Copulae in Practice

Ostap Okhrin

Chair of Econometrics and Statistics,
esp. Transportation
Faculty of Transportation
Dresden University of Technology
http://osv.vkw.tu-dresden.de

with contributions by
W. Härdle, Y. Okhrin, A. Ristig, S. Trimborn, P. Song, M. Zhou, S. Zhang,

Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon 02.23.09

In the mid-'80s, *Wall Street* turned to the quants – *brainy financial engineers* – to invent new ways to boost profits.
Their methods for minting money worked brilliantly...
until one of the them devastated the global economy.

Here's what killed your 401(k). *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick and fatally flawed way to assess risk.*

$$\Pr[T_A < 1, T_B < 1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Probability - Specifically, this is a joint default probability–the likelihood that any two members of the pool (A and B) will both default. It's what investors are looking for, and the rest of the formula provides the answer.

Here's what killed your 401(k). *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick and fatally flawed way to assess risk.*

$$\Pr[T_A < 1, T_B < 1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Survival times - The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.

Here's what killed your 401(k). *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick and fatally flawed way to assess risk.*

$$\Pr[T_A < 1, T_B < 1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Distribution functions - The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.

Here's what killed your 401(k). *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick and fatally flawed way to assess risk.*

$$\Pr[T_A < 1, T_B < 1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Copula - This couples (hence the Latine term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.

Here's what killed your 401(k). *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick and fatally flawed way to assess risk.*

$$\Pr[T_A<1, T_B<1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Gamma - The all-powerful correlation parameter, which reduces correlation to a single constant-something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.

Example

- we pay 200 EUR for the chance to win 1000 EUR, if DAX returns decrease by 2%

$$\mathrm{P}_{DAX}(r_{DAX} \leq -0.02) = F_{DAX}(-0.02) = 0.2$$

- we pay 200 EUR for the chance to win 1000 EUR, if DJ returns decrease by 1%

$$\mathrm{P}_{DJ}(r_{DJ} \leq -0.01) = F_{DJ}(-0.01) = 0.2$$

Example

- we get 1000 EUR if DAX and DJ indices decrease simultaneously by 2% and 1% respectively.
  how much are we ready to pay in this case?

$$
\begin{aligned}
&\mathrm{P}\{(r_{DAX} \leq -0.02) \wedge (r_{DJ} \leq -0.01)\} \\
&\quad = F_{DAX,DJ}(-0.02, -0.01) \\
&\quad = C\{F_{DAX}(-0.02), F_{DJ}(-0.01)\} \\
&\quad = C(0.2, 0.2).
\end{aligned}
$$

with C being the copula.

Outline

Univariate Case

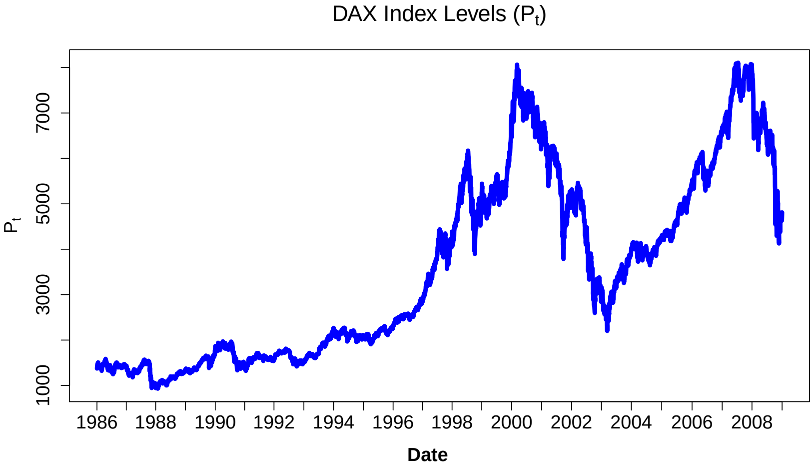
Let $x_1, \dots, x_n$ be realizations of the random variable X
$X \sim F$, where F is unknown

Example

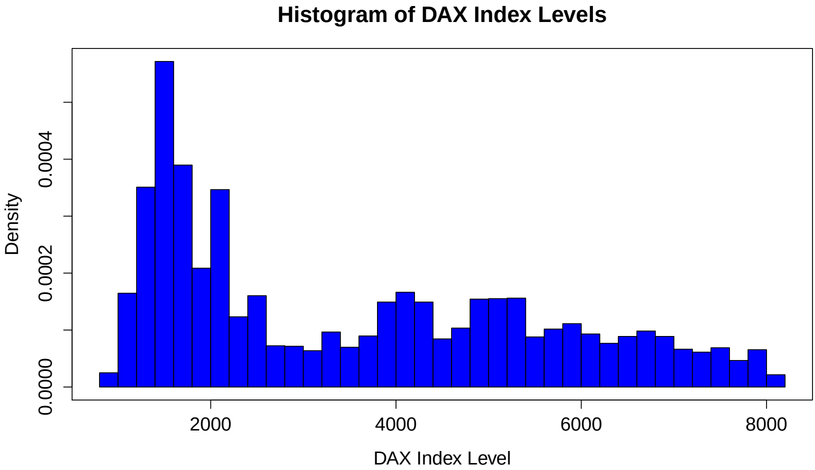
- x_i are returns of the asset for one firm at the day t_i
- x_i are numbers of sold albums *The Man Who Sold the World* by *David Bowie* at day t_i

What is a good approximation of F ?

`traditional` or `modern` approach

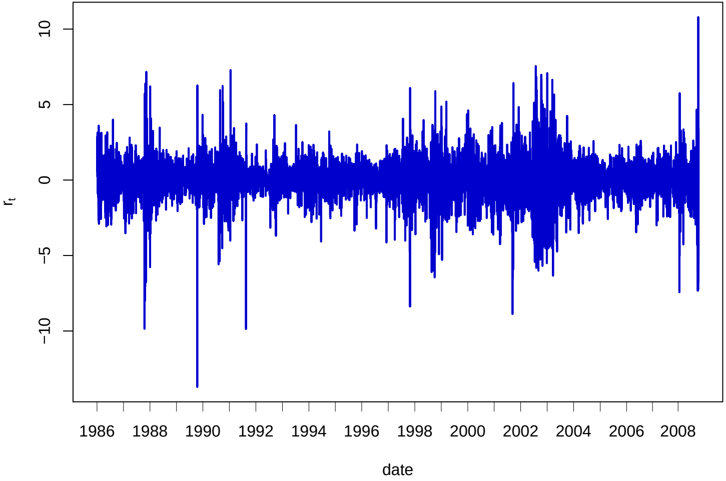


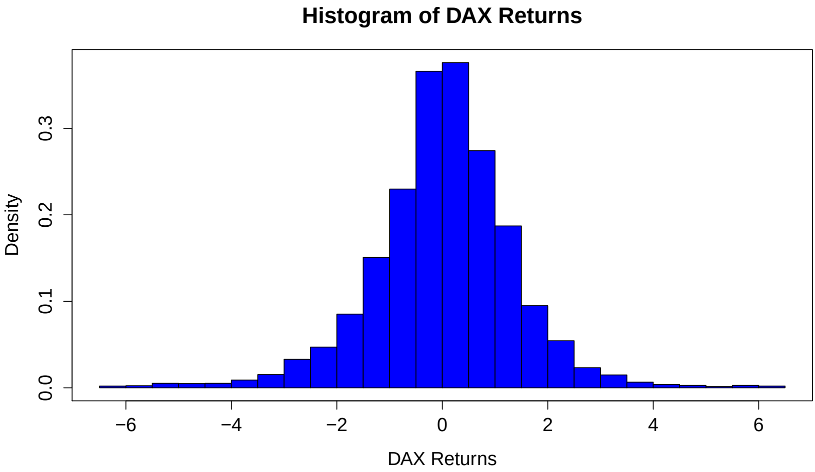
COPdaxtimeseries



COPdaxhistogram

DAX returns ($r_t = \log \frac{P_t}{P_{t-1}}$)





COPdaxreturnhist

Traditional approach:

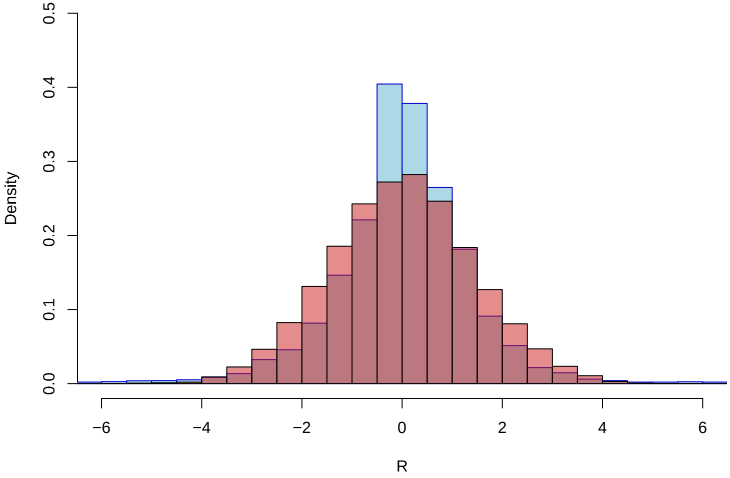
F_0 – known distribution

- parameters of F_0 are estimated from the sample $x_1, \ldots, x_n$
  - $F_0 = \mathrm{N}(\mu, \sigma^2) \;\Rightarrow\; (\mu, \sigma)$, here $\widehat{\mu} = \overline{x}$, $\widehat{\sigma}^2 = \widehat{s}^2$
  - $F_0 = St(\alpha, \beta, \mu, \sigma^2) \;\Rightarrow\; (\alpha, \beta, \mu, \sigma)$ are estimated by Hull Estimator, Tail Exponent Estimation, etc.
- check the appropriateness of F_0 by a test (KS type)

$$H_0 : F = F_0 \quad \text{vs} \quad H_1 : F \neq F_0$$

- if test confirm F_0, use $\widehat{F_0}$

Fit of the Normal distribution to DAX returns ($\widehat{\mu} = 0.0002113130$, $\widehat{\sigma}^2 = 0.0002001865$)



Modern approach: calculate the edf

$$\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{x_i \leq x\},$$

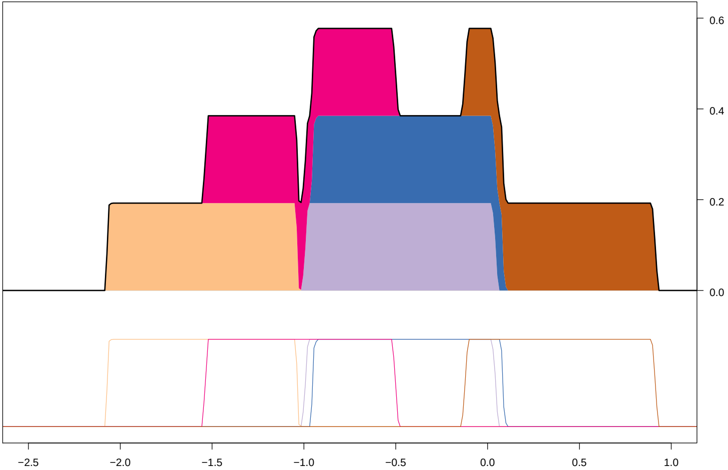
or the nonparametric kernel smoother

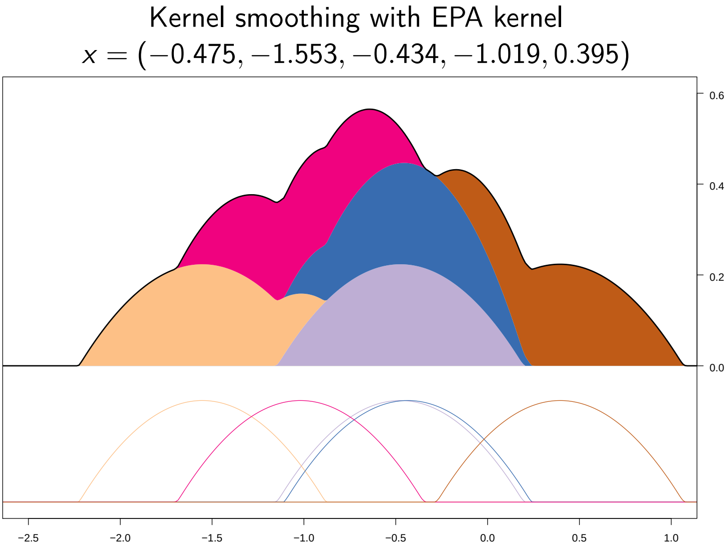
$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

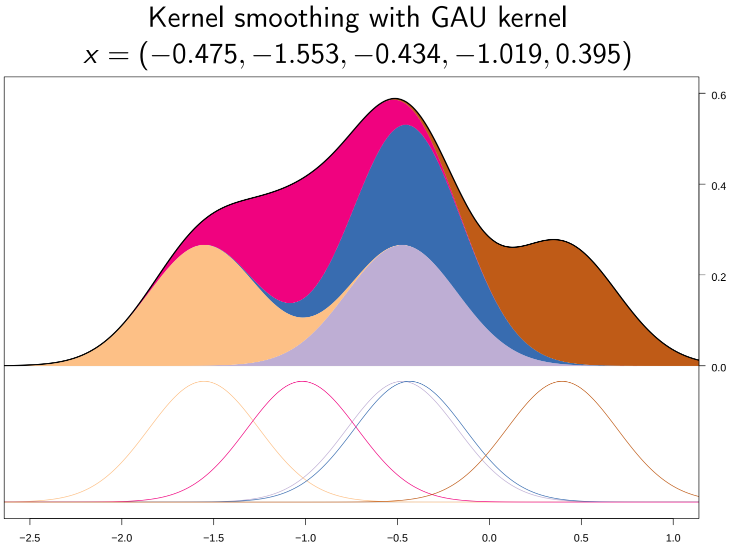
name	$K(u)$
Uniform	$\frac{1}{2}\mathbf{I}\{\lvert u\rvert \leq 1\}$
Epanechnikov	$\frac{3}{4}(1-u^2)\mathbf{I}\{\lvert u\rvert \leq 1\}$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}u^2\right\}$

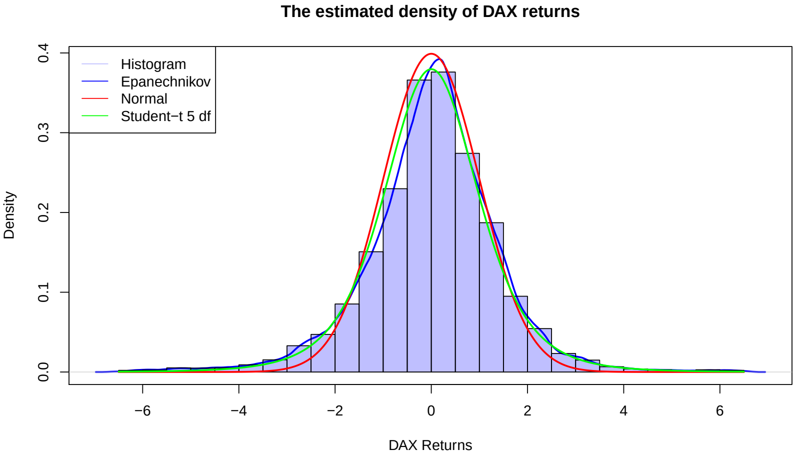
with **I** as indicator function.

Kernel smoothing with UNI kernel
$x = (-0.475, -1.553, -0.434, -1.019, 0.395)$




Kernel smoothing with EPA kernel
$x = (-0.475, -1.553, -0.434, -1.019, 0.395)$
0.6
0.4
0.2
0.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0


Kernel smoothing with GAU kernel
$x = (-0.475, -1.553, -0.434, -1.019, 0.395)$
0.6
0.4
0.2
0.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0



COPdensitydaxreturn

Multivariate Case

$\{x_{1i}, \dots, x_{di}\}_{i=1,\dots,n}$ is the realization of the vector $(X_1, \dots, X_d) \sim F$, where F is unknown.

Example

- $\{x_{1i}, \dots, x_{di}\}_{i=1,\dots,n}$ are returns of the d assets in the portfolio at day t_i
- $(x_{1i}, x_{2i})^\top$ are numbers of sold albums *The Man Who Sold The World* by *David Bowie* and singles *I Saved The World Today* by *Eurythmics* at day t_i

Multivariate Case

What is a good approximation of F ?

`traditional` or `modern` approach

Very flexible approximation to F is challenging in high dimension due to curse of dimensionality.

Traditional approach: Mainly restricted to the class of elliptical distributions: Normal or *t* distributions

$$f_N(x_1,\dots,x_d) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} \exp\left\{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right\}$$

Drawbacks of the elliptical distributions:

1. does not often describe financial data properly
2. huge number of parameters to be estimated

   f.e. for Normal distribution: $\underbrace{\frac{d(d-1)}{2}}_{in\ dependency} + \underbrace{2d}_{in\ margins}$

3. ellipticity

Simulate $X \sim \mathrm{N}(\mu, \Sigma)$ with the sample size $n = 1000$ and estimate the parameters $(\widehat{\mu}, \widehat{\Sigma})$

$$\Sigma = \begin{pmatrix} 1.5 & 0.7 & 0.2 \\ 0.7 & 1.3 & -0.4 \\ 0.2 & -0.4 & 0.3 \end{pmatrix} \Rightarrow \widehat{\Sigma} = \begin{pmatrix} 1.461 & 0.726 & 0.181 \\ 0.726 & 1.335 & -0.408 \\ 0.181 & -0.408 & 0.301 \end{pmatrix}$$

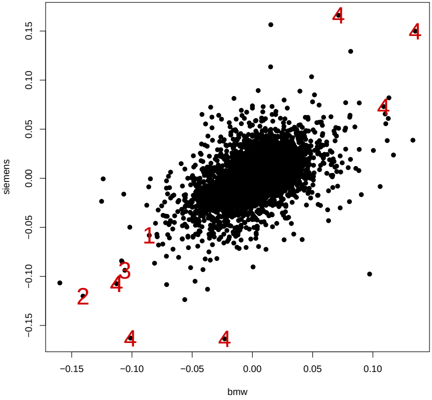
$$\mu = (0,\ 0,\ 0) \Rightarrow \widehat{\mu} = (0.0175,\ -0.0022,\ 0.0055)$$

$\widehat{\Sigma}$ and Σ are not close to each other for only 3 dimensions and quiet big sample

"Extreme, **synchronized rises and falls** in financial markets occur infrequently but they do occur. The problem with the models is that they did not assign a high enough chance of occurrence to the scenario in which **many things go wrong at the same time**
- the "perfect storm" scenario"

(Business Week, September 1998)

Correlation



1. 19.10.1987
   Black Monday
2. 09.11.1989
   Berlin Wall
3. 19.08.1991
   Kremlin
4. 17.03.2008, 19.09.2008, 10.10.2008, 13.10.2008, 15.10.2008, 29.10.2008
   Crisis

Correlation

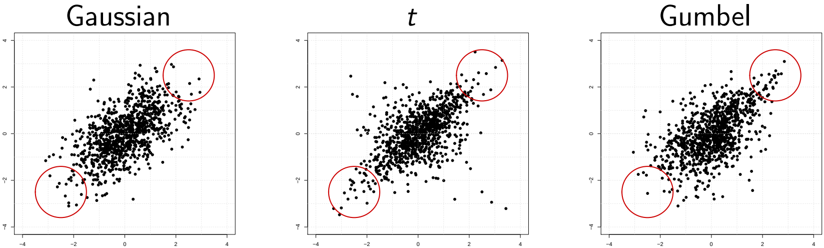


Figure 1: Scatterplots for two distributions with $\rho = 0.4$

- same linear correlation coefficient ($\rho = 0.4$)
- same marginal distributions
- rather big difference

Copula

Books:

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

Nelsen, R. B. (2006). An Introduction to Copulas, Springer Verlag, New York.

