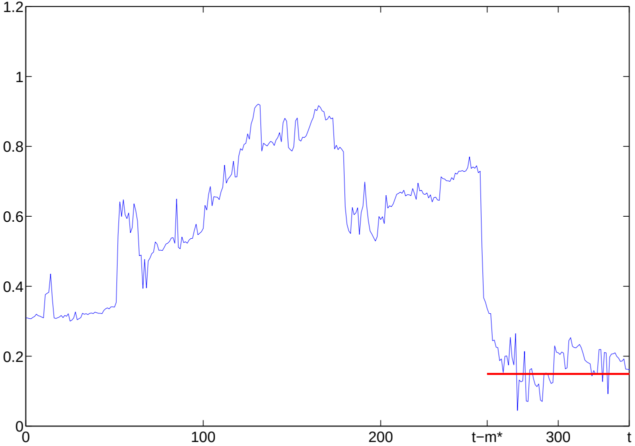


Figure 44: Parameter θ_t (blue), size of homogeneity interval at t (black).

Outline

1. Motivation ✓
2. Copulae and Value-at-Risk
3. Adaptive Copulae
4. Applications

Copulae

1. a copula $C : [0,1]^d \to [0,1]$ is a d-variate distribution with marginal distributions being $U(0,1)$
2. the copula associated with distribution F and its marginals F_j is the distribution C that satisfies

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}$$

Theorem (Sklar's theorem)

For a distribution function F with marginals $F_{X_1}, \ldots, F_{X_d}$. There exists a copula $C : [0,1]^d \to [0,1]$ with

$$F(x_1, \ldots, x_d) = C\{F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\}$$

If $F_{X_1}, \ldots, F_{X_d}$ are cts, then C is unique. If C is a copula and $F_{X_1}, \ldots, F_{X_d}$ are cdfs, then the function F defined in (1) is a joint cdf with marginals $F_{X_1}, \ldots, F_{X_d}$.

With copula density

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}$$

the density function of $F(x_1, \ldots, x_d)$ is

$$f(x_1, \ldots, x_d) = c\{F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\} \prod_{j=1}^{d} f_j(x_j)$$

where $u_j = F_{X_j}(x_j)$ and $f_j(x_j) = F'_{X_j}(x_j)$, $j = 1 \ldots d$

1. Gaussian Copula

$$C_{\Psi}^{Ga}(u_1, \ldots, u_d) = \Phi_{\Psi}\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\}$$

Φ univariate standard normal cdf

Φ_{Ψ} d-dimensional standard normal cdf with correlation matrix Ψ

- Gaussian copula contains *the dependence structure*
- *normal* marginal distributions + Gaussian copula = multivariate normal distributions
- *non-normal* marginal distributions + Gaussian copula = *meta-Gaussian* distributions

Explicit expression for the Gaussian copula

$$
\begin{aligned}
C_{\Psi}^{Ga}(u_1,\dots,u_d) &= \Phi_{\Psi}\{\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d)\} \\
&= \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_d)} 2\pi^{-\frac{d}{2}} \mid \Psi \mid^{-\frac{1}{2}} e^{(-\frac{1}{2}r^{\top}\Psi^{-1}r)} dr_1 \dots dr_d
\end{aligned}
$$

where

$$r = (r_1,\dots,r_d)^{\top},\; u_j = \Phi(x_j)$$

- $C_{\Psi}^{Ga}(u_1,\dots,u_d)$ allows to generate joint symmetric dependence, but no tail dependence (i.e., there are no joint extreme events)

2. Frank Copula, $0 < \theta \leq \infty$

$$C_\theta(u_1, \ldots, u_d) = -\frac{1}{\theta} \log \left[1 + \frac{\prod_{j=1}^{d} \{\exp(-\theta u_j) - 1\}}{\{\exp(-\theta) - 1\}^{d-1}} \right]$$

- dependence becomes maximal when $\theta \longrightarrow \infty$
- independence is achieved when $\theta \longrightarrow 0$

3. Gumbel-Hougaard copula, $1 \leq \theta \leq \infty$

$$C_\theta(u_1, \ldots, u_d) = \exp\left[-\left\{\sum_{j=1}^{d}(-\log u_j)^\theta\right\}^{\theta^{-1}}\right]$$

- for $\theta > 1$ allows to generate dependence in the upper tail (Schmidt, 2005)
- For $\theta = 1$ reduces to the product copula, i.e. $C_\theta(u_1, \ldots, u_d) = \prod_{j=1}^{d} u_j$.
- for $\theta \longrightarrow \infty$, we obtain the Fréchet-Hoeffding upper bound:

$$C_\theta(u_1, \ldots, u_d) \stackrel{\theta\to\infty}{\longrightarrow} \min(u_1, \ldots, u_d).$$

4. Ali-Mikhail-Haq copula, $-1 \leq \theta < 1$

$$C_\theta(u_1, \dots, u_d) = \frac{\prod_{j=1}^{d} u_j}{1 - \theta \left\{ \prod_{j=1}^{d} (1 - u_j) \right\}}$$

- independence is achieved when $\theta = 0$
- the Fréchet-Hoeffding bounds are not achieved

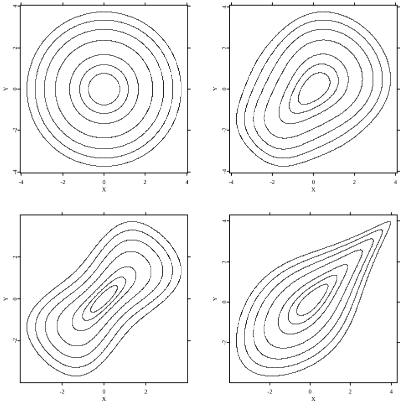


Figure 45: Pdf contour plots, $F(x_1, x_2) = C\{\Phi(x_1), \Phi(x_2)\}$ with Gaussian ($\rho = 0$), AMH ($\theta = 0.9$), Frank ($\theta = 8$), Gumbel ($\theta = 2$) copulae.