Copula

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

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

for a while $d = 2$

A little bit of history

- 1940s: *Wassilij Hoeffding* studies properties of multivariate distributions

Wassilij Hoeffding

1914–91, b. Mustamäki, Finland; d. Chapel Hill, NC
gained his PhD from U Berlin in 1940
1924–45 work in U Berlin

A little bit of history

- 1940s: *Wassilij Hoeffding* studies properties of multivariate distributions
- 1959: The word **copula** appears for the first time (*Abe Sklar*)
- 1999: Introduced to financial applications (*Paul Embrechts*, *Alexander McNeil*, *Daniel Straumann* in RISK Magazine)
- 2000: Paper by *David Li* in *Journal of Derivatives* on application of copulae to CDO
- 2006: Several insurance companies, banks and other financial institutions apply copulae as a risk management tool

Applications

Practical Use:

1. medicine (Vandenhende (2003), ...)
2. hydrology (Genest and Favre (2006), Durante and O2 (2015), ...)
3. biometrics (Wang and Wells (2000, JASA), Chen and Fan (2006, CanJoS), ...)
4. economics
    - portfolio selection (Patton (2004, JoFE), Hennessy and Lapan (2002, MathFin), ...)
    - time series (Chen and Fan (2006a, 2006b, JoE), Fermanian and Scaillet (2003, JoR), Lee and Long (2005, JoE), ...)
    - risk management (Junker and May (2002, EJ), Breyman et. al. (2003, QF), ...)

Special Copulae

Theorem

Let C be a copula. Then for every $(u_1, u_2) \in [0,1]^2$

$$\max(u_1 + u_2 - 1, 0) \le C(u_1, u_2) \le \min(u_1, u_2),$$

*where bounds are called **lower and upper Fréchet-Hoeffdings bounds**. When they are copulae they represent perfect negative and positive dependence respectively.*

The simplest copula is the **product copula**

$$\Pi(u_1, u_2) = u_1 u_2$$

characterize the case of independence.

Copula Classes

1. elliptical
   - implied by well-known multivariate df's (Normal, t), derived through Sklar's theorem
   - do not have closed form expressions and are restricted to have radial symmetry
2. Archimedean

$$C(u_1, u_2) = \phi^{-1}\{\phi(u_1) + \phi(u_2)\}$$

   - allow for a great variety of dependence structures
   - closed form expressions
   - several useful methods for multivariate extension
   - not derived from mv df's using Sklar's theorem

Copula Examples 1

Gaussian copula

$$\begin{aligned} C_\delta^G(u_1, u_2) &= \Phi_\delta\{\Phi^{-1}(u_1), \Phi^{-1}(u_2)\} \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\delta^2}} \exp\left\{\frac{-(s^2 - 2\delta st + t^2)}{2(1-\delta^2)}\right\} ds\, dt, \end{aligned}$$

- Gaussian copula contains the dependence structure
- *normal* marginal distribution + Gaussian copula = multivariate normal distributions
- *non-normal* marginal distribution + Gaussian copula = meta-Gaussian distributions
- allows to generate joint symmetric dependence, but no tail dependence

Copula Examples 2

Gumbel (1960) copula

$$C_\theta^{Gu}(u_1, u_2) = \exp\left[-\left\{(-\log u_1)^{1/\theta} + (-\log u_2)^{1/\theta}\right\}^\theta\right], \ 1 \leq \theta < \infty.$$

- $\phi(x, \theta) = \exp\{-x^{1/\theta}\}, \ x \in [0, \infty)$
- independence for $\theta = 1$
- upper Frèchet-Hoeffding for $\theta \to \infty$
- asymmetric dependence and upper tail dependence, but no lower tail dependence
- the only extreme value Archimedean copula

Copula Examples 3

Clayton (1978) copula

$$C_\theta^{Cl}(u_1, u_2) = \left\{\max\left(u_1^{-\theta} + u_2^{-\theta} - 1, 0\right)\right\}^{-\frac{1}{\theta}}, \ -1 \leq \theta < \infty, \theta \neq 0$$

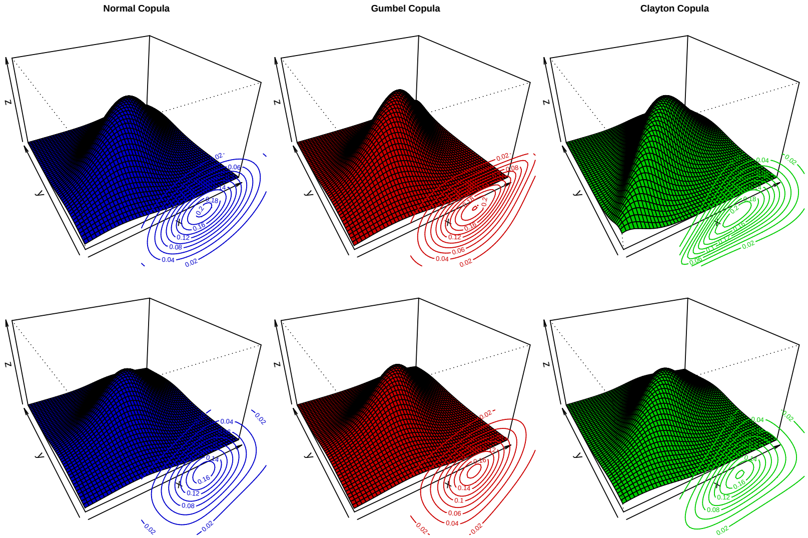
- $\phi(x, \theta) = (\theta x + 1)^{-\frac{1}{\theta}}, \ x \in [0, \infty)$
- lower Fréchet-Hoeffding bound for $\theta \to -1$
- independence for $\theta = 0$
- upper Fréchet-Hoeffding bound for $\theta \to \infty$
- asymmetric dependence and lower tail dependence, but no upper tail dependence
- the only Archimedean copula with truncated property

Copula Examples 4

Frank (1979) copula

$$C_\theta^{Fr}(u_1, u_2) = -\frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right\}$$
$$-\infty < \theta < \infty,\ \theta \neq 0.$$

- $\phi(x, \theta) = \theta^{-1}\log\left\{1 - (1 - e^{-\theta})e^{-x}\right\},\ x \in [0, \infty)$
- lower Fréchet-Hoeffding bound for $\theta \to -\infty$
- independence for $\theta = 0$
- upper Fréchet-Hoeffding bound for $\theta \to \infty$
- the only elliptically contoured Archimedean copula


Normal Copula
Gumbel Copula
Clayton Copula

Dependencies, Pearson's rho

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

- Sensitive to outliers
- Measures the 'average dependence' between X_1 and X_2
- Invariant under strictly increasing linear transformations
- May be misleading in situations where multivariate df is not elliptical
- `cor(x, y, method = 'pearson')` or `cor(x, y)`

Dependencies, Kendall's tau

Definition

If F is continuous bivariate cdf and let (X_1, X_2), (X_1', X_2') be independent random pairs with distribution F. Then **Kendall's tau** is

$$\tau = \boldsymbol{P}\{(X_1 - X_1')(X_2 - X_2') > 0\} - \boldsymbol{P}\{(X_1 - X_1')(X_2 - X_2') < 0\}$$

- Less sensitive to outliers
- Measures the 'average dependence' between X and Y
- Invariant under strictly increasing transformations
- Depends only on the copula of (X_1, X_2)
- For elliptical copulae: $\delta(X_1, X_2) = \sin\left(\frac{\pi}{2}\tau\right)$
- `cor(x, y, method = 'kendall')`

Dependencies, Spearmans's rho

Definition

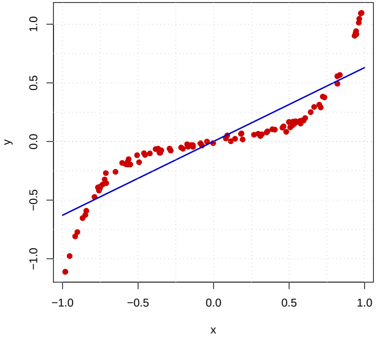
If F is a continuous bivariate cumulative distribution function with marginal F_1 and F_2 and let $(X_1, X_2) \sim F$. Then **Spearmans's rho** is a correlation between $F_1(X_1)$ and $F_2(X_2)$

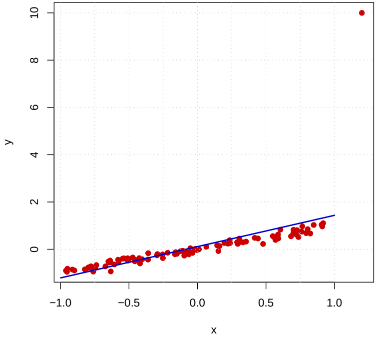
$$\rho = \frac{Cov\{F_1(X_1), F_2(X_2)\}}{\sqrt{Var\{F_1(X_1)\}Var\{F_2(X_2)\}}}.$$

- Less sensitive to outliers
- Measures the 'average dependence' between X_1 and X_2
- Invariant under strictly increasing transformations
- Depends only on the copula of (X_1, X_2)
- For elliptical copulae: $\delta(X_1, X_2) = 2\sin\left(\frac{\pi}{6}\rho\right)$
- `cor(x, y, method = 'spearman')`

$$\begin{aligned} \delta &= 0.892, \\ \tau &= 0.956, \\ \rho &= 0.996 \end{aligned}$$

$$\begin{aligned} \delta &= 0.659, \\ \tau &= 0.888, \\ \rho &= 0.982 \end{aligned}$$





Method of Moments Estimation

Gaussian copula

$$
\begin{aligned}
\rho &= \frac{6}{\pi}\arcsin\frac{\delta}{2}, \\
\tau &= \frac{2}{\pi}\arcsin\delta,
\end{aligned}
$$

where δ is the Pearson linear correlation coefficient.
Gumbel copula with parameter θ.

$$
\begin{aligned}
\rho &- \text{no closed form}, \\
\tau &= 1 - \frac{1}{\theta}.
\end{aligned}
$$

Later `irho` or `itau`

Multivariate Copula Definition

Definition

The **copula** is a multivariate distribution with all univariate margins being $U(0,1)$.

Theorem (Sklar, 1959)

Let $X_1, \dots, X_d$ be random variables with marginal distribution functions $F_1, \dots, F_d$ and joint distribution function F. Then there exists a d-dimensional copula $C : [0,1]^d \to [0,1]$ such that $\forall x_1, \dots, x_d \in \overline{\mathbb{R}} = [-\infty, \infty]$

$$F(x_1, \dots, x_d) = C\{F_1(x_1), \dots, F_d(x_d)\} \tag{1}$$

If the margins $F_1, \dots, F_d$ are continuous, then C is unique. Otherwise C is uniquely determined on $F_1(\overline{\mathbb{R}}) \times \dots \times F_d(\overline{\mathbb{R}})$ Conversely, if C is a copula and $F_1, \dots, F_d$ are distribution functions, then the function F defined in (1) is a joint distribution function with margins $F_1, \dots, F_d$.

Copula Density

The copula density:

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

Joint density function based on copula

$$_c f(x_1, \ldots, x_d) = c\{F_1(x_1), \ldots, F_d(x_d)\} \cdot f_1(x_1) \cdot \ldots \cdot f_d(x_d),$$

where $f_1(\cdot), \ldots, f_d(\cdot)$ are marginal density functions.

Special Copulae

Theorem

Let C be a copula. Then for every $(u_1, \dots, u_d) \in [0,1]^d$

$$\max\left(\sum_{i=1}^{d} u_i + 1 - d, 0\right) \leq C(u_1, \dots, u_d) \leq \min(u_1, \dots, u_d),$$

where bounds are called ***lower and upper Fréchet-Hoeffdings bounds****. When they are copulae they represent perfect negative and positive dependence respectively.*

The simplest copula is the **product copula**

$$\Pi(u_1, \dots, u_d) = \prod_{i=1}^{d} u_i$$

characterize the case of independence.

Archimedean Copula

Multivariate Archimedean copula $C : [0,1]^d \to [0,1]$ defined as

$$C(u_1, \dots, u_d) = \phi\{\phi^{-1}(u_1) + \cdots + \phi^{-1}(u_d)\}, \qquad (2)$$

where $\phi : [0, \infty) \to [0,1]$ is continuous and strictly decreasing with $\phi(0) = 1$, $\phi(\infty) = 0$ and ϕ^{-1} its pseudo-inverse.

Advantages: single parameter, closed form
Disadvantages: too restrictive: single parameter, exchangeable

d-dimensional Gumbel copula

The generator function and the copula function are given as follows:

$$\phi(x,\theta) = \exp\left\{-x^{\frac{1}{\theta}}\right\}, \quad 1 \le \theta < \infty, x \in [0,\infty),$$

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

d-dimensional Clayton copula

The generator function and the copula function are given as follows:

$$\phi(x,\theta) = (\theta x + 1)^{-\frac{1}{\theta}},$$
$$-1/(d-1) \leq \theta < \infty, \theta \neq 0, x \in [0,\infty),$$
$$C_\theta^{Cl}(u_1,\dots,u_d) = \left\{ \left(\sum_{j=1}^{d} u_j^{-\theta} \right) - d + 1 \right\}^{-\theta^{-1}}.$$

d-dimensional Frank copula

The generator function and the copula function are given as follows,

$$\phi(x,\theta) = \theta^{-1} \log\left\{1-(1-e^{-\theta})e^{-x}\right\}, \\ -\infty < \theta < \infty,\ \theta \neq 0,\ x \in [0,\infty),$$

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

R packages for copula

- `copula` - most powerful copula package (!), Yan (2007), Hofert and Maechler (2011), Kojadinovic and Yan (2010)
- `fCopulae` - learning purposes, bivariate, Wuertz et al. (2009a)
- `fgac` - generalized Archimedean copulas, Gonzalez-Lopez (2009)
- `gumbel` - functions for Gumbel copulas, Caillat et al. (2008)
- `HAC` - inference for hierarchical Archimedean copulae, Okhrin and Ristig (2012)
- `VineCopula` - inference for vine copulae, Czado et al. (2015)
- `gofCopula` - goodness-of-fit tests for copulae, Trimborn, Okhrin, Zhang, Zhou (2015)
- `sbgcop` - Gaussian copula with margins being nuisance parameters, Hoff (2010)
- ...

R packages for copula

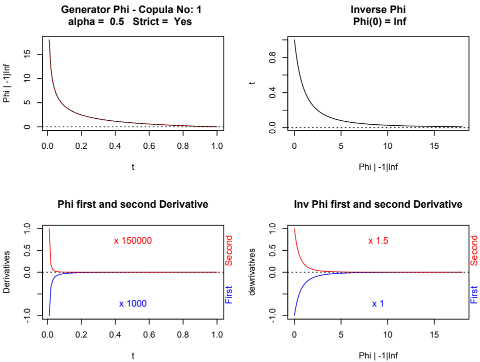
- copula - most powerful copula package (!), Yan (2007), Hofert and Maechler (2011), Kojadinovic and Yan (2010)
- fCopulae - learning purposes, bivariate, Wuertz et al. (2009a)
- fgac - generalized Archimedean copulas, Gonzalez-Lopez (2009)
- gumbel - functions for Gumbel copulas, Caillat et al. (2008)
- HAC - inference for hierarchical Archimedean copulae, Okhrin and Ristig (2012)
- VineCopula - inference for vine copulae, Czado et al. (2015)
- gofCopula - goodness-of-fit tests for copulae, Trimborn, Okhrin, Zhang, Zhou (2015)
- sbgcop - Gaussian copula with margins being nuisance parameters, Hoff (2010)
- ...

fCopulae get a feeling using sliders

copula method: 1 - Clayton, 4 - Gumbel, 5 - Frank

- Generator functions

```
PhiSlider()
```

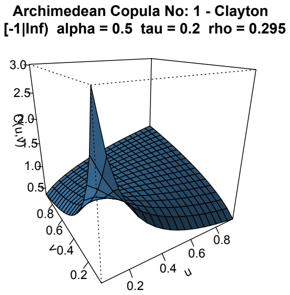


fCopulae get a feeling using sliders

copula method: 1 - Clayton, 4 - Gumbel, 5 - Frank

- Copula density

```
darchmSlider()
```



copula **objects**

- Gaussian copula with $\rho = 0.75$ and $F_1 = N(0, 2)$ and $F_2 = Exp(2)$

```
> ga.c          = normalCopula(0.75, dim = 2)
> ga.c
Normal copula family
Dimension:  2
Parameters:
   rho.1  =  0.75
> mvdc.ga.c = mvdc(ga.c, c('norm','exp'),
     paramMargins = list(list(mean=0, sd=2),
                         list(rate=2)))
> mvdc.gauss.n.e
Multivariate Distribution Copula based ("mvdc")
 @ copula:
.........
```

Simulation

Frees and Valdez, (1998, NAAJ), Whelan, (2004, QF), Marshal and Olkin, (1988, JASA), Hofert (2008, CSDA)

Conditional inversion method:

Let $C = C(u_1, \dots, u_d)$, $C_i = C(u_1, \dots, u_i, 1, \dots, 1)$ and $C_d = C(u_1, \dots, u_d)$. Conditional distribution of U_i is given by

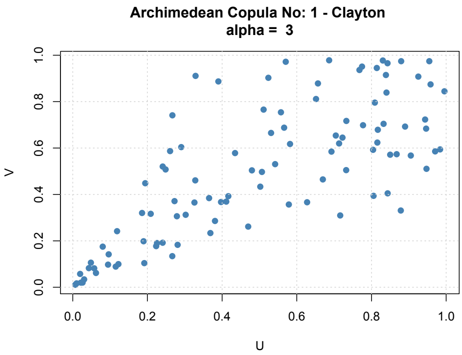
$$\begin{aligned} C_i(u_i|u_1, \dots, u_{i-1}) &= \boldsymbol{P}\{U_i \leq u_i | U_1 = u_1 \dots U_{i-1} = u_{i-1}\} \\ &= \frac{\partial^{i-1} C_i(u_1, \dots, u_i)}{\partial u_1 \dots \partial u_{i-1}} \Big/ \frac{\partial^{i-1} C_{i-1}(u_1, \dots, u_{i-1})}{\partial u_1 \dots \partial u_{i-1}} \end{aligned}$$

- Generate i.r.v. $v_1, \dots, v_d \sim U(0,1)$
- Set $u_1 = v_1$
- $u_i = C_d^{-1}(v_i | u_1, \dots, u_{i-1}), \ \forall i = 2, \dots, d$

fCopulae get a feeling using sliders

copula method: 1 - Clayton, 4 - Gumbel, 5 - Frank

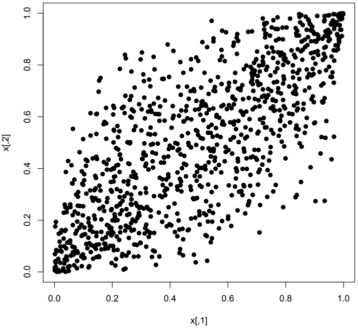
```
rarchmCopula(1000, alpha = 0.5, type = "4")
rarchmSlider()
```



copula simulation

- Simulation from a copula

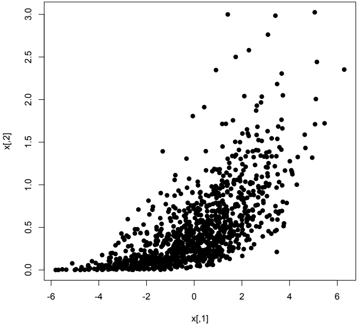
```
x = rCopula(1000, gauss.c)
plot(x, pch = 19)
```



copula simulation

- Simulation from a copula-based distribution

```
x = rMvdc(1000, mvdc.ga.c)
plot(x, pch = 19)
```



Estimation Issues - Margins

$$\widehat{F}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{I}(x_{ji} \leq x),$$

```r
apply(x, 2, FUN = rank) / (nrow(x) + 1)
```

$$\widetilde{F}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} K\left(\frac{x - x_{ji}}{h}\right)$$

for $j = 1, \dots, k$, where $\varkappa : \mathbb{R} \to \mathbb{R}$, $\int \varkappa = 1$, $K(x) = \int_{-\infty}^{x} \varkappa(t)dt$ and $h > 0$ is the bandwidth.

Estimation Issues - Margins

$$F_j(x;\widehat{\alpha}_j) = F_j\left\{x; arg\max_\alpha \sum_{i=1}^{n} \log\ f_j(x_{ji},\alpha)\right\},$$

```
optimise(
  f = function(a){
    sum(log(dexp(x[, 2], rate = a)))
  },
  interval = c(0, 10),
  maximum  = TRUE)
```

$$\check{F}_j(x) \in \{\widehat{F}_j(x), \widetilde{F}_j(x), F_j(x;\widehat{\alpha}_j)\}$$

Full maximum likelihood estimation

- The log-likelihood function:

$$\ell(\alpha; x_1, \dots, x_n) = \sum_{i=1}^{n} \log c\{F_1(x_{1i}; \alpha_1), \dots, F_d(x_{di}; \alpha_d); \theta\} + \sum_{i=1}^{n} \sum_{j=1}^{d} \log f_j(x_{ji}; \alpha_j).$$

- The efficient and asymptotically normal estimator:

$$\widehat{\alpha}_{FML} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_d, \widehat{\theta})^\top = \arg\min_{\alpha} \ell(\alpha).$$

IFM (inference for margins) method

Steps:

1. Estimate the parameter α_j from the margins
2. Estimate the dependence parameter θ

- Maximize the pseudo log-likelihood function over θ to get the dependence parameter estimate $\widehat{\theta}$,

$$\ell(\theta, \widehat{\alpha}_1, \dots, \widehat{\alpha}_d) = \sum_{i=1}^{n} \log \; c\{F_1(x_{1i}; \widehat{\alpha}_1), \dots, F_d(x_{di}; \widehat{\alpha}_d); \theta\}.$$

CML (canonical maximum likelihood) method

- Normalize the empirical cdf not by n but by $n+1$

$$\widehat{F}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} I(x_{ji} \leq x) \text{ or } \widetilde{F}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} K\left(\frac{x - x_{ji}}{h}\right)$$

- The copula parameter estimator $\widehat{\theta}_{CML}$ is given by:

$$\begin{aligned}
\widehat{\theta}_{CML} &= \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} \log\ c\{\widehat{F}_1(x_{1i}), \dots, \widehat{F}_d(x_{di}); \theta\}, \\
&\text{or} \\
\widehat{\theta}_{CML} &= \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} \log\ c\{\widetilde{F}_1(x_{1i}), \dots, \widetilde{F}_d(x_{di}); \theta\}.
\end{aligned}$$

Empirical Copula

$$C_n(u_1,\dots,u_d) = \frac{1}{n}\sum_{i=1}^{n} I\{\check{F}_1(x_{i1}) \le u_1,\dots,\check{F}_d(x_{id}) \le u_d\}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} I\{\check{F}_j(x_{ij}) \le u_j\}$$

```
emp.copula = function(u, data){
    n = dim(data)[1]
    d = dim(u)[2]
    embed = function(u){
        comp = function(X){(X <= u)}
        Comp = function(X){apply(X, 1, comp)}
        cumsum(apply(t(Comp(data)), 1, prod)) / n
    }
    apply(u, 1, embed)[n,]
}
```

Estimation in R

- copula package

```
u = apply(x, 2, FUN = rank) / (nrow(x) + 1)
fitCopula(ga.c, u, "mpl")
fitCopula(ga.c, u, "itau")
fitCopula(ga.c, u, "irho")
```

- fCopulae package

```
ellipticalCopulaFit(u[,1], u[,2], type = "norm")
```

- copula package (fits whole distribution)

```
fitMvdc(x, mvdc.ga.c, start = c(0.5, 3, 3, 0.5),
  hideWarnings = FALSE)
```

Goodness-of-Fit Tests

Papers:

Zhang, S., Okhrin, O., Zhou, Q., and Song, P., Goodness-of-fit Test For Specification of Semiparametric Copula Dependence Models, forthcoming in *Journal of Econometrics*

Trimborn, S., Zhang, S., Okhrin, O., and Zhou, Q., `gofCopula` package for `R`

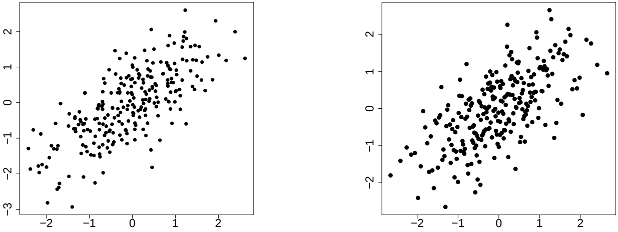
Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199-213.

Applications

1. medicine (Vandenhende (2003), ...)
2. hydrology (Genest and Favre (2006), ...)
3. biometrics (Wang and Wells (2000, JASA), Chen and Fan (2006, CanJoS), ...)
4. economics
   - portfolio selection (Patton (2004, JoFE), Hennessy and Lapan (2002, MathFin), ...)
   - time series (Chen and Fan (2006a, 2006b, JoE), Fermanian and Scaillet (2003, JoR), Lee and Long (2005, JoE), ...)
   - risk management (Junker and May (2002, EJ), Breyman et. al. (2003, QF), ...)
5. ...

How to be sure, that one uses a proper copula?

Different tests ⇒ Different outcomes



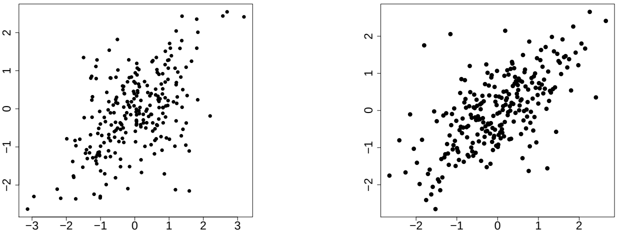
(a) Gaussian copula (b) year 2004

Figure 2: Sample from Gauss copula with $\mathrm{N}(0,1)$ margins, $\theta = 0.71$, $N = 250$ and residuals transformed to standard normal for Citygroup/BoA for 2004.

Visually - Gaussian copula

Test 1: Gumbel, Test 2: Gauss, Test 3: Gauss

Different tests $\Rightarrow$ Different outcomes



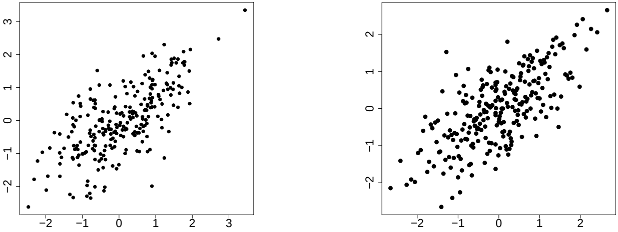
(a) t-copula (b) year 2006

Figure 3: Sample from t-copula with $\mathrm{N}(0,1)$ margins, $\theta = 0.6$, $N = 250$ and residuals transformed to standard normal for Citygroup/BoA for 2006.

Visually - t-copula

Test 1: t-copula, Test 2: Gauss, Test 3: t-copula

Different tests ⇒ Different outcomes



(a) Gumbel copula (b) year 2009

Figure 4: Sample from Gumbel copula with N(0, 1) margins, $\theta = 2$, $N = 250$ and residuals transformed to standard normal for Citygroup/BoA for 2009.

Visually - Gumbel copula

Test 1: Gumbel, Test 2: Gumbel, Test 3: Gauss

Goodness-of-Fit Tests

$$\mathcal{H}_0 : C_0 \in \mathcal{C} \quad vs. \quad \mathcal{H}_1 : C_0 \notin \mathcal{C}$$

where $\mathcal{C} = \{C(\cdot;\theta) : \theta \in \Theta\}$.

- $X_1 = (x_{11}, \dots, x_{d1})^\top, \dots, X_n = (x_{1n}, \dots, x_{dn})^\top$ random sample of size n drawn from multivariate distribution $H(x) = H(x_1, x_2, \dots, x_d)$
- Continuous marginal cdf $F(x) = \{F_1(x_1), \dots, F_d(x_d)\}$

$$H(x_1, x_2, \dots, x_d) = C_0\{F(x)\} = C_0\{F_1(x_1), \dots, F_d(x_d)\}.$$

PIOS test, I

Define $\ell\{\widehat{F}(X_i);\theta\} = \log c\{\widehat{F}_1(x_{1i}),\dots,\widehat{F}_d(x_{di});\theta\}$ and $\widehat{\theta}$ be the two-step pseudo maximum likelihood method (PMLE) of θ given by

$$\widehat{\theta} = \operatorname*{argmax}_{\theta\in\Theta} \sum_{i=1}^{n} \ell\{\widehat{F}(X_i);\theta\}.$$

Compute delete-one-block PLMEs $\widehat{\theta}_{-b}$, $1 \leq b \leq B$:

$$\widehat{\theta}_{-b} = \operatorname*{argmax}_{\theta\in\Theta} \sum_{b'\neq b}^{B} \sum_{i=1}^{m} \ell\{\widehat{F}(X_i^{b'});\theta\}, \quad b = 1,\dots,B,$$

PIOS test, `gofPIOSTn`

Comparing "in-sample" and "out-of-sample" pseudo-likelihoods with the following test statistic:

$$T_n(m) = \sum_{b=1}^{B} \sum_{i=1}^{m} \left[\ell\{\widetilde{F}(X_i^b); \widehat{\theta}\} - \ell\{\widetilde{F}(X_i^b); \widehat{\theta}_{-b}\} \right].$$

Challenge: needed $[\frac{n}{m}]$ *dependence parameters*
Solution: test statistic which is asymptotically equivalent.

PIOS test, III

- Under suitable regularity conditions and under assumption, that $\exists\ \theta^* \in \Theta$ with $\widehat{\theta} \xrightarrow{p} \theta^*$ for $n \to \infty$:

$$T_n(m) \xrightarrow{p} \mathrm{tr}\{S(\theta^*)^{-1} V(\theta^*)\}$$

with

$$S(\theta) = -\,\mathsf{E}_0 \left[\frac{\partial^2}{\partial\theta\partial\theta^\top} \ell\{F(X_1);\theta\} \right],$$
$$V(\theta) = \mathsf{E}_0 \left[\frac{\partial}{\partial\theta} \ell\{F(X_1);\theta\} \frac{\partial}{\partial\theta} \ell^\top\{F(X_1);\theta\} \right].$$

PIOS test, `gofPIOSRn`

- Under a correct model specification, it holds: $V(\theta^*) = S(\theta^*)$.
- Then is $\operatorname{tr}\{S(\theta^*)^{-1}V(\theta^*)\} = p$.
- Asymptotic test statistic:

$$R_n = \operatorname{tr}\left\{\widehat{S}(\widehat{\theta})^{-1}\widehat{V}(\widehat{\theta})\right\}$$

where $\widehat{S}(\widehat{\theta})$ and $\widehat{V}(\widehat{\theta})$ are the empirical counterparts to $S(\theta)$ and $V(\theta)$.

Similar to `gofWhite` where one tests if $V(\theta^*) - S(\theta^*) = 0$, see White (1982)

Law of Large Numbers

Theorem

Under assumptions A1 and A2 hold

$$R_n \xrightarrow{p} tr\left\{S(\theta^*)^{-1}V(\theta^*)\right\}, \quad as \quad n \to \infty,$$

where θ^* is the limiting value of PMLE $\widehat{\theta}$.

▸ Assumptions

Central Limit Theorem

Theorem

- *Under the null hypothesis, if A2 and B1 - B3 hold, then*

$$\sqrt{n}\,(R_n - p) \xrightarrow{d} N(0, \sigma_R^2), \quad as\ n \to \infty,$$

  where σ_R^2 is the asymptotic variance.
- *Under assumptions A2, B1 - B3 and C1,*

$$R_n - T_n(m) = o_p(n^{-1/2}).$$

▸ Assumption $A2$ ▸ Assumptions $B1 - B3$ ▸ Assumption $C1$

Local Power, I

- Asymptotic power of R_n against a local alternative in the Pitman sense for a constant $\delta > 0$:

$$H_{1,n} : P_n^{C_1,\delta}(x) = C_0\{F(x);\theta_0\} + \frac{\delta}{\sqrt{n}}\left[C_1\{F(x)\} - C_0\{F(x);\theta_0\}\right]$$

- Assume $C_1\{F(x)\} \geq C_0\{F(x);\theta_0\}$ for all $x \in \mathbb{R}^d$
  - Ensures that $P_n^{C_1,\delta}(x)$ is a copula for $0 < \delta \leq n^{1/2}$ and the departure from the null $C_0\{F(x);\theta_0\}$ increases as δ increases.

Local Power, II

Theorem

Suppose D1 holds in addition to the assumptions A2 and B1 - B3. Then under $H_{1,n}$

$$\sqrt{n}\,(R_n - p) \xrightarrow{\mathcal{L}} \mathrm{N}\{\delta m(c_0, c_1), \sigma_R^2\}$$

where

$$m(c_0, c_1) = \mathrm{E}_{c_0}\left[W(X_t) g\left\{F(X_t); \theta_0\right\}\right],$$

and $\mathrm{E}_{c_0}(\cdot)$ denotes the expectation under the null distribution c_0 or P_0, and $W(\cdot)$ as a weighting function. That is, $m(c_0, c_1)$ is a weighted expectation of $g\left\{F(X_t); \theta_0\right\}$ under P_0.

Assumption $A2$ Assumptions $B1 - B3$ Assumption $D1$

Local Power, III

- Implication: as long as $m(c_0, c_1) \neq 0$
  - R_n will yield power locally
  - The asymptotic local power increases to 1 as δ increases to infinity
  - → R_n is a consistent test
  - T_n has the same asymptotic local power function as R_n
  - → T_n is also a consistent test

Local power, Simulation Study I

- Asymptotic power of R_n under alternatives in the Pitman sense
- Two settings: Clayton copula under H_0, and Gaussian copula under H_0
- $n = 500,\ N = 1000$
- Margins $F(\cdot)$ uniform on $[0, 1]$
- $(\tau_1, \tau_2) = (0.4, 0.8)$
- $\delta \in [0.0; 0.5]$

Results

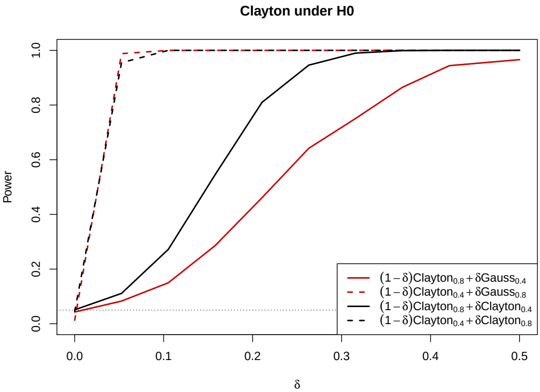


Figure 5: Local Power curves for the R_n test with Clayton copula being under H_0 and four different cases of true mixture copulas.