5. Clayton copula, $\theta > 0$

$$C_\theta(u_1, \dots, u_d) = \left\{ \left(\sum_{j=1}^{d} u_j^{-\theta} \right) - d + 1 \right\}^{-\theta^{-1}}$$

- dependence becomes maximal when $\theta \longrightarrow \infty$
- independence is achieved when $\theta \longrightarrow 0$
- the distribution tends to the lower Fréchet-Hoeffding bound when $\theta \longrightarrow 1$
- allows to generate asymmetric dependence and lower tail dependence, but no upper tail dependence

Copula estimation

The distribution of $X = (X_1, ..., X_d)^\top$ with marginals $F_{X_j}(x_j, \delta_j)$, $j = 1, ..., d$ is given by:

$$F_X(x_1, \ldots, x_d) = C\{F_{X_1}(x_1; \delta_1), \ldots, F_{X_d}(x_d; \delta_d); \theta\}$$

and its density is given by

$$f(x_1, \ldots, x_d; \delta_1, \ldots, \delta_d, \theta)$$

$$= c\{F_{X_1}(x_1; \delta_1), \ldots, F_{X_d}(x_d; \delta_d); \theta\} \prod_{j=1}^{d} f_j(x_j; \delta_j)$$

where c is a copula density.

For a sample of observations $\{x_t\}_{t=1}^T$ and $\vartheta = (\delta_1, \dots, \delta_d, \theta)^\top \in \mathbb{R}^{d+1}$ the likelihood function is

$$L(\vartheta; x_1, \dots, x_T) = \prod_{t=1}^{T} f(x_{1,t}, \dots, x_{d,t}; \delta_1, \dots, \delta_d, \theta)$$

and the corresponding log-likelihood function

$$\ell(\vartheta; x_1, \dots, x_T) = \sum_{t=1}^{T} \log c\{F_{X_1}(x_{1,t}; \delta_1), \dots, F_{X_d}(x_{d,t}; \delta_d); \theta\}$$

$$+ \sum_{t=1}^{T} \sum_{j=1}^{d} \log f_j(x_{j,t}; \delta_j)$$

Full Maximum Likelihood (FML)

- FML estimates vector of parameters ϑ in one step through

$$\widetilde{\vartheta}_{FML} = \arg\max_{\vartheta} \ell(\vartheta).$$

- the estimates $\widetilde{\vartheta}_{FML} = (\widetilde{\delta}_1, \ldots, \widetilde{\delta}_d, \widetilde{\theta})^\top$ solve

$$(\partial\ell/\partial\delta_1, \ldots, \partial\ell/\partial\delta_d, \partial\ell/\partial\theta) = 0.$$

- Drawback: with an increasing dimension the algorithm becomes too burdensome computationally.

Inference for Margins (IFM)

1. estimate parameters δ_j from the marginal distributions:

$$\widehat{\delta}_j = \arg\max_{\delta} \ell_j(\delta_j) = \arg\max_{\delta} \left\{ \sum_{t=1}^{T} \log f_j(x_{j,t}; \delta_j) \right\}$$

2. estimate the dependence parameter θ by maximizing the *pseudo log-likelihood* function

$$\ell(\theta, \widehat{\delta}_1, \ldots, \widehat{\delta}_d) = \sum_{t=1}^{T} \log c\{F_{X_1}(x_{1,t}; \widehat{\delta}_1), \ldots, F_{X_d}(x_{d,t}; \widehat{\delta}_d); \theta\}$$

- The estimates $\widehat{\vartheta}_{IFM} = (\widehat{\delta}_1, \ldots, \widehat{\delta}_d, \widehat{\theta})^\top$ solve

$$(\partial\ell_1/\partial\delta_1, \ldots, \partial\ell_d/\partial\delta_d, \partial\ell/\partial\theta) = 0.$$

- Advantage: numerically stable.

Canonical Maximum Likelihood (CML)

- CML maximizes the *pseudo log-likelihood* function with *empirical* marginal distributions

$$\ell(\theta) = \sum_{t=1}^{T} \log c\{\widehat{F}_{X_1}(x_{1,t}), \ldots, \widehat{F}_{X_d}(x_{d,t}); \theta\}$$

$$\widehat{\vartheta}_{CML} = \arg\max_{\theta} \ell(\theta)$$

where

$$\widehat{F}_{X_j}(x) = \frac{1}{T+1} \sum_{t=1}^{T} \mathbf{1}\{X_j, t \leq x\}$$

- Advantage: no assumptions about the parametric form of marginal distributions.

Value-at-Risk with Copulae

The process $\{X_t\}_{t=1}^T$ of log-returns can be modelled as

$$X_{j,t} = \sigma_{j,t}\varepsilon_{j,t}$$

with

$$\sigma_{j,t}^2 = E[X_{j,t}^2 \mid \mathcal{F}_{t-1}]$$

where $\mathcal{F}_t$ is the available information at time t.

The standardized innovations

$$\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{d,t})^\top$$

are independent with distribution function

$$F(\varepsilon_t) = C\{F_1(\varepsilon_1), \dots, F_d(\varepsilon_d); \theta_t\}$$

where

1. C is copula with parameter θ
2. ε_j have continuous marginal distributions F_j, $j = 1, \dots, d$

VaR Estimation with Copulae

For log-returns $\{x_{j,t}\}_{t=1}^{T}$, $j = 1, \dots, d$, estimation of VaR at level α:

1. determine innovations $\widehat{\varepsilon}_t$ (e.g. by deGARCHing)
2. specificy and estimate marginal distributions $F_j(\widehat{\varepsilon}_j)$
3. specificy a copula C and estimate dependence parameter θ
4. simulate innovations ε and losses L
5. determine $\widehat{VaR}(\alpha)$, the empirical α-quantile of F_L.

Moving Window

- use static windows of size $w = 250$ scrolling in time t for VaR estimation:

$$\{x_t\}_{t=s-w+1}^{s}$$

for $s = w, \dots, T$

- the VaR estimation procedure generates a time series $\{\widehat{VaR}_t\}_{t=w}^{T}$ and $\{\widehat{\theta}_t\}_{t=w}^{T}$ of dependence parameters estimates.

Adaptive Copula Estimation I

Estimation of dependence parameter by moving window does not fine tune changes in dependency

1. window too small: high variability of estimator
2. window too large: poor sensitivity to changes in parameter, high delay in detections

Adaptive Copula Estimation II

1. *"oracle" choice*: largest interval $I = [t_0 - m_{k^*}, t_0]$ where small modelling bias condition (SMB)

$$\Delta_I(\theta) = \sum_{t \in I} \mathcal{K}(P_{\theta_t}, P_\theta) \leq \Delta$$

   where

$$\mathcal{K}(\vartheta, \vartheta') = E_\vartheta \log \frac{p(y, \vartheta)}{p(y, \vartheta')}$$

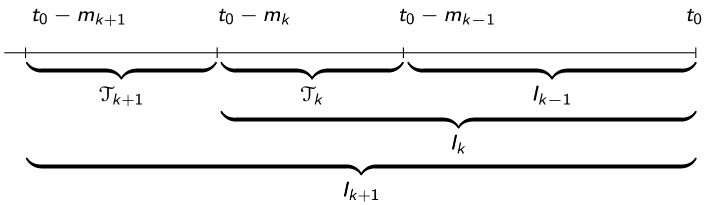
   holds
2. θ is ideally estimated from $I = [t_0 - m_{k^*}, t_0]$

Adaptive Copula Estimation III

1. *"oracle" choice* depends on unknown parameter θ_t
2. adaptively estimate largest interval where homogeneity hypothesis is accepted
3. *Local Change Point detection (LCP)*, Mercurio, Spokoiny (2004): sequentially test θ_t is constant (i.e. $\theta_t = \theta$) within some interval I
4. complete theory in Spokoiny (2007)

Local Change Point procedure (LCP)

1. define family of nested intervals $I_0 \subset I_1 \subset I_2 \subset \ldots \subset I_K$ with length m_k as $I_k = [t_0 - m_k, t_0]$
2. $\mathcal{T}_k = I_k \setminus I_{k-1} = [t_0 - m_k, t_0 - m_{k-1}]$



LCP

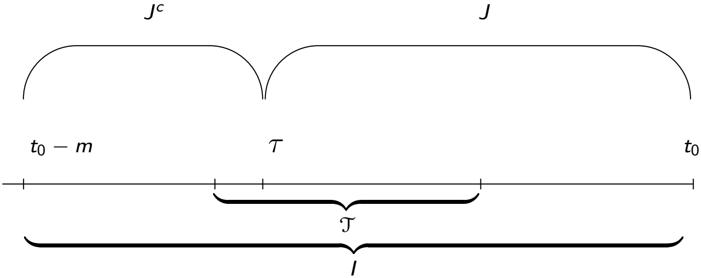
Start with $k = 1$ and

1. test homogeneity hypothesis $H_{0,k}$ against change point alternative within $\mathcal{T}_k$ using testing interval I_{k+1};
2. if no change points were found in $\mathcal{T}_k$, accept I_k. Take next interval $\mathcal{T}_{k+1}$ and repeat the previous step until homogeneity is rejected or the largest possible interval $I_K = [t_0 - m_K, t_0]$ is accepted;
3. if $H_{0,k}$ is rejected for $\mathcal{T}_k$, the estimated interval of homogeneity is last accepted interval $\widehat{I} = I_{k-1}$.
4. if the largest possible interval I_K is accepted, $\widehat{I} = I_K$.

Test of homogeneity against a change point alternative

Interval $I = [t_0 - m, t_0]$, $\mathcal{T} \subset I$

$$H_0 : \forall \tau \in \mathcal{T}, \quad \theta_t = \theta \ \forall t \in J = [\tau, t_0], \quad \forall t \in J^c = I \setminus J$$
$$H_1 : \exists \tau \in \mathcal{T}, \quad \theta_t = \theta_1 \ \forall t \in J, \quad \theta_t = \theta_2 \neq \theta_1 \ \forall t \in J^c$$



Likelihood ratio test for the fixed change point location:

$$
\begin{aligned}
T_{I,\tau} &= \max_{\theta_1,\theta_2} \{L_J(\theta_1) + L_{J^c}(\theta_2)\} - \max_{\theta} L_I(\theta) \\
&= \widehat{L}_J + \widehat{L}_{J^c} - \widehat{L}_I
\end{aligned}
$$

Test statistics for unknown change point location:

$$T_I = \max_{\tau \in \mathcal{T}_I} T_{I,\tau}$$

Reject H_0 if $T_I > \mathfrak{z}$

For critical values $\mathfrak{z}_1, \ldots, \mathfrak{z}_K$ define the random sets

1. $$\mathcal{C}_k = \{ T_{I_k} \le \mathfrak{z}_k \}$$

2. I_k accepted:
$$\mathcal{A}_k = \mathcal{C}_1 \cap \ldots \cap \mathcal{C}_k$$

3. I_k accepted, I_{k+1} rejected:
$$\mathcal{B}_k = \mathcal{A}_k \backslash \mathcal{C}_{k+1}$$

Adaptive Estimator $\widehat{\theta}$

1. adaptive estimator $\widehat{\theta}_k$ at step k

$$\widehat{\theta}_k(\mathfrak{z}_1,\ldots,\mathfrak{z}_k) = \widetilde{\theta}_k \mathbf{1}(\mathcal{A}_k) + \sum_{l=1}^{k-1} \widetilde{\theta}_l \mathbf{1}(\mathcal{B}_l)$$

2. adaptive estimator $\widehat{\theta}$

$$\widehat{\theta} = \widehat{\theta}_K$$

where $\widetilde{\theta}_k$ is weak (MLE) estimator at I_k

Choice of Critical Values $\mathfrak{z}$

Desirable in the homogeneous situation

$$\widehat{\theta}_k = \widetilde{\theta}_k$$

"False alarm": $\widehat{I}_k \subset I_k$, the estimated homogeneity interval is too small and

$$L_{I_k}(\widetilde{\theta}_k, \widehat{\theta}_k) > 0$$

Select $\mathfrak{z}_1, \ldots, \mathfrak{z}_K$ such that risk associated with "false alarm" is bounded.