PIOS for the time series models, I

- Semi-Parametric Copula based Multivariate DYnamic model (SCOMDY), Chen and Fan (2006), for time series data

$$Y_t = \mu_t(\eta_1^0) + \Sigma_t^{1/2}(\eta^0)\epsilon_t,$$

- $Y_t = (Y_{t1}, \dots, Y_{td})^\top$
- $\mu_t(\eta_1^0) = \left\{\mu_{t1}(\eta_1^0), \dots, \mu_{td}(\eta_1^0)\right\}^\top = \mathsf{E}\left(Y_t | \mathcal{F}_{t-1}\right)$
- $\mathcal{F}_t$ is sigma-field generated by $(Y_{t-1}, Y_{t-2}, \dots; Z_t, Z_{t-1}, \dots)$, and Z_t is a vector of predetermined or exogenous variables.
- $\Sigma_t(\eta^0) = diag\left\{\Sigma_{t1}(\eta^0), \dots, \Sigma_{td}(\eta^0)\right\}$, where $\Sigma_{tj}(\eta^0) = \mathsf{E}\left[\left\{Y_{tj} - \mu_{tj}(\eta_1^0)\right\}^2 | \mathcal{F}_{t-1}\right]$, $j = 1, \dots, d$,
- $\epsilon_t = (\epsilon_{t1}, \dots, \epsilon_{td})^\top$, $t = 1, \dots, n$ with $\epsilon_t \overset{iid}{\sim} \mathcal{L}(0, 1)$

PIOS for the time series models, II

- Special cases of SCOMDY:
  - VAR
  - Multivariate ARMA
  - Multivariate GARCH
  - ...
- Estimation:
  - Performed with three-stage procedure
- Resulting residuals are used to construct PIOS test to test the specification of a parametric copula.

Estimation, I

1. Univariate quasi ML with $\epsilon \sim \mathrm{N}(0,1)$ to estimate $\eta = \left(\eta_1^\top, \eta_2^\top\right)^\top$:

$$\widehat{\eta}_1 = \arg\min_{\eta_1 \in \Psi_1} \left[\frac{1}{n} \sum_{t=1}^{n} \left\{ Y_t - \mu_t(\eta_1) \right\}^\top \left\{ Y_t - \mu_t(\eta_1) \right\} \right]$$

and

$$\widehat{\eta}_2 = \arg\min_{\eta_2 \in \Psi_2} \left(\frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{d} \left[\Sigma_{tj}^{-1}(\widehat{\eta}_1, \eta_2) \left\{ Y_t - \mu_t(\widehat{\eta}_1) \right\}^2 + \log \Sigma_{tj}(\widehat{\eta}_1, \eta_2) \right] \right)$$

Estimation, II

2. Estimate marginal distribution $F_j(\cdot)$ of $\widetilde{\epsilon}_{tj}$

$$\widetilde{\epsilon}_{tj} = \Sigma_{tj}^{-1/2}(\widehat{\eta})\left\{y_{tj} - \mu_{tj}(\widehat{\eta}_1)\right\}, \quad j = 1, \dots, d; \quad t = 1, \dots, n$$

by

$$\check{F}_j(x) = \frac{1}{n+1}\sum_{t=1}^{n} \mathbf{I}\left\{\widetilde{\epsilon}_{tj} \le x\right\}, \; x \in \mathbb{R}, \; j = 1, \dots, d.$$

Estimation, III

3. Estimate θ by

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} \ell\{\check{F}(\widetilde{\epsilon}_t);\theta\},$$

   where $\ell(\cdot;\cdot) = \log c(\cdot;\cdot)$.

- use residuals to estimate T_n and R_n

Theorem

(i) *Under conditions A1 - A2 and E1 - E4, we have*

$$\widetilde{R}_n \xrightarrow{p} tr\left\{S(\theta^*)^{-1}V(\theta^*)\right\}, \quad as \quad n \to \infty.$$

(ii) *Under the null hypothesis, if A2, B1 - B3 and conditions E1 - E4 hold, we have*

$$\sqrt{n}\left(\widetilde{R}_n - p\right) \xrightarrow{d} N(0, \widetilde{\sigma}_R^2), \quad as \quad n \to \infty,$$

where $\widetilde{\sigma}_R^2$ is the asymptotic variance.

(iii) *Under assumptions A2, B1 - B3, C1 and E1 - E4, we have*

$$\widetilde{R}_n - \widetilde{T}_n(m) = o_p(n^{-1/2}).$$

Assumption $A1 - A2$ Assumptions $B1 - B3$ Assumptions $C1$ Assumption $E1 - E4$

Other important tests to discuss here

- gofSn
- gofRn
- gofKendallCvM
- gofKendallKS
- gofRosenblattSnB
- gofRosenblattSnC
- gofADChisq
- gofADGamma
- Chen et al. (2004)
- gofKernel

gofSn and gofRn and KS test

- Use *empirical process*

$$\mathbb{C}_n(u_1,\dots,u_d)=\sqrt{n}\{C_n(u_1,\dots,u_d)-C_{\widehat{\theta}}(u_1,\dots,u_d)\}$$

- Cramér-von Mises (CvM)

$$S_n^E=\int_{[0,1]^d}\mathbb{C}_n(u_1,\dots,u_d)^2 dC_n(u_1,\dots,u_d)$$

- weighted CvM, with tuning params. $m\geq 0$ and $\zeta_m\geq 0$

$$R_n^E=\int_{[0,1]^d}\left\{\frac{\mathbb{C}_n(u_1,\dots,u_d)}{[C_{\widehat{\theta}}(u_1,\dots,u_d)\{1-C_{\widehat{\theta}}(u_1,\dots,u_d)\}+\zeta_m]^m}\right\}^2 dC_n(u_1,\dots,u_d)$$

- Kolmogorov-Smirnov

$$T_n^E=\sup_{\{u_1,\dots,u_d\}\in[0,1]^d}|\mathbb{C}_n(u_1,\dots,u_d)|$$

Tests based on Kendalls's transform

- Having

$$(X_1, \dots, X_d) \sim F(x_1, \dots, x_d) = C_\theta\{F_1(x_1), \dots, F_d(x_d)\},$$

$X_i \sim F_i(x)$, $F_i(X_i) \sim U(0,1)$

$$C_\theta\{F_1(X_1), \dots, F_d(X_d)\} \sim K_\theta(v)$$

where $K_\theta(v)$ is the *univariate* Kendall's distribution.
- $K_\theta(v)$ is the distribution of the copula as the random variable
- K can be estimated nonparametrically as

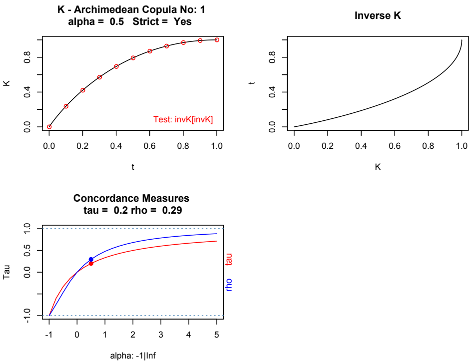
$$K_n(v) = \frac{1}{n}\sum_{i=1}^{n} I(C_n\{\check{F}_1(x_{i1}), \dots, \check{F}_d(x_{id})\} \leq v), \quad v \in [0,1]$$

fCopulae get a feeling using sliders

copula method: 1 - Clayton, 4 - Gumbel, 5 - Frank

- *K*-distribution

```
KfuncSlider()
```



gofKendallCvM and gofKendallKS

- Test $H_0'' : K \in \mathcal{K}_0 = \{K_\theta : \theta \in \Theta\}$
- Use empirical process $\mathbb{K}_n = \sqrt{n}(K_n - K_{\widehat{\theta}})$
- Note that $H_0 \subset H_0'' \Rightarrow$ tests are not generally consistent
- Usual distances
  - Cramér- von Mises

$$S_n^{(K)} = \int_0^1 \mathbb{K}_n(v)^2 dK_{\theta_n}(v)$$

  - Kolmogorov-Smirnov

$$T_n^{(K)} = \sup_{v \in [0,1]} |\mathbb{K}_n(v)|$$

- for bivariate Archimedean copulas H_0'' and H_0 are equivalent

Tests based on Rosenblatt's transform

Recall conditional inverse simulation method, where conditional distribution of U_i is given by

$$\begin{aligned} C_d(u_i|u_1,\ldots,u_{i-1}) &= \boldsymbol{P}\{U_i \le u_i|U_1 = u_1 \ldots U_{i-1} = u_{i-1}\} \\ &= \frac{\partial^{i-1} C(u_1,\ldots,u_i,1,\ldots,1)/\partial u_1 \ldots \partial u_{i-1}}{\partial^{i-1} C(u_1,\ldots,u_{i-1},1,\ldots,1)/\partial u_1 \ldots \partial u_{i-1}} \end{aligned}$$

Definition

Rosenblatt's probability integral transform of a copula C is the mapping $\mathfrak{R} : (0,1)^d \to (0,1)^d$, $\mathfrak{R}(u_1,\ldots,u_d) = (e_1,\ldots,e_d)$ with $e_1 = u_1$ and $e_i = C_d(u_i|u_1,\ldots,u_{i-1})$, $\forall i = 2,\ldots,d$

gofRosenblattSnB and gofRosenblattSnC

- Tests on direct Rosenblatt transformed data, see Genest, Rémillard and Beaudoin (2009, IME)
  - Cramér-von Mises I:

$$S_n = n \int_{[0,1]^d} \{D_n(u) - \Pi(u)\}^2 du$$ – best following Genest et al. (2009)

  - Cramér-von Mises II:

$$S_n^{(C)} = n \int_{[0,1]^d} \{D_n(u) - \Pi(u)\}^2 dD_n(u)$$

  - where the empirical distribution function

$$D_n(u) = D_n(u_1, \dots, u_d) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} I(e_{ij} \le u_j)$$

should be "close" to product copula Π under H_0.

gofADGamma and gofADChisq

- *Anderson-Darling* test statistics:

$$T_n = -n - \sum_{i=1}^{n} \frac{2i-1}{n} [\log G_{(i)} + \log\{1 - G_{(n+1-i)})\}]$$

- where for Gamma

$$G_i = \Gamma_d \left\{ \sum_{j=1}^{d} (-\log e_{ij}) \right\},$$

where $\Gamma_d(\cdot)$ is the Gamma distribution with shape d and scale 1

- where for Chisq

$$G_i = \chi^2_d \left[\sum_{j=1}^{d} \{\Phi^{-1}(e_{ij})\}^2 \right]$$

where χ^2_d Chi-Squared distribution with d degrees of freedom and Φ is standard normal distribution

Chen et al. (2004)

- Test statistics (Chen et al., 2004):

$$C_n^{Ch} = \frac{n\sqrt{h}\widehat{J}_n - c_n}{\sigma} \to \boldsymbol{N}(0,1)$$

where c_n and σ are normalization factors and

$$\widehat{W}_i = \sum_{j=1}^{d} \{\Phi^{-1}(e_{ji})\}^2,$$

$$\widehat{g}_W(w) = \frac{1}{n}\sum_{i=1}^{n} K_h\{w, F_{\chi^2_d}(\widehat{W}_i)\},$$

$$\widehat{J}_n = \int_0^1 \{\widehat{g}_W(w) - 1\}^2 dw$$

gofKernel

- J_n from Scaillet (2007, JoMA)
  - Kernel-based GoF test statistic with fixed smoothing parameter

$$J_n = \int_{[0,1]^d} \{\widehat{c}(u) - K_H * c(u;\widehat{\theta})\} w(u) du,$$

with $*$ convolution operator and $w(u)$ a weight function.
  - $K_H(y) = K(H^{-1}y)/\det(H)$ with K bivariate quadratic kernel
  - $H = 2.6073 n^{-1/6}\widehat{\Sigma}^{1/2}$ with $\widehat{\Sigma}$ sample covariance matrix
  - The copula density is estimated as

$$\widehat{c}(u) = \frac{1}{n}\sum_{t=1}^{n} K_H[u - \{\widetilde{F}_1(X_{t1}), \dots, \widetilde{F}_d(X_{td})\}^\top].$$

Residual-based Bootstrap

Step 1. Generate bootstrap sample $\left\{\epsilon_t^{(k)}, t = 1, \ldots, n\right\}$ from copula $C(u; \widehat{\theta})$ under H_0 with PMLE $\widehat{\theta}$ and estimated marginal distribution $\check{F}$ obtained from original data;

Step 2. Based on $\left\{\epsilon_t^{(k)}, t = 1, \ldots, n\right\}$ from Step 1, estimate θ of the copula under H_0 by the two-step PMLE method, and compute R_n, denoted by R_n^k;

Step 3. Repeat Steps 1 - 2 N times and obtain N statistics $R_n^k, k = 1, \ldots, N$;

Step 4. Compute empirical p-value as $p_e = \frac{1}{N}\sum_{k=1}^{N} \mathbf{I}\left(|R_n^k| \geq |R_n|\right)$.

Simulation Study - Fixed true model setup

- Tests used in the study:
  - S_n
  - J_n
  - R_n
  - $T_n(1)$ and $T_n(3)$
- Copulae: Gaussian, t, Clayton and Gumbel
- $\tau \in \{0.25;\ 0.50;\ 0.75\}$
- $n \in \{100;\ 300\}$
- Rounds of simulation $N = 1000$
- Bootstrap sample paths in every simulation $M = 1000$

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$
$n = 300$	Ga.	Ga.	5.5	4.3	4.4	4.5	4.2
	t	t	4.3	5.1	5.5	4.6	6.2
	Cl.	Cl.	5.0	5.9	6.6	6.5	5.0
	Gu.	Gu.	4.5	3.3	5.2	5.2	5.2
	Ga.	t	5.1	12.4	66.0	61.7	22.4
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3
	t	Ga.	65.7	12.3	95.6	96.3	88.1
	t	Cl.	98.3	100.0	98.0	98.0	86.5
	t	Gu.	88.3	24.7	71.4	72.6	52.7
	Cl.	Ga.	100.0	100.0	100.00	99.8	97.2
	Cl.	t	100.0	98.5	36.6	97.7	75.9
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5

Table 1: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$
$n = 300$	Ga.	Ga.	**5.5**	**4.3**	**4.4**	**4.5**	**4.2**
	t	t	**4.3**	**5.1**	**5.5**	**4.6**	**6.2**
	Cl.	Cl.	**5.0**	**5.9**	**6.6**	**6.5**	**5.0**
	Gu.	Gu.	**4.5**	**3.3**	**5.2**	**5.2**	**5.2**
	Ga.	t	5.1	12.4	66.0	61.7	22.4
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3
	t	Ga.	65.7	12.3	95.6	96.3	88.1
	t	Cl.	98.3	100.0	98.0	98.0	86.5
	t	Gu.	88.3	24.7	71.4	72.6	52.7
	Cl.	Ga.	100.0	100.0	100.00	99.8	97.2
	Cl.	t	100.0	98.5	36.6	97.7	75.9
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5

Table 2: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$
$n = 300$	Ga.	Ga.	**5.5**	**4.3**	**4.4**	**4.5**	**4.2**
	t	t	**4.3**	**5.1**	**5.5**	**4.6**	**6.2**
	Cl.	Cl.	**5.0**	**5.9**	**6.6**	**6.5**	**5.0**
	Gu.	Gu.	**4.5**	**3.3**	**5.2**	**5.2**	**5.2**
	Ga.	t	5.1	12.4	66.0	61.7	22.4
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3
	t	Ga.	65.7	12.3	95.6	96.3	88.1
	t	Cl.	98.3	100.0	98.0	98.0	86.5
	t	Gu.	88.3	24.7	71.4	72.6	52.7
	Cl.	Ga.	100.0	100.0	100.00	99.8	97.2
	Cl.	t	100.0	98.5	36.6	97.7	75.9
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5

Table 3: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$
$n = 300$	Ga.	Ga.	**5.5**	**4.3**	**4.4**	**4.5**	**4.2**
	t	t	**4.3**	**5.1**	**5.5**	**4.6**	**6.2**
	Cl.	Cl.	**5.0**	**5.9**	**6.6**	**6.5**	**5.0**
	Gu.	Gu.	**4.5**	**3.3**	**5.2**	**5.2**	**5.2**
	Ga.	t	5.1	12.4	66.0	61.7	22.4
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3
	t	Ga.	65.7	12.3	95.6	96.3	88.1
	t	Cl.	98.3	100.0	98.0	98.0	86.5
	t	Gu.	88.3	24.7	71.4	72.6	52.7
	Cl.	Ga.	100.0	100.0	100.00	99.8	97.2
	Cl.	t	100.0	98.5	36.6	97.7	75.9
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5

Table 4: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Hybrid Test, I

- Different tests + different situations = Different power
- Hybrid test combines several test methods
- Consider q test statistics $T_n^{(1)}, T_n^{(2)}, \dots, T_n^{(q)}$
- Common H_0 hypothesis and given significance level α
- Hybrid test statistic, T_n^{hybrid}, will have p-value

$$p_n^{\text{hybrid}} = \min\{q \times \min\{p_n^{(1)}, \dots, p_n^{(q)}\}, 1\}$$

- Rejection rule: $p_n^{\text{hybrid}} \leq \alpha$

Hybrid Test, II

- Type I error:

$$P(p_n^{(hybrid)} \leq \alpha | H_0) \leq \alpha$$

- Type II error:

$$P(p_n^{hybrid} \leq \alpha | H_1) \geq \max \left\{ \beta_n^1(\alpha/q), \dots, \beta_n^q(\alpha/q) \right\}$$

- Implication: If at least one test is consistent, hybrid test is consistent as well
- Simulation study shows that the Hybrid Test behaves more desirably than the individual tests

Simulation Study - cont.