Choice of Critical Values $\mathfrak{z}$

The critical values $\mathfrak{z}_1, \ldots, \mathfrak{z}_K$ are the minimal values providing

$$\mathbf{E}_{\theta^*} \left| L_{I_k}(\widetilde{\theta}_k, \widehat{\theta}_k) \right|^{1/2} \le \rho \mathfrak{R}(\theta^*), \qquad k = 2, \ldots, K, \qquad \theta^* \in \Theta \quad (15)$$

where $0 < \rho \le 1$ and $\mathfrak{R}(\theta^*)$ is the risk of the non-adaptive estimate $\widetilde{\theta}_k$:

$$\mathfrak{R}(\theta^*) = \max_{k \ge 1} \mathbf{E}_{\theta^*} \left| L_{I_k}(\widetilde{\theta}_k, \theta^*) \right|^{1/2}.$$

Sequential Choice of $\mathfrak{z}$

Sequentially select $\mathfrak{z}_l$ for $l = 0, \ldots, K-1$ such that

$$\max_{k>l\geq 0} \mathbf{E}_{\theta^*} \left| L_{I_k}(\widetilde{\theta}_k, \widehat{\theta}_l) \right|^{1/2} \mathbf{1}(\mathcal{B}_l) \quad \leq \quad \frac{\rho \mathfrak{R}(\theta^*)}{K-1} \tag{16}$$

by Monte Carlo simulation under $H_0 : \theta_t = \theta^*, \forall t \in I_K$.

Simulated Examples
1. Clayton copula: sudden jump in dependence

Simulated sets of observations from 6-dimensional Clayton copula with parameter

$$\theta_t = \begin{cases} \vartheta_a & \text{if} \quad -390 \leq t \leq 10 \\ \vartheta_b & \text{if} \quad 10 < t \leq 210 \end{cases}$$

for different values of the pair $(\vartheta_a, \vartheta_b)$.

Selection of intervals I and $\mathcal{T}_I$

1. set of numbers m_k defining the length of I_k and $\mathcal{T}_k$ in form of a geometric grid.
2. $m_k = [m_0 c^k]$ for $k = 1, 2, \ldots, K$, $m_0 = 20$ and $c = 1.25$ where $[x]$ means the integer part of x.
3. $I_k = [t_0 - m_k, t_0]$ and $\mathcal{T}_k = [t_0 - m_k, t_0 - m_{k-1}]$ for $k = 1, 2, \ldots, K$.

Critical Values

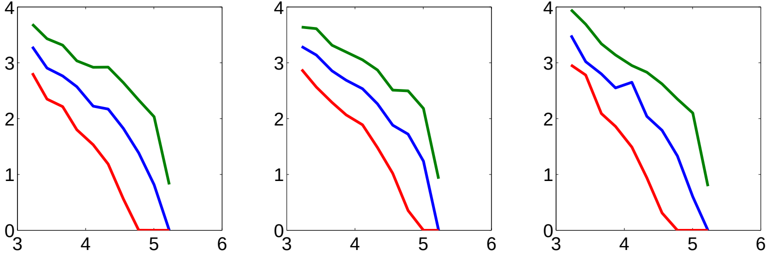


Figure 46: Critical values $\mathfrak{z}_k(\rho, \theta^*)$ (vertical axis), $log(m_k)$ (horizontal axis) for $\rho = 0.2, 0.5$ and 1.0 (top to bottom), $\theta^* = 0.5, 1.0$ and 1.5 (left to right). Based on 5000 simulations from Clayton copula, $m_0 = 20$, $c = 1.25$, $k = 1, \ldots, 11$

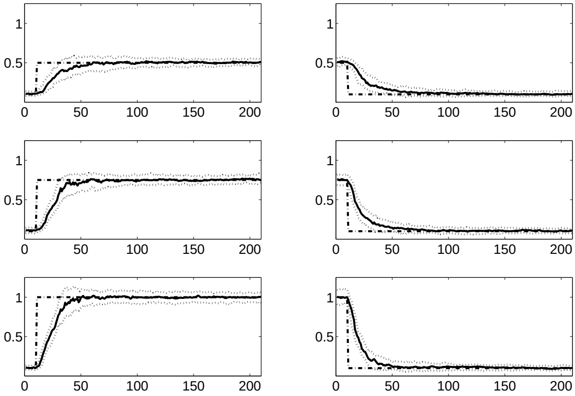


Figure 47: Pointwise median (full), 0.25 and 0.75 quantiles (dotted) from $\widehat{\theta}_t$. True parameter θ_t (dashed) with $\vartheta_a = 0.10$, $\vartheta_b = 0.50$, 0.75 and 1.00 (left, top to bottom) and $\vartheta_b = 0.10$, $\vartheta_a = 0.50$, 0.75 and 1.00 (right, top to bottom). Based on 100 simulations from Clayton copula, estimated with LCP, $m_0 = 20$, $c = 1.25$ and $\rho = 0.5$

Detection Delay

Detection delay δ at rule $r \in [0,1]$ to jump $\gamma = \theta_t - \theta_{t-1}$ at t

$$\delta(t, \gamma, r) = \min\{u \geq t : \widehat{\theta}_u = \theta_{t-1} + r\gamma\} - t$$

1. number of steps for estimated parameter to reach r-fraction of jump in real parameter
2. proportional to type II error (accept homogeneity when jump), decreasing in power and increasing in $\mathcal{K}(H_0, H_1)$.