- Bootstrap technique to numerically establish the null distribution of the test statistics

- Applied single tests:
  - S_n
  - J_n
  - R_n
  - $T_n(1)$ and $T_n(3)$

- Applied hybrid tests:
  - SR_n
  - $ST_n(1)$
  - JR_n
  - $JT_n(1)$
  - SJR_n
  - $SJT_n(1)$

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$	SR_n	$ST_n(1)$	JR_n	$JT_n(1)$	SJR_n	$SJT_n(1)$
$n = 300$	Ga.	Ga.	5.5	4.3	4.4	4.5	4.2	4.7	4.2	4.3	4.1	5.6	5.7
	t	t	4.3	5.1	5.5	4.6	6.2	5.6	4.5	4.7	5.1	5.1	4.7
	Cl.	Cl.	5.0	5.9	6.6	6.5	5.0	5.5	5.5	3.5	3.5	3.2	3.2
	Gu.	Gu.	4.5	3.3	5.2	5.2	5.2	4.4	4.3	4.5	4.3	5.1	5.1
	Ga.	t	5.1	12.4	66.0	61.7	22.4	55.3	46.4	58.3	50.3	51.2	42.9
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5	98.3	98.3	100.0	100.0	100.0	100.0
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3	49.5	49.1	26.8	26.9	57.9	57.9
	t	Ga.	65.7	12.3	95.6	96.3	88.1	92.9	93.7	93.2	94.0	91.9	92.5
	t	Cl.	98.3	100.0	98.0	98.0	86.5	99.6	99.6	100.0	100.0	100.0	100.0
	t	Gu.	88.3	24.7	71.4	72.6	52.7	88.3	88.3	67.9	68.1	83.1	83.1
	Cl.	Ga.	100.0	100.0	100	99.8	97.2	100.0	100.0	100.0	100.0	100.0	100.0
	Cl.	t	100.0	98.5	36.6	97.7	75.9	100.0	100.0	97.9	99.6	100.0	100.0
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4	83.1	80.0	82.8	82.1	79.7	78.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9	32.2	31.8	19.6	19.5	30.4	29.2
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5	100.0	100.0	100.0	100.0	100.0	100.0

Table 5: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

PIOS for SCOMDY model

- True data-generating processes are GARCH(1,1):

$$x_{it} = \sigma_{it}\varepsilon_{it}$$
$$\sigma^2_{it} = \omega + \alpha x^2_{i,t-1} + \beta\sigma^2_{i,t-1}, \quad \text{for } i = 1, 2$$

  with $\{\varepsilon_{1t}, \varepsilon_{2t}\} \sim C\{F_1(\cdot), F_2(\cdot); \theta\}$, $\varepsilon_{i,t} \perp \varepsilon_{i,t-1}$ for $i = 1, 2$.
- $\omega = 10^{-1}$, $\alpha = 0.1$ and $\beta = 0.8$

1. Simulated *iid* samples in bootstrap loop
2. Bootstrap loop with time series structure

Observation-based Bootstrap

Step 1. Generate time series $\{Y_t^{(k)}, t = 1, \ldots, n\}$ from SCOMDY model with $\widehat{\eta}_1$ and $\widehat{\eta}_2$ estimated from original data, and with innovation process generated from assumed copula under H_0 with $\widehat{\theta}$ and marginal distribution $\check{F}$.

Step 2. Based on $\{Y_t^{(k)}, t = 1, \ldots, n\}$, estimate $\widehat{\eta}_1^{(k)}$ and $\widehat{\eta}_2^{(k)}$. Estimate residuals $\widetilde{\epsilon}_{tj}^{(k)} = \{y_{tj}^{(k)} - \mu_{tj}(\widehat{\eta}_1^{(k)})\}/\Sigma_{tj}^{1/2}(\widehat{\eta}_2^{(k)})$.

Step 3. Based on $\{\widetilde{\epsilon}_t^{(k)}, t = 1, \ldots, n\}$, estimate θ of copula under H_0 by two-step PMLE method and compute R_n^k;

Step 4. Repeat Steps 1- 3 N times and obtain N statistics $R_n^k, k = 1, \ldots, N$;

Step 5. Compute empirical p-value as $p_e = \frac{1}{N}\sum_{k=1}^{N} \mathbf{I}\left(|R_n^k| \geq |R_n|\right)$.

SCOMDY, I

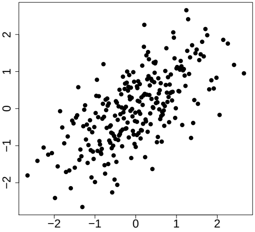
True	H_0	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
		R_n	$T_n(1)$	R_n	$T_n(1)$	R_n	$T_n(1)$
Ga	Ga	*0.062*	*0.059*	*0.058*	*0.066*	*0.085*	*0.088*
		0.058	0.061	0.046	0.043	0.042	0.041
Cl	Cl	*0.058*	*0.052*	*0.061*	*0.068*	*0.113*	*0.113*
		0.053	0.057	0.038	0.039	0.050	0.050
t	t	*0.054*	*0.053*	*0.048*	*0.044*	*0.062*	*0.043*
		0.042	0.043	0.052	0.060	0.049	0.046
Gu	Gu	*0.054*	*0.056*	*0.055*	*0.052*	*0.070*	*0.069*
		0.052	0.055	0.048	0.049	0.046	0.045

Table 6: Percentages of rejection of H_0 by various tests from different copula models for $n = 300$, $N = 300$, $M = 1000$ for the GARCH(1,1) dependent data. Type I errors were obtained using residual-based (in italic) and observation-based bootstrap procedures.

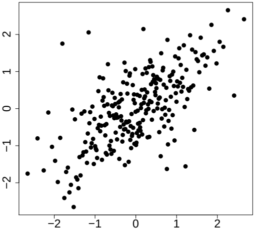
Application: Structural changes in the dependency

- Daily returns of Citigroup and Bank of America
- Period 2004 – 2013
- Apply GARCH(1,1) to each year separately
- Chosen is the copula dependency with the largest p-value for each year

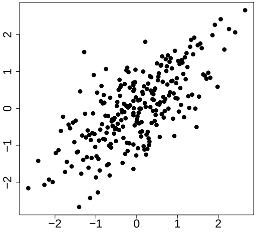
Scatterplots



(a) year 2004



(b) year 2006



(c) year 2009

Figure 6: Scatterplots of residuals transformed to the standard normal for Citygroup/Bank of America for 2004, 2006 and 2009.

Results

	$T_n(1)$	R_n	S_n	J_n	$ST_n(1)$	SR_n	$JT_n(1)$	JR_n	$SJT_n(1)$	SJR_n
2004	Gu.	**Gu.**	**Ga.**	**Ga.**	Ga.	Ga.	Gu.	Gu.	**Ga.**	**Ga.**
2005	Gu.	Gu.	*t*	*t*	Gu.	Gu.	Gu.	Gu.	Gu.	Gu.
2006	*t*	**t**	**Ga.**	**t**	*t*	*t*	*t*	*t*	**t**	**t**
2007	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
2008	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
2009	Gu.	**Gu.**	**Gu.**	**Ga.**	Gu.	Gu.	Gu.	Gu.	**Gu.**	**Gu.**
2010	*t*	*t*	Gu.	*t*	Gu.	Gu.	*t*	*t*	Gu.	Gu.
2011	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
2012	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
2013	*t*	*t*	*t*	Gu.	*t*	*t*	*t*	*t*	*t*	*t*

Table 7: Copulas that are preferred in each time period by each goodness-of-fit test for the Citigroup / Bank of America.

R-package `gofCopula`

- Most tests in one package
- Several margin structures
- Each test at least 3 dim
- Automatized parallelization (in progress)

Covered single tests

- gofRosenblattSnB
- gofRosenblattSnC
- gofKendallCvM
- gofKendallKS
- gofPIOSRn
- gofPIOSTn
- gofADChisq
- gofADGamma
- gofSn
- gofRn
- gofKernel
- gofWhite

Hybrid test - computation

- Random 2 dim sample

```
x = cbind(rnorm(100), rnorm(100))
```

- Hybrid test for normality with 2 tests

```
gofHybrid("gaussian", x, testset = c("gofRosenblattSnB"
  , "gofRosenblattSnC"), M = 1000)
```

- Computation time

```
[1] "The computation will take approximately 0 d, 0 h,
  2 min and 14 sec."
```

Hybrid test - result

```
Tests results:
p.value test statistic
RosenblattSnB 0.9515485      0.05416868
RosenblattSnC 0.8786214      0.07630096
hybrid(1, 2)  1.0000000             NaN
```

- Automatic margin estimation

```
Warning message:
In gofHybrid("gaussian", x, testset = c("
  gofRosenblattSnB", "gofRosenblattSnC"),  :
The observations are not in [0,1]. The margins will be
  estimated by the ranks of the observations.
```

Tests for copulae

- Available tests for gaussian?

```
gofWhich("gaussian", d = 2)
[1] "gofHybrid" "gofRosenblattSnB"
[3] "gofRosenblattSnC" "gofADChisq" "gofADGamma"
[6] "gofSn" "gofRn" "gofPIOSRn" "gofPIOSTn"
[10] "gofKernel" "gofWhite" "gofKendallCvM"
[13] "gofKendallKS"
```

- Use all for hybrid test

```
gofHybrid("gaussian", x, testset = gofWhich("gaussian",
   d = 2)[-1], M = 1000)
[1] "The computation will take approximately 0 d, 2 h,
  20 min and 15 sec."
```

Flexible testing structure

- Adjust margins

```
margins = "gaussian"
gofHybrid("gaussian", x,
  testset = c("gofRosenblattSnB",
              "gofRosenblattSnC"),
  M       = 1000,
  margins = margins)
```

Flexible testing structure

- Fix parameter

```
parameter = 0.2
gofHybrid("gaussian", x,
  testset    = c("gofRosenblattSnB",
                 "gofRosenblattSnC"),
  M          = 1000,
  param.est  = FALSE,
  param      = parameter)
```

Copulae for tests

- Available copulae for test?

```
gofWhichCopula("gofRosenblattSnB")
[1] "gaussian" "t" "clayton" "frank" "gumbel"

gofWhichCopula("gofRosenblattSnC")
[1] "gaussian" "t" "clayton" "frank" "gumbel"
```

Copulae for tests

- Use a test with all copulae

```
copulae = gofWhichCopula("gofRosenblattSnB")
for (i in copulae){
  print(gofHybrid(i, x,
    testset = c("gofRosenblattSnB",
                "gofRosenblattSnC"),
    M       = 10))
}
```

- OR

```
copulae = gofWhichCopula("gofRosenblattSnB")
gof(x,
  copula = copulae,
  tests  = c("gofRosenblattSnB", "gofRosenblattSnC"),
  M      = 10)
```

gof

- Options of gof
  - priority $\in$ {"tests", "copula"}
    - "tests": all tests for their shared copulae
    - "copula": all tests which support {"gaussian", "t", "gumbel", "clayton", "frank"}
  - copula: which copulae to use
  - tests: which tests to use
- priority just in effect if copula = tests = NULL

Interface to copula package

- Connection to copula package
- Usage of both packages may be desirable

```
copulaobject = normalCopula(param = 0.2, dim = 2)
gofco(copulaobject, x,
  testset = c("gofRosenblattSnB", "gofRosenblattSnC"),
  M = 1000)

Tests results:
p.value test statistic
RosenblattSnB 0.9475524     0.05416868
RosenblattSnC 0.8366633     0.07630096
hybrid(1, 2)  1.0000000            NaN
```

Multivariate Copula families

- Gaussian copula
  - No tail dependence and correlation matrix.
- t-copula
  - One parameter for all tail areas plus correlation matrix.
- Factor copula, Oh and Patton (2014)
  - Flexible, but no density/conditional quantile.
- Vines, Kurowicka and Joe (2011)
  - Flexible, but need $d(d-1)/2$ parameters.
- d-dimensional Archimedean Copulae
  - too restrictive: single parameter, exchangeable
- **HAC**

Hierarchical Archimedean copulae

Papers:

Okhrin, O., Okhrin, Y. and Schmid, W., Determining the structure and estimation of hierarchical Archimedean copulas, *Journal of Econometrics* 173(2), 2013, pp. 189-204.

Okhrin, O. and Ristig, A., Hierarchical Archimedean Copulae: The HAC Package, *Journal of Statistical Software* 58(4), 2014, pp 1-20.

Okhrin, O., Ristig, A., Sheen, J., and Trueck, S., Conditional Systemic risk with penalized copula, *working paper*

Hofert, M., Sampling Archimedean copulas. *Computational Statistics and Data Analysis* 52, 2008, pp. 5163-5174.

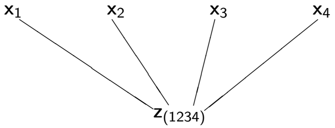
Main Idea of HAC

- combine interpretability with flexibility without loosing statistical precision
- determine the optimal structure of HAC
- convenient and useful probabilistic properties of the HAC

Hierarchical Archimedean Copulae

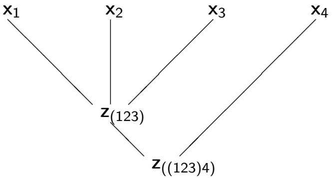
Simple AC with s=(1234)

$$C(u_1, u_2, u_3, u_4) = C_1(u_1, u_2, u_3, u_4)$$



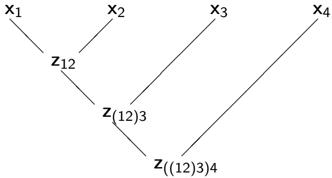
AC with s=((123)4)

$$C(u_1, u_2, u_3, u_4) = C_1\{C_2(u_1, u_2, u_3), u_4\}$$



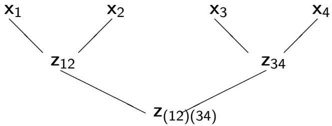
Fully nested AC with s=(((12)3)4)

$$C(u_1, u_2, u_3, u_4) = C_1[C_2\{C_3(u_1, u_2), u_3\}, u_4]$$



Partially Nested AC with s=((12)(34))

$$C(u_1, u_2, u_3, u_4) = C_1\{C_2(u_1, u_2), C_3(u_3, u_4)\}$$



Hierarchical Archimedean Copula

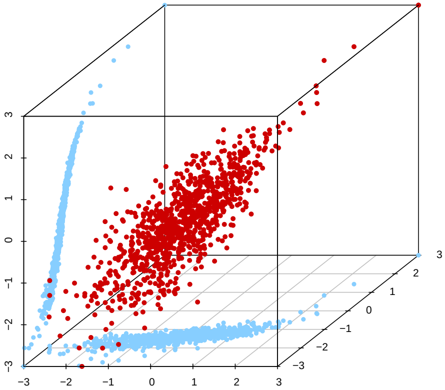


Figure 7: Scatterplot of the $C_{Gumbel}[C_{Gumbel}\{\Phi(x_1), t_2(x_2); \theta_1 = 10\}, \Phi(x_3); \theta_2 = 2]$, $s = ((12)_{10}3)_2$

Hierarchical Archimedean Copula

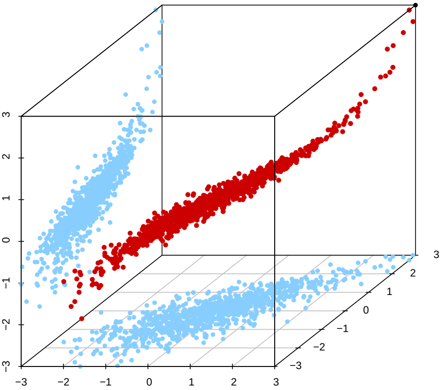


Figure 8: Scatterplot of the $C_{Gumbel}[\Phi(x_2), C_{Gumbel}\{t_2(x_1), \Phi(x_3); \theta_1 = 10\}; \theta_2 = 2]$, $s = (2_{10}(13)_2)$

Hierarchical Archimedean Copula

Advantages of HAC:

- flexibility and wide range of dependencies:
  for $d = 10$ more than $2.8 \cdot 10^8$ structures
- dimension reduction:
  $d - 1$ parameters to be estimated
- subcopulae are also HAC

Theoretical motivation

Let M be the cdf of a positive random variable and ϕ denotes its Laplace transform, i.e. $\phi(t) = \int_0^\infty \exp^{-tw} dM(w)$. For an arbitrary cdf F there exists a unique cdf G, such that

$$F(x) = \int_0^\infty G^\alpha(x) dM(\alpha) = \phi\{-\log G(x)\}.$$

Now consider a d-variate cdf F with margins $F_1, \ldots, F_d$. Then it holds for $G_j = \exp\{-\phi^{-1}(F_j)\}$ that

$$\int_0^\infty G_1^\alpha(x_1) \cdot \cdots \cdot G_d^\alpha(x_d) dM(\alpha) = \phi\left\{-\sum_{j=1}^d \log G_j(x_j)\right\} = \phi\left[\sum_{j=1}^d \phi^{-1}\{F_j(x_j)\}\right].$$

$$C(u_1, \ldots, u_d) =$$

$$\int_0^\infty \ldots \int_0^\infty G_1^{\alpha_1}(u_1) G_2^{\alpha_1}(u_2) dM_1(\alpha_1, \alpha_2)\; G_3^{\alpha_2}(u_3) dM_2(\alpha_2, \alpha_3) \ldots G_d^{\alpha_{d-1}}(u_d) dM_{d-1}(\alpha_{d-1}).$$

Recovering the structure (theory)

To guarantee that C is a HAC we assume that $\phi_{d-i}^{-1} \circ \phi_{d-j} \in \mathcal{L}^*$, $i < j$ with

$$\mathcal{L}^* = \{\omega : [0, \infty) \to [0, \infty) \,|\, \omega(0) = 0,\ \omega(\infty) = \infty,\ (-1)^{j-1}\omega^{(j)} \geq 0,\ j \geq 1\}.$$

For most of the generator functions the parameters should decrease from the lowest level to the highest

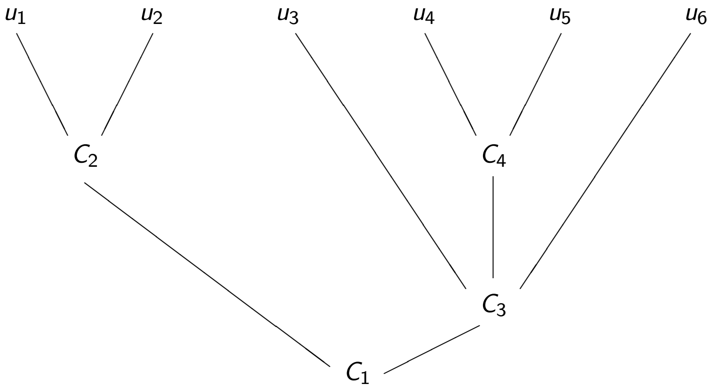
Theorem

Let F be an arbitrary multivariate distribution function based on HAC. Then F can be uniquely recovered from the marginal distribution functions and all bivariate copula functions.

$$C(u_1, \dots, u_6) = C_1[C_2(u_1, u_2), C_3\{u_3, C_4(u_4, u_5), u_6\}].$$

The bivariate marginal distributions are then given by

$$\begin{array}{lll}
(U_1, U_2) \sim C_2(\cdot,\cdot), & (U_2, U_3) \sim C_1(\cdot,\cdot), & (U_3, U_5) \sim C_3(\cdot,\cdot), \\
(U_1, U_3) \sim C_1(\cdot,\cdot), & (U_2, U_4) \sim C_1(\cdot,\cdot), & (U_3, U_6) \sim C_3(\cdot,\cdot), \\
(U_1, U_4) \sim C_1(\cdot,\cdot), & (U_2, U_5) \sim C_1(\cdot,\cdot), & (U_4, U_5) \sim C_4(\cdot,\cdot), \\
(U_1, U_5) \sim C_1(\cdot,\cdot), & (U_2, U_6) \sim C_1(\cdot,\cdot), & (U_4, U_6) \sim C_3(\cdot,\cdot), \\
(U_1, U_6) \sim C_1(\cdot,\cdot), & (U_3, U_4) \sim C_3(\cdot,\cdot), & (U_5, U_6) \sim C_3(\cdot,\cdot).
\end{array}$$


u_1
u_2
u_3
u_4
u_5
u_6
C_2
C_4
C_3
C_1

$$\mathcal{C}_2\{\boldsymbol{N}(C)\} = \{C_1(\cdot,\cdot), C_2(\cdot,\cdot), C_3(\cdot,\cdot), C_4(\cdot,\cdot)\}.$$

- each variable belongs to at least one bivariate margin C_1 $\rightsquigarrow$ the distribution of $u_1, \dots, u_6$ has C_1 at the top level.
- C_3 covers the largest set of variables u_3, u_4, u_5, u_6 $\rightsquigarrow$ C_3 is at the top level of the subcopula containing u_3, u_4, u_5, u_6.

$$U_1, \dots, U_6 \sim C_1\{u_1, u_2, C_3(u_3, u_4, u_5, u_6)\}.$$

- C_2 and C_4 and they join u_1, u_2 and u_4, u_5 respectively.

$$(U_1, \dots, U_6) \sim C_1[C_2(u_1, u_2), C_3\{u_3, C_4(u_4, u_5), u_6\}]$$

Let for each bivariate copula $C^* \in \mathcal{C}_2\{\boldsymbol{N}(C)\}$, $I(C)$ be the set of indices $i \in \{1, \dots, k\}$ such that $(U_i, U_j) \sim C^*$ for at least one $j \in \{1, \dots, k\} \backslash \{i\}$.

$I(C_1) = \{1, \dots, 6\},\ I(C_2) = \{1, 2\},\ I(C_3) = \{3, 4, 5, 6\},\ I(C_4) = \{4, 5\}$

The family of sets $I(C^*)$, as C^* ranges over $\mathcal{C}_2\{\boldsymbol{N}(C)\}$, is partially ordered by inclusion

$$I(C_1) \supset \begin{cases} I(C_2), \\ I(C_3) \supset I(C_4). \end{cases}$$

Recovering the structure (practice)

$(12) \rightsquigarrow \widehat{\theta}_{12}$
$(13) \rightsquigarrow \widehat{\theta}_{13}$
$(14) \rightsquigarrow \widehat{\theta}_{14}$
$(23) \rightsquigarrow \widehat{\theta}_{23}$
$(24) \rightsquigarrow \widehat{\theta}_{24}$
$(34) \rightsquigarrow \widehat{\theta}_{34}$

$(123) \rightsquigarrow \widehat{\theta}_{123}$
$(124) \rightsquigarrow \widehat{\theta}_{124}$
$(234) \rightsquigarrow \widehat{\theta}_{234}$
$(134) \rightsquigarrow \widehat{\theta}_{134}$
$(1234) \rightsquigarrow \widehat{\theta}_{1234}$

best fit (13) $\rightsquigarrow$

$$z_{(13),i} = \widehat{C}\{\widehat{F}_1(x_{1i}), \widehat{F}_3(x_{3i})\}$$

$(13)2 \rightsquigarrow \widehat{\theta}_{(13)2}$
$(13)4 \rightsquigarrow \widehat{\theta}_{(13)4}$
$24 \rightsquigarrow \widehat{\theta}_{24}$

$(13)24 \rightsquigarrow \widehat{\theta}_{(13)24}$

best fit ((13)4) $\rightsquigarrow$

$$z_{((13)4),i} = \widehat{C}\{z_{(13)i}, \widehat{F}_4(x_{4i})\}$$

$((13)4)2 \rightsquigarrow \widehat{\theta}_{((13)4)2}$

Estimation: multistage MLE with nonparametric and parametric margins

Criteria for grouping: goodness-of-fit tests, parameter-based method, etc.

Estimation Issues - Multistage Estimation

$$\left(\frac{\partial \mathcal{L}_1}{\partial \boldsymbol{\theta}_1^\top}, \dots, \frac{\partial \mathcal{L}_p}{\partial \boldsymbol{\theta}_p^\top}\right)^\top = \mathbf{0},$$

$$\begin{aligned} \text{where} \quad \mathcal{L}_j &= \sum_{i=1}^n l_j(\mathbf{X}_i) \\ l_j(\mathbf{X}_i) &= \log\left(c(\{\phi_\ell, \boldsymbol{\theta}_\ell\}_{\ell=1,\dots,j};\, s_j)\left[\{\check{F}_m(x_{mi})\}_{m\in s_j}\right]\right) \\ &\quad \text{for } j = 1,\dots,p. \end{aligned}$$

Theorem

Under regularity conditions, estimator $\widehat{\boldsymbol{\theta}}$ is consistent and

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{\mathcal{L}}{\to} N(\mathbf{0}, \mathrm{B}^{-1}\Sigma\mathrm{B}^{-1})$$

Criteria for grouping

Alternatives:

- goodness-of-fit tests
  - dimension dependent
  - computationally complicated
- distance measures
  - dimension dependent
- parameter-based methods

  Note that, if the true structure is (123) then

  $$\theta_{(12)} = \theta_{(13)} = \theta_{(23)} = \theta_{(123)}.$$

  - heuristic methods
  - test-based methods
- tests on exchangeability

Criteria for grouping based on θ's

I. For all subsets perform tests of the kind

$$
\begin{aligned}
H_0: &\quad \theta_{(12)} = \theta_{(13)} = \theta_{(23)} = \theta_{(123)} \\
H_1: &\quad \text{at least one equality is not fulfilled}
\end{aligned}
$$

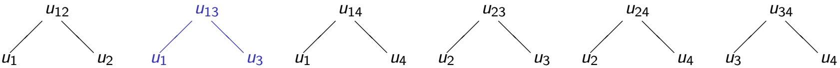
II.

$$
\Delta = \min_{I_{ki},\, |I_{ki}| \geq 3} \; \max_{I_{|I_{ki}|,j} \subset I_{ki}} |\theta(I_{ki}) - \theta(I_{|I_{ki}|,j})|,
$$

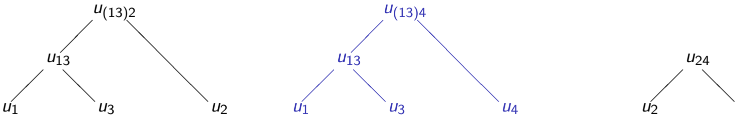
where $j = 1, \dots, 2^{|I_{ki}|} - |I_{ki}| - 1$ and $\{I_{ki}\}_{i=1,\dots 2^k - k - 1}$ denote the subsets of the initial set of size k, excluding empty set and single element sets.

$$
I^* = \begin{cases} I_\Delta, & \Delta \leq \delta \\ \max_{I_{ki},\, |I_{ki}|=2} \theta(I_{ki}), & \Delta > \delta \end{cases}.
$$

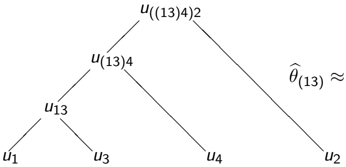
Estimation of HAC



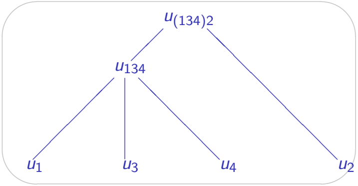
$$\max\{\widehat{\theta}_{12},\widehat{\theta}_{13},\widehat{\theta}_{14},\widehat{\theta}_{23},\widehat{\theta}_{24},\widehat{\theta}_{34}\}=\widehat{\theta}_{13} \quad \Rightarrow$$



$$\max\{\widehat{\theta}_{(13)2},\widehat{\theta}_{(13)4},\widehat{\theta}_{24}\}=\widehat{\theta}_{(13)4} \quad \Rightarrow$$



$$\widehat{\theta}_{(13)} \approx \widehat{\theta}_{(13)4} \Rightarrow$$



Simulation, I

Method	Copula structure(s)	%	KL	Kendall τ	λ_U
Gauss			0.2847 (0.0403)	0.0789 (0.0299)	3.0261 (0.0000)
t			0.1989 (0.0283)	0.0682 (0.0268)	1.3398 (0.1500)
sAC	$(12345)_{2.32}$	100.0	0.8038 (0.0711)	0.5249 (0.0097)	0.4710 (0.0078)
$\tau_{\triangle\tau>0}$	$((45)_{3.01}(123)_{4.14})_{2.23}$	100.0	0.0183 (0.0124)	0.1825 (0.0710)	0.1740 (0.0675)
Chen	$(45(123)_{4.01})_{2.02}$	18.3	0.7261 (0.2880)	0.5460 (0.1505)	0.4929 (0.1337)
	$(12(345)_{2.18})_{2.01}$	14.8			
	$(23(145)_{2.19})_{2.03}$	13.3			
θ_{binary}	$((45)_{3.01}(3(12)_{4.11})_{3.91})_{2.27}$	34.7	0.0215 (0.0073)	0.2117 (0.0550)	0.2024 (0.0520)
	$((45)_{3.01}(1(23)_{4.11})_{3.91})_{2.28}$	33.4			
	$((45)_{3.01}(2(13)_{4.11})_{3.90})_{2.28}$	31.9			
θ_{PML}	$((123)_{4.04}(45)_{3.01})_{1.97}$	82.2	-0.0029 (0.0030)	0.0542 (0.0310)	0.0515 (0.0312)
	$((1(23)_{4.2})_{3.94}.(4.5)_{3.03})_{1.99}$	6.4			
	$((3(12)_{4.23})_{3.96}.(4.5)_{2.99})_{1.97}$	5.8			

Table 8: Model fit for the true structure $((123)_4(45)_3)_2$.

Simulation, II

Method	Copula structure(s)	%	KL	Kendall τ	λ_U
Gauss			0.2896 (0.0418)	0.6417 (0.0298)	3.0882 (0.0000)
t			0.1992 (0.0281)	0.6442 (0.0279)	1.4610 (0.1430)
sAC	$(12345)_{2.37}$	100.0	0.4963 (0.0664)	0.4338 (0.0147)	0.3938 (0.0127)
$\tau_{\triangle\tau>0}$	$(5(12(34)_{4.03})_{3.33})_{2.18}$	98.8	0.0318 (0.0219)	0.1627 (0.0582)	0.1488 (0.0539)
	$(5(1234)_{3.58})_{2.17}$	1.1			
	$(5(1(2(34)_{3.97})_{3.74})_{3.50})_{2.24}$	0.1			
Chen	$(15(234)_{3.20})_{2.10}$	12.1	0.4512 (0.1377)	0.4939 (0.0891)	0.4503 (0.082)
	$(13(245)_{2.18})_{2.09}$	11.0			
	$(34(125)_{2.18})_{2.08}$	10.9			
θ_{binary}	$(5((34)_{4.00}(12)_{3.07})_{3.07})_{1.78}$	38.9	0.0312 (0.0163)	0.2196 (0.0562)	0.2152 (0.0562)
	$(5(2(1(34)_{4.03})_{3.06})_{2.59})_{1.75}$	32.3			
	$(5(1(2(34)_{4.03})_{3.06})_{2.59})_{1.75}$	28.8			
θ_{PML}	$(5(12(34)_{4.01})_{3.02})_{2.00}$	81.1	-0.0025 (0.0032)	0.0509 (0.0253)	0.0472 (0.0244)
	$(5((12)_{3.12}(34)_{4.03})_{2.87})_{1.99}$	17.5			
	$(5(1(2(34)_{4.17})_{3.19})_{2.97})_{2.02}$	0.9			

Table 9: Model fit for the true structure $((12(34)_4)_3 5)_2$.

Simulation, III

	$\widehat{\theta}_3,\ (\theta_3 = 4.0)$	$\widehat{\theta}_2,\ (\theta_2 = 3.0)$	$\widehat{\theta}_1,\ (\theta_1 = 2.0)$	Time (in s)
		Structure $((123)_4(45)_3)_2$		
MStage	4.028 (0.103)	3.010 (0.112)	1.967 (0.058)	0.496 (0.032)
Full	4.002 (0.100)	3.010 (0.111)	2.002 (0.058)	0.949 (0.060)
		Structure $((12(34)_4)_35)_2$		
MStage	3.983 (0.148)	2.995 (0.078)	2.003 (0.061)	1.995 (0.372)
Full	3.980 (0.141)	3.004 (0.070)	2.005 (0.061)	2.740 (0.326)

Table 10: The average parameters and computational times for multistage ML and full ML estimation based on 1000 simulated samples of size 500. Standard errors provided in brackets.

Misspecification

Let $H(x_1, \dots, x_k)$ – true df with density h. Since H is unknown we specify $F(x_1, \dots, x_k, \boldsymbol{\eta})$ with density f.

- F is correctly specified:
  $\exists \boldsymbol{\eta}_0 : F(x_1, \dots, x_k, \boldsymbol{\eta}_0) = H(x_1, \dots, x_k),\ \forall (x_1, \dots, x_k)$ then $\widehat{\boldsymbol{\eta}}$ is consistent for $\boldsymbol{\eta}_0$.
- F is not correctly specified:
  $\nexists \boldsymbol{\eta}_0 : F(x_1, \dots, x_k, \boldsymbol{\eta}_0) = H(x_1, \dots, x_k),\ \forall (x_1, \dots, x_k)$, then $\widehat{\boldsymbol{\eta}}$ is an estimator for $\boldsymbol{\eta}_*$ which minimizes the Kullback–Leibler divergence between f and h as

$$\mathcal{K}(h, f, \boldsymbol{\eta}) = \mathsf{E}_h\{\log[h(x_1, \dots, x_k)/f(x_1, \dots, x_k, \boldsymbol{\eta})]\},$$

Misspecification, I

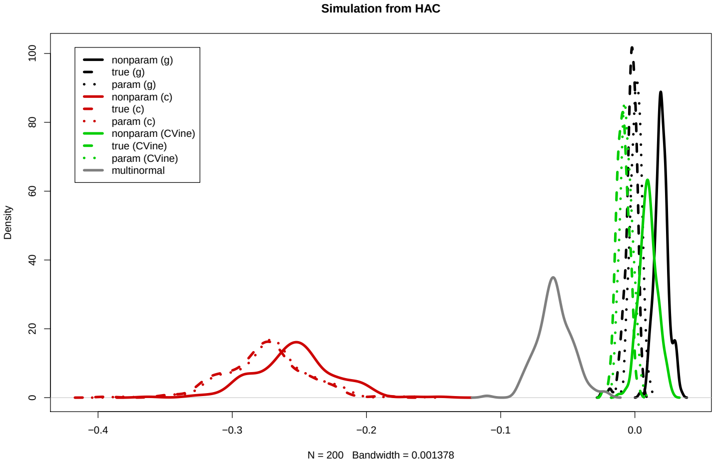


Figure 9: Kullback-Leibler divergences for the simulated samples, HAC with Gumbel generators, $\theta_1 = 2.0, \theta_2 = 1.5, N = 200, n = 1000$

Misspecification, II

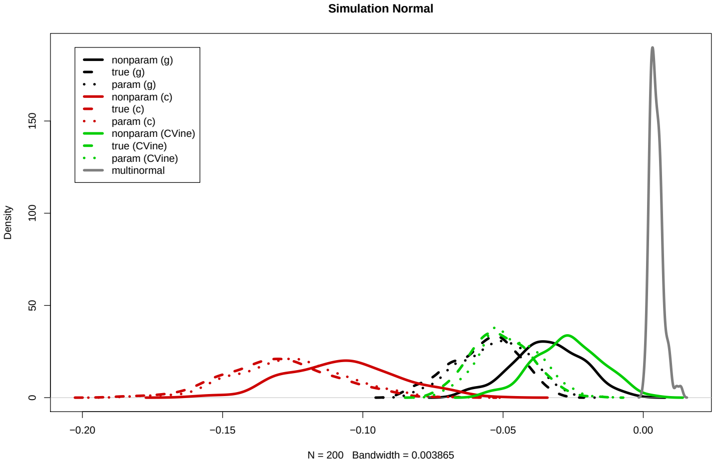
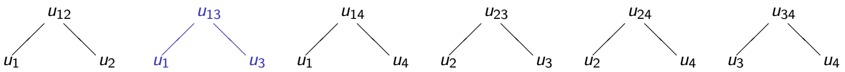
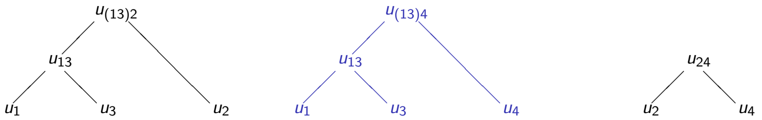


Figure 10: Kullback-Leibler divergences for the simulated samples, HAC, Σ equal to first model, $N = 200, n = 1000$

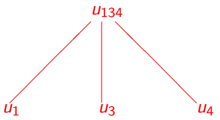
Penalized estimation of HAC



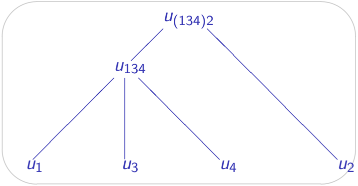
$$\max\{\widehat{\theta}_{12}, \widehat{\theta}_{13}, \widehat{\theta}_{14}, \widehat{\theta}_{23}, \widehat{\theta}_{24}, \widehat{\theta}_{34}\} = \widehat{\theta}_{13} \quad \Rightarrow$$

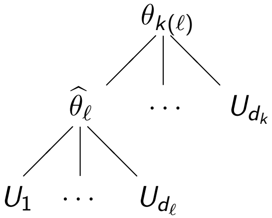


$$\max\{\widehat{\theta}_{(13)2}, \widehat{\theta}_{(13)4}, \widehat{\theta}_{24}\} = \widehat{\theta}_{(13)4}, \quad \text{if } \widehat{\theta}_{13} - \widehat{\theta}_{(13)4} < \epsilon_n \quad \Rightarrow$$



$\Longrightarrow$





- Build $\ell_i(\theta_{k(\ell)}) = \log\{c(U_{i1}, \dots, U_{id_k}; \theta_{k(\ell)})\}$. ▸ Assumptions
- Penalized log-likelihood

$$\mathcal{Q}(\theta_\ell, \theta_{k(\ell)}) = \sum_{i=1}^{n} \ell_i(\theta_{k(\ell)}) - np_{\lambda_n}(\theta_\ell - \theta_{k(\ell)}),$$

  c.f. Cai and Wang (2014, JASA), Fan and Li (2001, JASA), Tibshirani et al. (2005, JRSSB).
- Let $\widehat{\theta}^{\lambda_n}_{k(\ell)}$ be the maximizer of $\mathcal{Q}(\widehat{\theta}_\ell, \theta_{k(\ell)})$.