Detection Delay and Clayton Copula

For Clayton copula

1. KL is asymmetric: $\theta_0 < \theta_1$, $\mathcal{K}_d(\theta_0, \theta_1) > \mathcal{K}_d(\theta_1, \theta_0)$
2. detection delay δ is decreasing in $|\gamma|$ and higher for downward than for upward jumps.

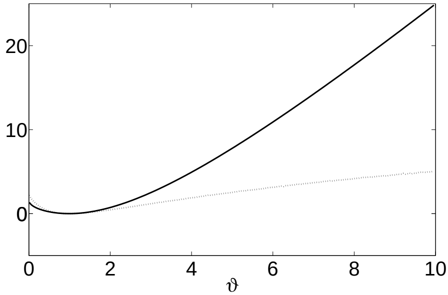


Figure 48: Kullback-Leibler divergences $\mathcal{K}(0.10, \vartheta)$ (full) and $\mathcal{K}(\vartheta, 0.10)$ (dashed), 6-dimensional Clayton copula

Detection Delay Statistics

$(\vartheta_a, \vartheta_b)$	r	mean	std dev.	max	min
(0.50, 0.10)	0.25	9.06	7.28	56	0
	0.50	13.64	9.80	60	0
	0.75	21.87	14.52	89	3
(0.75, 0.10)	0.25	5.16	4.24	21	0
	0.50	8.85	5.55	25	0
	0.75	16.72	10.37	64	3
(1.00, 0.10)	0.25	4.47	2.94	12	0
	0.50	7.94	4.28	22	0
	0.75	14.79	7.38	62	5
(0.10, 0.50)	0.25	8.94	6.65	36	0
	0.50	14.21	9.06	53	0
	0.75	21.43	12.15	68	0
(0.10, 0.75))	0.25	9.00	4.80	25	0
	0.50	14.30	5.96	40	3
	0.75	21.00	10.97	75	6
(0.10, 1.00)	0.25	7.39	3.67	19	0
	0.50	13.10	4.13	22	2
	0.75	20.13	7.34	55	10

Table 10: Statistics for detection delay δ at rule r, based on 100 simulations from Clayton copula, $m_0 = 20$, $c = 1.25$ and $\rho = 0.5$

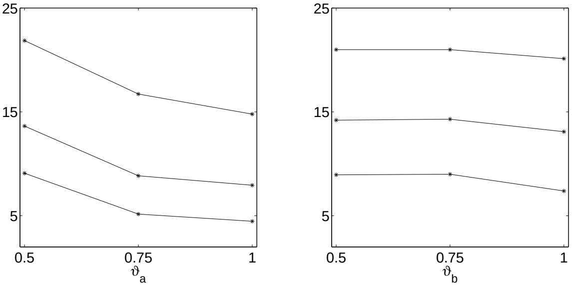


Figure 49: Mean detection delays (crosses) at $r = 0.75$, 0.50 and 0.25 from top to bottom. Left: $\vartheta_b = 0.10$(upward jump). Right: $\vartheta_a = 0.10$ (downward jump), Clayton Copula

2. Clayton copula: smooth change in dependence

Simulated sets of observations from 6-dimensional Clayton copula with parameter

$$\theta_t = \begin{cases} \vartheta_a & \text{if} \quad -350 \le t \le 50 \\ \vartheta_a + \frac{t-50}{100}(\vartheta_b - \vartheta_a) & \text{if} \quad 50 < t \le 150 \\ \vartheta_b & \text{if} \quad 150 < t \le 350 \end{cases}$$

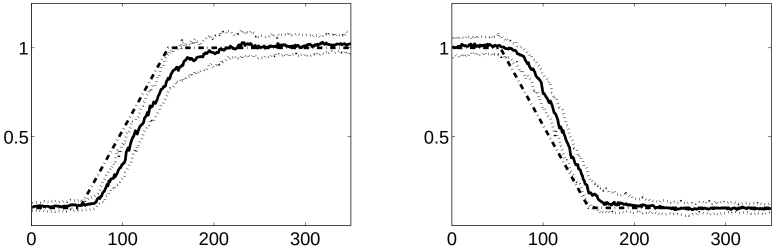


Figure 50: Pointwise median (full), 0.25, 0.75 quantiles (dotted) of estimated parameter $\widehat{\theta}_t$, true parameter θ_t (dashed). Based on 100 simulations, Clayton copula, $(\vartheta_a, \vartheta_b) = (1.00, 0.10)$, $m_0 = 20$ and $c = 1.25$

Applications: Value-at-Risk

Estimation of VaR from portfolios composed of 6 DAX stocks. Use 2 groups:

- group 1: Volkswagen, DaimlerChrysler, Allianz, Münchener Rückversicherung, Bayer and BASF (*high concentration*)
- group 2: Siemens, E.ON, ThyssenKrupp, Lufthansa, Schering and Henkel (*low concentration*)
- closing daily prices from 01.01.2000 to 31.12.2004 (1270 observations)
- data available in `http://sfb649.wiwi.hu-berlin.de/fedc`

VaR Estimation

1. log-returns from stock j modelled by

$$X_{t,j} = \sigma_{t,j}\varepsilon_{t,j}$$

2. $\sigma_{t,j}^2$ estimated at time t by exponential smoothing with $\lambda = \frac{1}{20}$

$$\widehat{\sigma}_{t,j}^2 = (e^{\lambda} - 1)\sum_{s<t} e^{-\lambda(t-s)} X_{s,j}^2$$

3. empirical distribution of the obtained residuals $\widehat{\varepsilon}_{t,j}$
4. 6-dim copula belongs to Clayton family

Backtesting

The *exceedances ratio* at level α for a portfolio w is given by

$$\widehat{\alpha}_w(\alpha) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{l_t < \widehat{VaR}_t(\alpha)\}$$

where $\{l_t\}$ are realizations of the P&L function.

VaR Estimation: Performance

Calculate $\widehat{\alpha}_w(\alpha)$ using *RiskMetrics* (RM) and time varying copula procedures moving window (MW) and Local Change Point (LCP)

1. for set $\mathcal{W} = \{w^*, w_n; n = 1, \ldots, 100\}$ of portfolios
2. $w^* = (w_1^*, \ldots, w_6^*)^\top$, is the equally weighted portfolio $w_i^* = \frac{1}{6}$, $i = 1, \ldots, 6$.
3. w_n is a realization of a random variable uniformly distributed on $\mathcal{S} = \{(x_1, \ldots, x_6) \in \mathbb{R}^6 : \sum_{i=1}^6 x_i = 1, x_i \geq 0\}$

VaR Estimation: Performance

1. average exceedance ratio ($A^{\mathcal{W}}$)

$$A^{\mathcal{W}} = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \widehat{\alpha}_w$$

2. squared sum of differences to level ($S^{\mathcal{W}}$)

$$S^{\mathcal{W}} = \sum_{w \in \mathcal{W}} (\widehat{\alpha}_w - \alpha)^2$$

3. relative distance to level ($D^{\mathcal{W}}$)

$$D^{\mathcal{W}} = \frac{1}{\alpha} \left\{ \sum_{w \in \mathcal{W}} (\widehat{\alpha}_w - \alpha)^2 \right\}^{1/2}$$

The performances in VaR estimation are compared based on $D^{\mathcal{W}}$.

Results Group 1

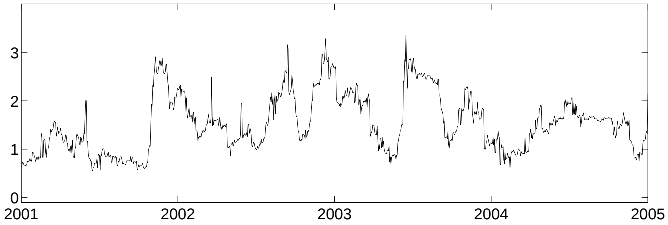


Figure 51: Estimated copula parameter $\widehat{\theta}_t$ for group 1, LCP method, $m_0 = 20$, $c = 1.25$ and $\rho = 0.5$, Clayton copula

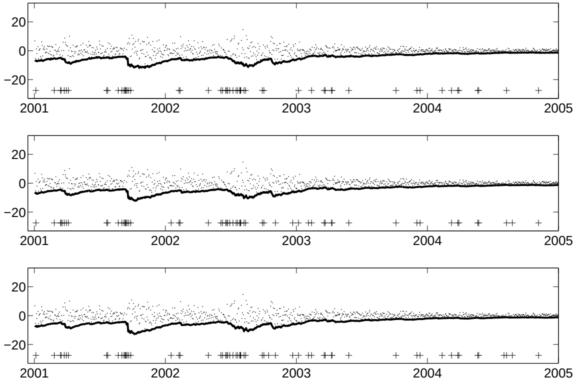


Figure 52: P&L (dots), Value-at-Risk at level $\alpha = 0.05$ (line), exceedances (crosses), estimated with LCP (above), MW (middle) and RM (below), for equally weighted portfolio w^*, group 1

	RM		MW		LCP		
	$\alpha(\times 10^2)$						
	5.00	1.00	5.00	1.00	5.00	1.00	
$\widehat{\alpha}_{w^*}$	6.21	1.28	5.81	0.69	5.52	0.69	
$\widehat{\alpha}_{w_1}$	6.40	1.28	6.40	0.69	6.31	0.79	
$\widehat{\alpha}_{w_2}$	5.32	1.58	5.91	0.99	5.62	0.99	$(\times 10^2)$
$A^{\mathcal{W}}$	5.92	1.42	5.52	0.64	5.41	0.66	
$S^{\mathcal{W}}$	0.94	0.20	0.51	0.15	0.38	0.14	
$D^{\mathcal{W}}$	1.92	4.52	1.43	3.83	1.23	3.74	

Table 11: Exceedance ratios for portfolios w^*, w_1 and w_2, average, sum of squared differences and relative distance to level across levels and methods, group 1

Results Group 2

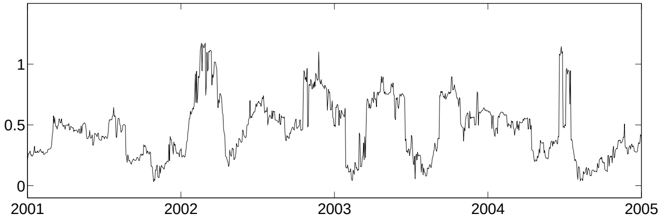


Figure 53: Estimated copula parameter $\widehat{\theta}_t$ for group 2, LCP method, $m_0 = 20$, $c = 1.25$ and $\rho = 0.5$, Clayton copula

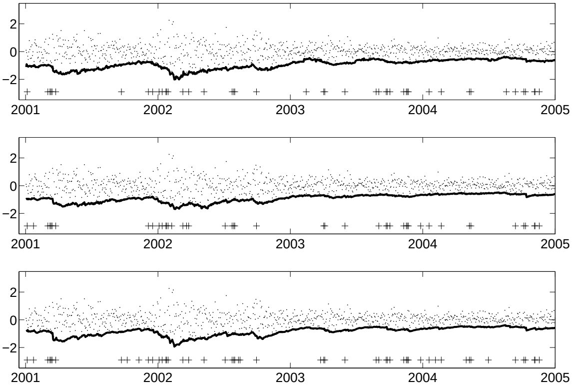


Figure 54: P&L (dots), Value-at-Risk at level $\alpha = 0.05$ (line), exceedances (crosses), estimated with LCP (above), MW (middle) and RM (below) for equally weighted portfolio w^*, group 2