Sparsity and oracle property

Proposition

Under Assumptions 1-3, if $n^{1/2}\lambda_n \to \infty$ *as* $n \to \infty$*, then*

$$\lim_{n\to\infty} \mathrm{P}(\widehat{\theta}^{\lambda_n}_{k(\ell)} = \theta_{\ell,0}) = 1.$$

Proposition

Under Assumptions 1-3, if $\lambda_n \to 0$ *as* $n \to \infty$*, then*

$$n^{1/2}\{\widehat{V}(\theta_{k(\ell),0}) + p''_{\lambda_n}(\theta_0^-)\}\big[(\widehat{\theta}^{\lambda_n}_{k(\ell)} - \theta_{k(\ell),0}) - \big\{\widehat{V}(\theta_{k(\ell),0}) + p''_{\lambda_n}(\theta_0^-)\big\}^{-1} p'_{\lambda_n}(\theta_0^-)\big] \xrightarrow{\mathcal{L}} \boldsymbol{N}\{0, V(\theta_{k(\ell),0})\},$$

where $\theta_0^- = \theta_{\ell,0} - \theta_{k(\ell),0}$.

ML representation

- Let $\widehat{\theta}_{k(\ell)}$ and $\widehat{\theta}_{\ell}$ be the MLE of Okhrin et al. (2013, JoE).
- Linear approximation of penalty function, Zou and Li (2008, Ann.).

Proposition

Under Assumptions 1-3, $\widehat{\theta}^{\lambda_n}_{k(\ell)} = \widehat{\theta}_{k(\ell)} + \epsilon_n$, *with*

$$\epsilon_n \stackrel{\text{def}}{=} \epsilon(\lambda_n, a_n) = \widehat{V}(\widehat{\theta}_{k(\ell)})^{-1} p'_{\lambda_n}(\widehat{\theta}_{\ell} - \widehat{\theta}_{k(\ell)}).$$

Practical issues

- Attain sparsity from

$$\widehat{\theta}_{k(\ell)} = \widehat{\theta}_{\ell}, \quad \text{if} \quad \widehat{\theta}_{\ell} - \widehat{\theta}_{k(\ell)} \leq \epsilon_n.$$

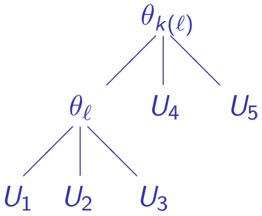
- Wang et al. (2007, Biometrica), determine $(\lambda, a)\top$ by

$$(\lambda_n, a_n)\top = \arg \max_{(\lambda,a)\top} 2 \sum_{i=1}^{n} \ell_i \left\{ \widehat{\theta}_{k(\ell)} + \epsilon(\lambda, a) \right\} - q_k \log(n).$$

- q_k parameters in HAC up to level k.

Setup

- Until $m = 1000$ structures correctly specified.
- Sample size $n = 250$.
- Let $\tau : \Theta_{k(\ell)} \to [0,1]$ transform the parameter $\theta_{k(\ell)}$ into Kendall's correlation coefficient.



- $\theta_\ell = \tau^{-1}(0.7)$ and $\theta_{k(\ell)} = \tau^{-1}(0.3)$.

Family	$s(\widehat{\theta}) = s(\theta_0)$	$\tau(\widehat{\theta}_1)$ (sd)	$\tau(\widehat{\theta}_2)$ (sd)	$\#\{\widehat{\theta}\}$
Clayton	0.82	0.70 (0.01)	0.30 (0.02)	3.04
Frank	0.85	0.70 (0.01)	0.30 (0.02)	3.03
Gumbel	0.85	0.70 (0.01)	0.30 (0.02)	3.02
Joe	0.88	0.70 (0.01)	0.30 (0.02)	3.04

Table 11: $s(\widehat{\theta}) = s(\theta_0)$ reports the fraction of correctly specified structures, $\tau(\widehat{\theta}_k)$ (sd), $k = 1, 2$, refers to the sample average of Kendall's $\tau(\cdot)$ evaluated at the estimates and sd to the sample standard deviation thereof. If the structure is misspecified, $\#\{\widehat{\theta}\}$ gives the number of parameters on average included in the misspecified HAC.

Data and Copula

- daily log returns of Apple (AAPL), Ford (F), Google (GOOG, Microsoft (MSFT) and Toyota Motors (TM)
- timespan = [03.01.2007 - 31.12.2010] ($n = 1008$)
- Gumbel and Clayton generators
- AR(1)–GARCH(1,1)-general error distributed residuals are conditionally distributed with estimated copula

$$\varepsilon \sim C\{F_1(x_1), \dots, F_d(x_d); \theta_t\}$$

where $F_1, \dots, F_d$ are marginal distributions taking to be nonparametrically and θ_t are the copula parameters.

Application II

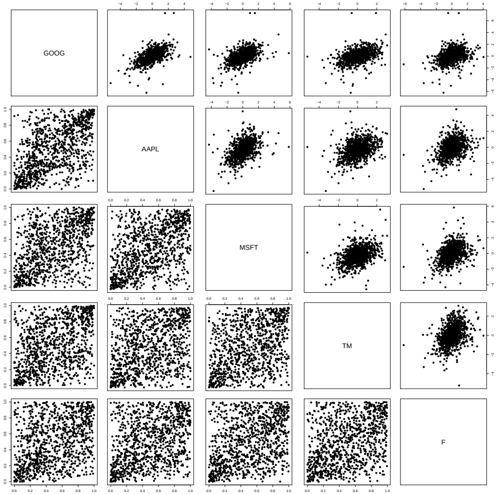


Figure 11: Residuals

Estimation Results

Generator	estimated structure and parameters
Gumbel	$((((GOOG.AAPL)_{1.692(0.078)}.MSFT)_{1.490(0.046)}.TM)_{1.344(0.028)}.F)_{1.280(0.025)}$
Clayton	$(((GOOG.AAPL)_{1.072(0.120)}.MSFT)_{0.762(0.089)}.(TM.F)_{0.554(0.058)})_{0.548(0.051)}$

Table 12: Estimation results for the fit of the HAC with Gumbel and Clayton generators to the residuals. The standard errors of the parameters are given in the parenthesis.

Model with Clayton generators

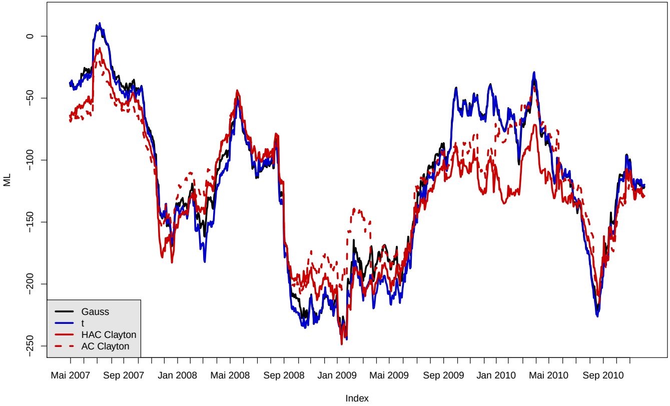


Figure 12: BIC for comparison of the model with Clayton generators with Gaussian and t-copula.

Model with Clayton generators

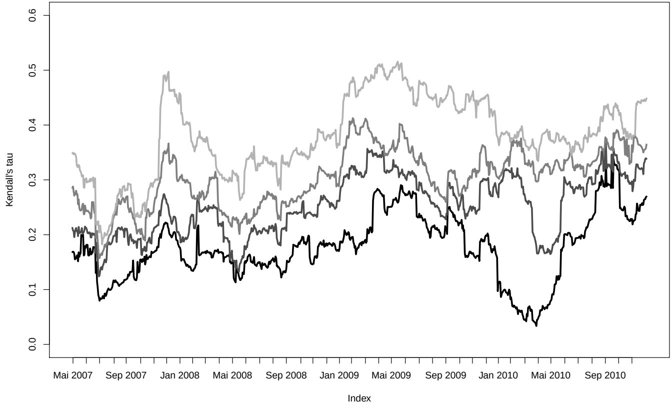


Figure 13: Parameters transformed to Kendall's τ of the model with Clayton generators.

Model with Clayton generators

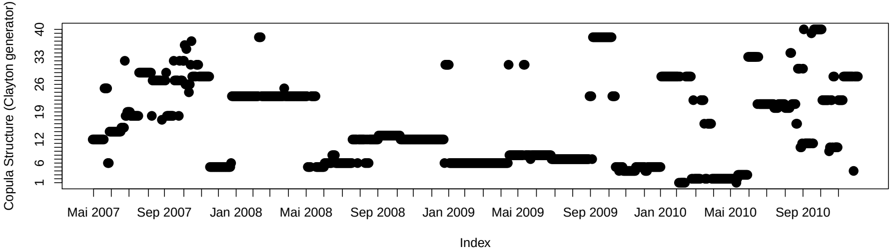


Figure 14: The structure from moving window estimation with window length 100 using Clayton generators.

VaR

The P&L function is $L_{t+1} = \sum_{i=1}^{3} w_i P_{it}(e^{R_{i,t+1}} - 1)$,
The VaR at level α is $VaR(\alpha) = F_L^{-1}(\alpha)$

$$\widehat{\alpha}_{\mathbf{w}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{I}\{L_t < \widehat{VaR}_t(\alpha)\}.$$

The distance between $\widehat{\alpha}$ and α

$$e_{\mathbf{w}} = (\widehat{\alpha}_{\mathbf{w}} - \alpha)/\alpha.$$

The performance of models is measured through

$$A_W = \frac{1}{|W|} \sum_{\mathbf{w} \in W} e_{\mathbf{w}}, \qquad D_W = \left\{ \frac{1}{|W|} \sum_{\mathbf{w} \in W} (e_{\mathbf{w}} - A_W)^2 \right\}^{1/2}.$$

VaR

	10%	5%	1%
HAC (Gumbel)	0.1070 (0.0048)	0.0586 (0.0061)	0.0173 (0.0020)
AC (Gumbel)	0.1094 (0.0070)	0.0639 (0.0067)	0.0186 (0.0028)
HAC (Clayton)	0.1017 (0.0050)	0.0473 (0.0034)	0.0107 (0.0012)
AC (Clayton)	0.1098 (0.0072)	0.0545 (0.0042)	0.0118 (0.0014)
Gauss	0.1015 (0.0053)	0.0542 (0.0034)	0.0139 (0.0016)
t	0.1015 (0.0048)	0.0530 (0.0033)	0.0128 (0.0015)
vineC (mixed)	0.1010 (0.0042)	0.0528 (0.0033)	0.0139 (0.0015)
vineC (Gumbel)	0.1084 (0.0053)	0.0614 (0.0049)	0.0157 (0.0029)
vineC (Clayton)	0.1028 (0.0044)	0.0492 (0.0033)	0.0119 (0.0016)
vineD (mixed)	0.1034 (0.0049)	0.0503 (0.0032)	0.0125 (0.0011)
vineD (Gumbel)	0.1074 (0.0054)	0.0602 (0.0057)	0.0157 (0.0028)
vineD (Clayton)	0.1060 (0.0047)	0.0496 (0.0030)	0.0105 (0.0016)

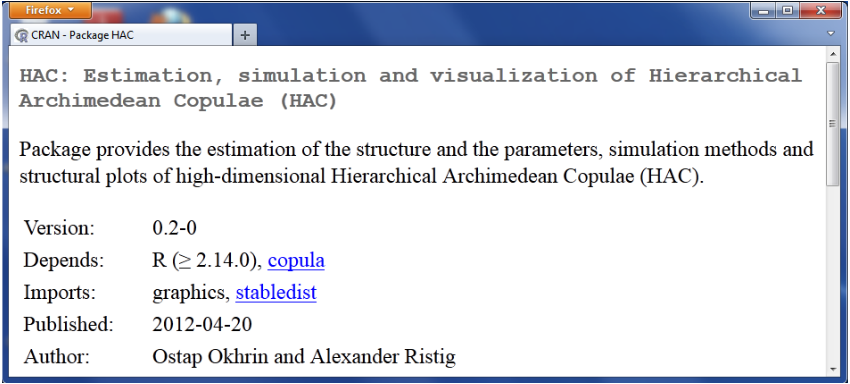
Table 13: The empirical quantiles $\widehat{VaR}(\alpha)$ and the standard deviation in parenthesis.

VaR

	10%	5%	1%
HAC (Gumbel)	0.0702 (0.0479)	0.1726 (0.1226)	0.7330 (0.2034)
AC (Gumbel)	0.0936 (0.0703)	0.2785 (0.1344)	0.8589 (0.2793)
HAC (Clayton)	0.0173 (0.0498)	-0.0548 (0.0685)	0.0702 (0.1173)
AC (Clayton)	0.0975 (0.0716)	0.0899 (0.0846)	0.1823 (0.1371)
Gauss	0.0150 (0.0528)	0.0845 (0.0680)	0.3871 (0.1601)
t	0.0153 (0.0476)	0.0602 (0.0655)	0.2842 (0.1485)
vineC (optimal)	0.0102 (0.0419)	0.0567 (0.0658)	0.3884 (0.1536)
vineC (Gumbel)	0.0837 (0.0533)	0.2280 (0.0983)	0.5709 (0.2877)
vineC (Clayton)	0.0276 (0.0445)	-0.0161 (0.0658)	0.1863 (0.1586)
vineD (optimal)	0.0339 (0.0489)	0.0067 (0.0640)	0.2504 (0.1087)
vineD (Gumbel)	0.0744 (0.0538)	0.2046 (0.1137)	0.5690 (0.2808)
vineD (Clayton)	0.0597 (0.0473)	-0.0078 (0.0591)	0.0462 (0.1581)

Table 14: The average exceedance A_W over all portfolios and its standard deviation D_W.

HAC meets R



HAC: Estimation, simulation and visualization of Hierarchical Archimedean Copulae (HAC)

Package provides the estimation of the structure and the parameters, simulation methods and structural plots of high-dimensional Hierarchical Archimedean Copulae (HAC).

Version:	0.2-0
Depends:	R ($\geq$ 2.14.0), copula
Imports:	graphics, stabledist
Published:	2012-04-20
Author:	Ostap Okhrin and Alexander Ristig

Figure 15: Website for downloading

Portfolio Management

- HAC can be applied to VaR estimation or assessing diversification effects.
- Four stocks: Chevron (CVX), Total (FP), Royal Dutch Sheel (RDSA) and Exxon (XOM).
- 2011-02-02 to 2012-03-19

```
> price = read.table("stocks")
> ret = diff(log(price), 1)
```

- Residuals of ARMA-GARCH models `res`
- Non-ellipticity? Joint extreme events?

```
> pairs(ret, pch = 20)
```

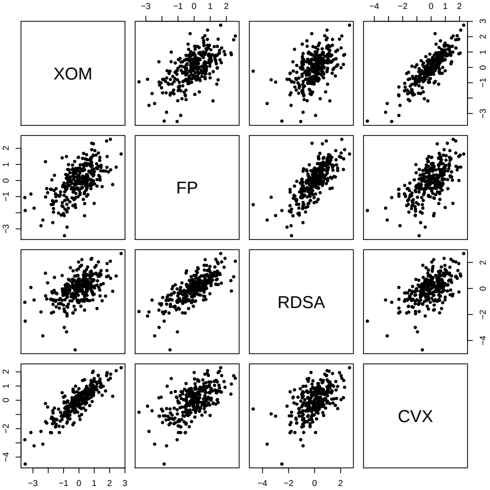


Figure 16: Dependencies of CVX, FP, RDSA and XOM

- Copula estimation based on uniformly distributed margins ures

```
> result = estimate.copula(ures)
> plot(result)
```

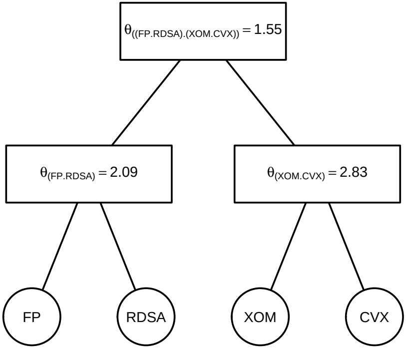


Figure 17: Estimated HAC of the portfolio

Estimation

- 3 computational blocks:
  1. Specification of the margins
  2. Estimation of the parameters and the structure
  3. Optional aggregation of the binary HAC
- Two estimation procedures: QML and Kendall's τ.
- `estimate.copula` returns a `hac` object.

```
> result1 = estimate.copula(res, margins = 'edf')
> plot(result1)
```

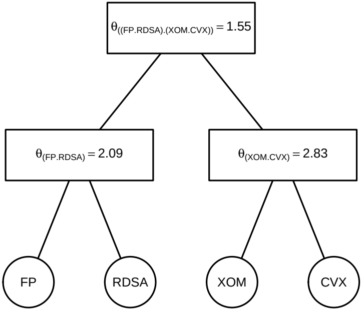


Figure 18: Estimation result

- Note, $C_{\theta_{\mathbf{1}(\mathbf{23})}}\left(C_{\theta_{\mathbf{23}}}\left(u_2, u_3\right), u_1\right) = C_{\theta_{\mathbf{123}}}\left(u_1, u_2, u_3\right)$, if $\left|\theta_{1(23)} - \theta_{23}\right| < \varepsilon, \varepsilon > 0$

- `epsilon = 0.3` leads to a non-binary structure

```
> result2 = estimate.copula(X = res,
+      type = 1, method = ML, epsilon = 0.3,
+      agg.method = "mean", margins = "edf")
> plot(result2)
```