	RM		MW		LCP		
	$\alpha(\times 10^2)$						
	5.00	1.00	5.00	1.00	5.00	1.00	
$\widehat{\alpha}_{w^*}$	5.42	1.58	4.53	0.39	4.53	0.30	
$\widehat{\alpha}_{w_1}$	6.01	1.67	5.22	0.69	5.02	0.69	
$\widehat{\alpha}_{w_2}$	5.22	1.48	4.93	0.69	4.63	0.59	$(\times 10^2)$
$A^{\mathcal{W}}$	5.43	1.50	4.53	0.55	4.43	0.51	
$S^{\mathcal{W}}$	0.34	0.28	0.55	0.22	0.58	0.26	
$D^{\mathcal{W}}$	1.17	5.32	1.49	4.73	1.53	5.10	

Table 12: Exceedance ratios for portfolios w^*, w_1 and w_2, average, sum of squared differences and relative distance to level across levels and methods, group 2

Expectation-Maximization Algorithm

The Expectation-Maximization (EM) algorithm is a general algorithm for maximum-likelihood estimation where the model depends on latent variables or where the data are incomplete.

Algorithm

1. Expectation step (E-step) computes the conditional expectation of the complete data likelihood given the current estimates of model parameters and the observed data.
2. Maximization step (M-step) re-estimates all the parameters by maximizing the expected likelihood.
3. The parameters found on the M-step are then used to begin another E-step. The process is repeated until the likelihood converges, i.e reaching a local maxima.

History

The EM algorithm was explained and given its name in 1977 in a paper by Arthur Dempster, Nan Laird and Donald Rubin *Maximum likelihood from incomplete data via the EM algorithm* in the Journal of the Royal Statistical Society.

Notation

$\mathbf{y}$	the incomplete (observed) data
$\mathbf{z}$	the latent variables

The complete data consist of the observed and missing variables.

$p(\mathbf{y}, \mathbf{z} \vert \theta)$	the joint probability density function (continuous case) or probability mass function (discrete case) of the complete data with parameters given by the vector θ

Conditional Distribution

Using the Bayes rule and the law of total probability the conditional distribution of the missing data given the observed can be expressed as:

$$p(\mathbf{z}|\mathbf{y},\theta) = \frac{p(\mathbf{y},\mathbf{z}|\theta)}{p(\mathbf{y}|\theta)} = \frac{p(\mathbf{y}|\mathbf{z},\theta)p(\mathbf{z}|\theta)}{\int p(\mathbf{y}|\widehat{\mathbf{z}},\theta)p(\widehat{\mathbf{z}}|\theta)d\widehat{\mathbf{z}}}.$$

E-step

The aim is to maximize the incomplete data likelihood $\log p(\mathbf{y}|\theta)$ through maximizing the expected complete data likelihood (since it is much easier to maximize) where expectation is taken over all possible values of the hidden variables:

$$\mathrm{Q}(\theta) = \mathbf{E}_{\mathbf{z}}\{\log p(\mathbf{y}, \mathbf{z}|\theta)|\mathbf{y}\}.$$

Q-function

Discrete case:

$$\mathrm{Q}(\theta) = \sum_{\mathbf{z}} p(\mathbf{z}|\mathbf{y},\theta) \log p(\mathbf{y},\mathbf{z}|\theta)$$

Continuous case:

$$\mathrm{Q}(\theta) = \int_{-\infty}^{\infty} p(\mathbf{z}|\mathbf{y},\theta) \log p(\mathbf{y},\mathbf{z}|\theta) d\mathbf{z}$$

M-step

EM starts with some initial guess of the parameter values θ_0 and then iteratively searches for better values for the parameters. Assume that the current estimate of the parameters is θ_n, our goal is to find another θ_{n+1} that can improve the expected value of the log-likelihood $\mathrm{Q}(\theta)$

$$\theta_{n+1} = \arg\max_{\theta} \mathrm{Q}(\theta)$$

Convergence

Convergence may be determined by examining when the parameters quit changing:

$$\|\theta_{n+1} - \theta_n\| < \epsilon$$

for some ϵ and measure $\|\cdot\|$.
EM algorithm does not guarantee that the convergence will be to a global maximum. For multimodal distributions, convergence will be to a local maximum which depends on the initial starting point θ_0.

Applications

- genetics: observed data (the phenotype) is a function of unobserved data (the genotype),
- signal processing: image reconstructions, pattern recognition,
- medical image reconstruction: tomography,
- spectroskopy,
- psychometrics: the item response theory models,
- portfolio risk management,
- speach recognition.

Mixture Copula

Let $C_1, \ldots, C_K$ be d-dimensional copulae and $w_1, \ldots, w_K$ positive weights so that $w_1 + \ldots + w_K = 1$.
Mixture copula:

$$C_{mix}(u_1, \ldots, u_d) = \sum_{i=1}^{K} w_i C_i(u_1, \ldots, u_d).$$

Mixture Copula

Joint density of $(Y_1, \ldots, Y_d)^\top$

$$f(\mathbf{y}_t|\psi) = \sum_{i=1}^{K} w_i f_i(\mathbf{y}_t|\theta_i),$$

where $\psi = (w_i, \theta_i)_{i=1}^{K}$, $t = 1, \ldots, T$ and

$$f_i(\mathbf{y}_t|\theta_i) = c_i\{F_1(y_{1,t}), \ldots, F_d(y_{d,t}); \theta_i\} \prod_{j=1}^{d} g_j(y_{j,t}),$$

where g_j, $j = 1, \ldots, d$, are densities of marginal distributions.

Indicator Variables

The log-likelihood function

$$\ell(\psi) = \sum_{t=1}^{T} \log \sum_{i=1}^{K} w_i f_i(\mathbf{y}_t|\theta_i).$$

When we observe data y_t we do not know from which component density f_i they were generated.
The indicator variables $\{z_t\}_{t=1}^{T}$

$$z_t = i \text{ when } y_t \text{ was generated from } f_i.$$

Q-function for Mixture Copula Model

$$
\begin{aligned}
\mathrm{Q}(\theta) &= \mathbf{E}_{\mathbf{z}}\{\log f(\mathbf{y},\mathbf{z}|\theta)|\mathbf{y}\} = \mathbf{E}_{\mathbf{z}}\{\log \prod_{t=1}^{T} f(\mathbf{y}_t,\mathbf{z}_t|\theta)|\mathbf{y}_t\} \\
&= \sum_{t=1}^{T}\sum_{i=1}^{K} f(\mathbf{z}_t = i|\mathbf{y}_t,\theta)\{\log c_i(\mathbf{u}_t;\theta_i) \\
&\quad + \sum_{j=1}^{d} \log g_j(y_{j,t}) + \log f(\mathbf{z}_t = i|\theta)\}.
\end{aligned}
$$

Lagrangian

To find the extremes of function $\mathrm{Q}(\theta)$ we add a Lagrange multiplier and define the Lagrangian:

$$J(\theta) = \sum_{t=1}^{T}\sum_{i=1}^{K} f(\mathbf{z}_t = i|\mathbf{y}_t, \theta)\{\log c_i(\mathbf{u}_t; \theta_i) + \sum_{j=1}^{d} \log g_j(y_{j,t}) + \log f(\mathbf{z}_t = i|\theta)\} - \lambda\{1 - \sum_{i=1}^{K} f(\mathbf{z}_t = i|\theta)\},$$

where λ is an unknown scalar.

E-step and Mixture Copula Model

The E-step computes and updates the conditional probability that our observation was from each component copula. Computing $(n+1)$-th time the iteration of weights we use values of parameters calculated in n-th estimation step

$$w_i^{n+1} = \frac{1}{T}\sum_{t=1}^{T}\frac{w_i^n c_i^n(\mathbf{u}_t;\theta_i^n)}{\sum_{k=1}^{K} w_k^n c_k^n(\mathbf{u}_t;\theta_k^n)}.$$

M-step and Mixture Copula Model

For any given estimates of weights w_i^n dependence parameters θ_i^n in our mixed copula model are not available in closed form. Consequently, in order to obtain the final results, we calculate parameters of the copulae using the complete-data maximum likelihood estimation.

Drawback of the Algorithm

The estimates are sensitive to initial values and the standard approach can lead us only to a local maximum.
Possible solutions of this problem:

- we try many different initial values and choose the solution that has the highest likelihood value,
- we use a simpler model to determine an initial value for more complex models.

Application

Example Consider daily standardized stock log-returns of BMW and Siemens. Time period: 1st January 1999 – 1st September 2006, 2000 data points. We fit to the data two mixture copulae. The margins are modeled with t-distribution.
The parameters estimated by the EM Algorithm:

Mixture	w	θ_1	θ_2
Gumbel-Clayton	0.109	3.92	$7 \cdot 10^{-18}$
Frank-Clayton	0.496	12.3	$2 \cdot 10^{-16}$

References

X. Chen and Y. Fan
Estimation and Model Selection of Semiparametric Copula-Based Multivariate Dynamic Models Under Copula Misspecification?
Journal of Econometrics, 135:125-154, 2006.

X. Chen and Y. Fan
Estimation of Copula-Based Semiparametric Time Series Models
Journal of Econometrics, 130:307-335, 2006.

X. Chen, Y. Fan and V. Tsyrennikov
Efficient Estimation of Semiparametric Multivariate Copula Models
Journal of the American Statistical Association, 101:1228-1240, 2006.

A. Dias and P. Embrechts
Dynamic Copula Models for Multivariate High-Frequency Data in Finance
working paper, 2004.

V. Durrleman, A. Nikeghbali and T. Roncalli
Which Copula is the Right One?
Groupe de Recherche Opérationnelle Crédit Lyonnais, 2000.

P. Embrechts, F. Lindskog and A. McNeil
Modelling Dependence with Copulas and Application to Risk Management
working paper, 2001.

P. Embrechts, A. McNeil and D. Straumann
Correlation and Dependence in Risk Management: Properties and Pitfalls
Risk Management: Value at Risk and Beyond, 1999.

J. Franke, W. Härdle and C. Hafner
Statistics of Financial Markets
Springer-Verlag, 2004.

W. Hoeffding
Masstabinvariante korrelationstheorie
Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin, 5(3):179-233, 1940.

E. Giacomini and W. Härdle
Nonparametric Risk Management with Adaptive Copulae
55th ISI, Sydney, 2005.

W. Härdle, N. Hautsch and L. Overbeck
Applied Quantitative Finance
Springer-Verlag, 2009.

W. Härdle, H. Herwartz and V. Spokoiny
Time Inhomogeneous Multiple Volatility Modeling
Journal of Financial Econometrics, 1(1): 55-95, 2003.

H. Joe
Multivariate Models and Dependence Concepts
Chapman & Hall, 1997.

F. Longin
From Value At Risk To Stress Testing: The Extreme Value Approach.
Journal of Banking and Finance, 24:1097-1130, 2000.

A. McNeil
Extreme value theory for risk managers.
Internal Modeling and CAD II, 93-113, 1999.

D. Mercurio, and V. Spokoiny
Estimation of time dependent volatility via local change point analysis
Annals of Statistics, 32:577-602, 2004.

R. Nelsen
An Introduction to Copulas
Springer-Verlag, 1998

R. Schmidt
Tail Dependence for Elliptically contoured Distributions
in P.Cízek, W. Härdle and R.Weron
Statistical Tools for Finance and Insurance
Springer-Verlag, 2005.

V. Spokoiny
Local parametric Methods in nonparametric estimation
Springer-Verlag, 2007.