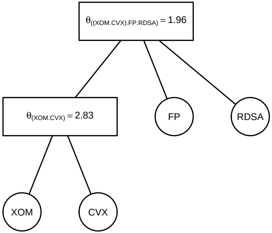


Figure 19: Results of the modified estimation

Objects of the class `hac`

- `hac` and `hac.full` create objects of the class `hac`.
- `hac.full` cannot construct partially nested AC.
- Consider a 5-dimensional fully nested Gumbel HAC:

```
> G1 = hac.full(type = 1,
+      y = c("X1", "X2", "X3", "X4", "X5"),
+      theta = c(1, 1.01, 2, 2.01))
> G1
Class: hac
Generator: Gumbel
((((X5.X4)_{2.01}.X3)_{2}.X2)_{1.01}.X1)_{1}
```

- It is smarter to aggregate the variables X1 and X2 in a first node and the variables X3, X4 and X5 in a second node.

```
> G2 = hac(type = 1,
+       tree = list(list("X3", "X4", "X5", 2.005),
+       "X2", "X1", 1.005))
```

- Substituting of variables for lists leads to arbitrary objects

```
> G3 = hac(type = 1,
+       tree = list(list("Y1", "Y2",
+       list("Z3", "Z4", 3), 2.5),
+       list("Z1", "Z2", 2),
+       list("X1", "X2", 2.4),
+       "X3", "X4", 1.5))
```

Graphics

```
> plot(G3)
```

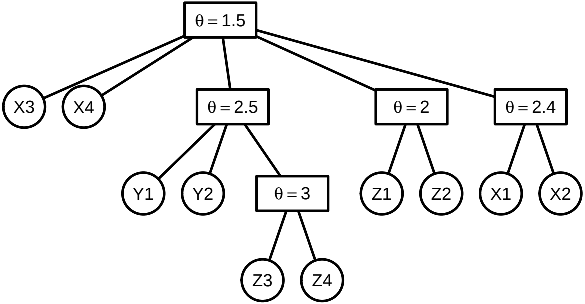


Figure 20: Structure of object G3

```
> plot(G3, digits = 2, theta = TRUE,
+      col = "blue3", fg = "red3",
+      bg = "white", col.t = "blue3")
```

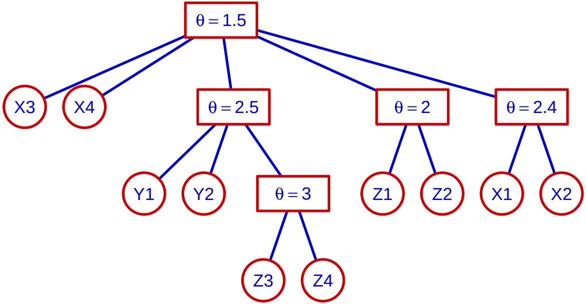


Figure 21: Colored structure of object G3

```
> tree2str(hac = G2, theta = TRUE
+              digits = 3)
[1] ‘‘((X3.X4.X5)_{2.005}.X2.X1)_{1.005}’’
> plot(G2, digits = 3, index = TRUE,
+       theta = FALSE)
```

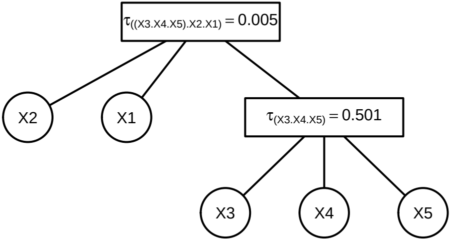


Figure 22: Structure of object G2

Simulation

- Simulation of HAC requires 2 arguments: the number of generated random vectors and a `hac` object.

```
> sample = rHAC(n = 1500, hac = G2)
```

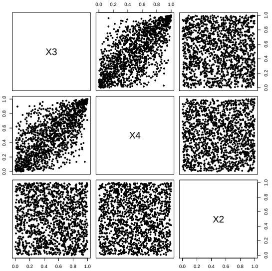


Figure 23: Scatterplot of `sample`

Distribution Functions

- pHAC computes the values of copulae.

```
> cf.values = pHAC(X = sample, hac = G2)
```

- emp.copula.self computes the empirical copula, i.e. $\widehat{C}(u_1, \ldots, u_d) = n^{-1} \sum_{i=1}^{n} \prod_{j=1}^{d} \mathbf{I}\left\{\widehat{F}_j(X_{ij}) \leq u_j\right\}$.

```
> ecf.values = emp.copula.self(x = sample,
+   proc = "M", sort = "none", na.rm = FALSE)
```

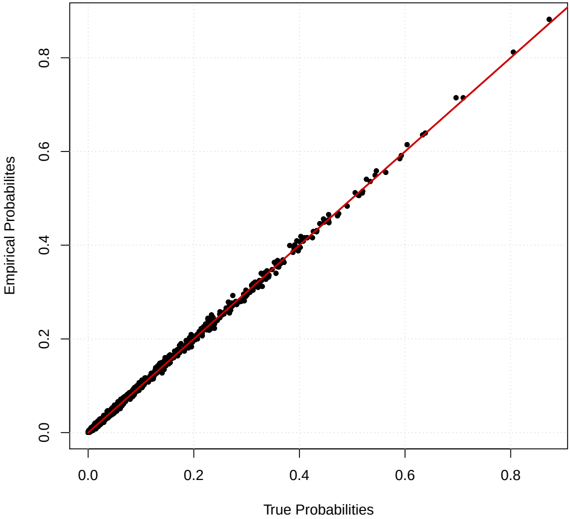


Figure 24: Values of `cf.values` on the x-axis against the values of the `ecf.values`

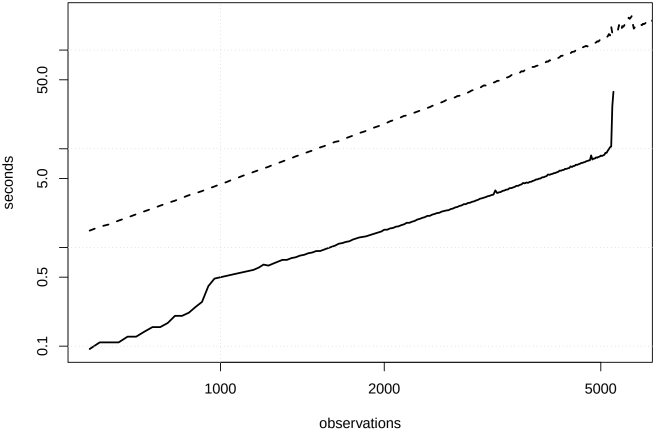


Figure 25: Runtimes of `emp.copula.self` for an increasing sample-size but fixed dimension $d = 5$ plotted on a log-log-scale

Density Functions

- d-dimensional copula density

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

- `dHAC` returns the values of the analytical density.
  - Requires a data matrix and a `hac` object as arguments.
- Construction of Likelihood functions by `to.logLik`.
- Random sampling using conditional inverse method.

Theoretical Properties of HAC

Papers:

Okhrin, O., Okhrin, Y. and Schmid, W., Properties of hierarchical Archimedean copulas. *Statistics and Risk Modeling* 30(1), 2013, pp. 21-53.

Charpentier, A., and Segers, J., Tails of multivariate Archimedean copulas. *Journal of Multivariate Analysis* 100, 2009, pp. 1521-1537

Barbe, Ph., Genest, Ch., Ghoudi, K., and Remillard, B., On Kendalls's process. *Journal of Multivariate Analysis* 58, 1996, pp. 197-229.

Distribution of HAC

Let $V = C\{F_1(X_1), \dots, F_d(X_d)\}$ and let $K(t)$ denote the cdf (K-distribution) of the rv V.

We consider a HAC of the form $C_1\{u_1, C_2(u_2, \dots, u_d)\}$.

Theorem

Let $U_1 \sim U[0,1]$, $V_2 \sim K_2$ and let $P(U_1 \leq x, V_2 \leq y) = C_1\{x, K_2(y)\}$ with $C_1(a,b) = \phi\left\{\phi^{-1}(a) + \phi^{-1}(b)\right\}$ for $a, b \in [0,1]$. Under certain regularity conditions the distribution function K_1 of the random variable $V_1 = C_1(U_1, V_2)$ is given by

$$K_1(t) = t - \int_0^{\phi^{-1}(t)} \phi'\left\{\phi^{-1}(t) + \phi^{-1} \circ K_2 \circ \phi(u) - u\right\} du \quad \text{for} \quad t \in [0,1].$$

Gumbel copula

$$\begin{aligned}\phi_\theta(t) &= \exp(-t^{1/\theta}),\\ \phi_\theta^{-1}(t) &= \{-\log(t)\}^\theta,\\ \phi'_\theta(t) &= -\frac{1}{\theta}\exp(-t^{1/\theta})t^{-1+1/\theta}.\end{aligned}$$

Following Genest and Rivest (1993), K for the simple 2-dim Archimedean copula with generator ϕ is given by $K(t) = t - \phi^{-1}(t)\phi'\{\phi^{-1}(t)\}$. Thus

$$K_2(t,\theta) = t - \frac{t}{\theta}\log(t)$$

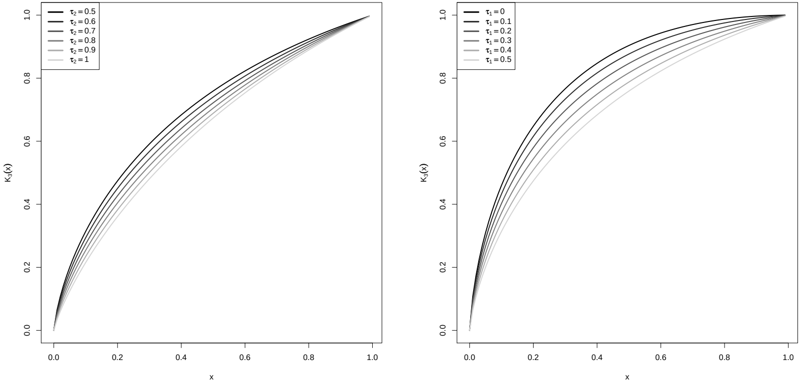


Figure 26: K distribution for three-dimensional HAC with Gumbel generators

Next consider $V_3 = C_3(V_4, V_5)$ with $V_4 = C_4(U_1, \dots, U_\ell)$ and $V_5 = C_5(U_{\ell+1}, \dots, U_d)$.

Theorem

Let $V_4 \sim K_4$ and $V_5 \sim K_5$ and $P(V_4 \le x, V_5 \le y) = C_3\{K_4(x), K_5(y)\}$ with $C_3(a, b) = \phi\left\{\phi^{-1}(a) + \phi^{-1}(b)\right\}$ for $a, b \in [0, 1]$. Under certain regularity conditions the distribution function K_3 of the rv $V_3 = C_3(V_4, V_5)$ is given by

$$\begin{aligned} K_3(t) &= K_4(t) - \\ &- \int_0^{\phi^{-1}(t)} \phi'\left[\phi^{-1} \circ K_5 \circ \phi(u)\right. \\ &\qquad \left. + \phi^{-1} \circ K_4 \circ \phi\{\phi^{-1}(t) - u\}\right] \, d\phi^{-1} \circ K_4 \circ \phi(u) \end{aligned}$$

for $t \in [0, 1]$.

Dependence orderings

C' is more **concordant** than C if
$(\overline{C}(u,v) = u + v - 1 + C(1-u, 1-v))$

$$C \prec_c C' \Leftrightarrow C(\mathbf{x}) \leq C'(\mathbf{x}) \text{ and } \overline{C}(\mathbf{x}) \leq \overline{C'}(\mathbf{x}) \ \forall \mathbf{x} \in [0;1]^d.$$

Theorem

If two feasible hierarchical Archimedean copulae C^1 and C^2 differ only by the generator functions on the top level satisfying the condition $\phi_1^{-1} \circ \phi_2 \in \mathcal{L}^*$, then $C^1 \prec_c C^2$.

Theorem

If two hierarchical Archimedean copulae $C^1 = C^1_{\boldsymbol{\phi}_1}(u_1, \dots, u_d)$ and $C^2 = C^2_{\boldsymbol{\phi}_2}(u_1, \dots, u_d)$ differ only by the generator functions on the level r as $\boldsymbol{\phi}_1 = (\phi_1, \dots, \phi_{r-1}, \phi, \phi_{r+1}, \dots, \phi_p)$ and $\boldsymbol{\phi}_2 = (\phi_1, \dots, \phi_{r-1}, \phi^*, \phi_{r+1}, \dots, \phi_p)$ with $\phi^{-1} \circ \phi^* \in \mathcal{L}^*$, then $C^1 \prec_c C^2$.

Theorem

(Deheuvels (1978)) Let $\{X_{1i}, \dots, X_{di}\}_{i=1,\dots,n}$ *be a sequence of the random vectors with the distribution function* F*, marginal distributions* $F_1, \dots, F_d$ *and copula* C*. Let also* $M_j^{(n)} = \max_{1 \le i \le n} X_{ji}$, $j = 1, \dots, d$ *be the componentwise maxima. Then*

$$\lim_{n\to\infty} \boldsymbol{P}\left\{ \frac{M_1^{(n)} - a_{1n}}{b_{1n}} \le x_1, \dots, \frac{M_d^{(n)} - a_{dn}}{b_{dn}} \le x_d \right\} = F^*(x_1, \dots, x_d), \quad \forall (x_1, \dots, x_d) \in \mathbb{R}^d$$

with $b_{jn} > 0,\ j = 1, \dots, d,\ n \ge 1$ *if and only if*

1. *for all* $j = 1, \dots, d$ *there exist some constants* a_{jn} *and* b_{jn} *and a non-degenerating limit distribution* F_j^* *such that*
$$\lim_{n\to\infty} \boldsymbol{P}\left\{ \frac{M_j^{(n)} - a_{jn}}{b_{jn}} \le x_j \right\} = F_j^*(x_j), \quad \forall x_j \in \mathbb{R};$$
2. *there exists a copula* C^* *such that*
$$C^*(u_1, \dots, u_d) = \lim_{n\to\infty} C^n(u_1^{1/n}, \dots, u_d^{1/n}).$$

Let F_{ds} be the class of d dimensional hierarchical Archimedean copulae with structure s.

Theorem

If $C \in F_{ds_1}$, $C^* \in F_{ds_2}$, $\forall \varphi_\theta \in \mathcal{N}(C)$, $\partial[\varphi_\ell^{-1}(t)/(\varphi_\ell^{-1})'(t)]/\partial t\big|_{t=1}$ *exists and is equal to* $1/\theta$ *and* $C \in MDA(C^*)$ *and* $C \in MDA(C^*)$ *then* $s_1 = s_2$, $\forall \phi_\theta \in \mathcal{N}(C^*)$, $\phi_\theta(x) = \exp\{-x^{1/\theta}\}$.

If the multivariate HAC C (under some minor condition) belongs to the domain of attraction of the HAC C^*. The extreme value HAC C^* has the same structure as the given copula C, with generators on all levels of the hierarchy being Gumbel generators, but with probably other parameters.

Tail dependency

The upper and lower tail indices of two random variables $X_1 \sim F_1$ and $X_2 \sim F_2$ are given by

$$\lambda_U = \lim_{u \to 1^-} P\{X_2 > F_2^{-1}(u) \mid X_1 > F_1^{-1}(u)\} = \lim_{u \to 1^-} \frac{\overline{C}(u,u)}{1-u}$$

$$\lambda_L = \lim_{u \to 0^+} P\{X_2 \leq F_2^{-1}(u) \mid X_1 \leq F_1^{-1}(u)\} = \lim_{u \to 0^+} \frac{C(u,u)}{u}.$$

Theorem (Nelsen (1997))

For a bivariate Archimedean copula with the generator ϕ it holds

$$\lambda_U = 2 - \lim_{u \to 1^-} \frac{1 - \phi\{2\phi^{-1}(u)\}}{1-u} = 2 - \lim_{w \to 0^+} \frac{1 - \phi(2w)}{1 - \phi(w)},$$

$$\lambda_L = \lim_{u \to 0^+} \frac{\phi\{2\phi^{-1}(u)\}}{u} = \lim_{w \to \infty} \frac{\phi(2w)}{\phi(w)}.$$

A function $\phi : (0, \infty) \to (0, \infty)$ is **regularly varying at infinity with tail index** $\lambda \in \mathbb{R}$ (written $RV_\lambda(\infty)$) if $\lim_{w \to \infty} \frac{\phi(tw)}{\phi(w)} = t^\lambda$ for all $t > 0$. $\phi \in RV_{-\infty}(\infty)$ if

$$\lim_{w \to \infty} \frac{\phi(tw)}{\phi(w)} = \begin{cases} \infty & \text{if} \quad t < 1 \\ 1 & \text{if} \quad t = 1 \\ 0 & \text{if} \quad t > 1 \end{cases}.$$

It holds for $\lambda \geq 0$ that if $\phi \in RV_{-\lambda}(\infty)$ then $\phi^{-1} \in RV_{-1/\lambda}(0)$. The function ϕ^{-1} is **regularly varying at zero with the tail index** γ, if $\lim_{w \to 0^+} \frac{\phi^{-1}(1-tw)}{\phi^{-1}(1-w)} = t^\gamma$. For the direct function $\lim_{w \to 0^+} \frac{1-\phi(tw)}{1-\phi(w)} = t^{1/\gamma}$.

$$\lim_{u\to 0^+} \mathrm{P}\{X_i \le F_i^{-1}(u_i u) \text{ for } i \notin \mathcal{S} \subset \mathcal{K} = \{1,\dots,k\} \mid X_j \le F_j^{-1}(u_j u) \text{ for } j \in \mathcal{S}\}$$

$$\lim_{u\to 0^+} \mathrm{P}\{X_i > F_i^{-1}(1-u_i u) \text{ for } i \notin \mathcal{S} \subset \mathcal{K} = \{1,\dots,k\} \mid X_j > F_j^{-1}(1-u_j u) \text{ for } j \in \mathcal{S}\}.$$

The above limits can be established via the limits

$$\lambda_L(u_1,\dots,u_k) = \lim_{u\to 0^+} \frac{1}{u} C(u_1 u,\dots,u_k u) \qquad \text{and}$$

$$\begin{aligned}\lambda_U(u_1,\dots,u_k) &= \lim_{u\to 0^+} \frac{1}{u} \overline{C}(1-u_1 u,\dots,1-u_k u)\\ &= \lim_{u\to 0^+} \sum_{s_1\in\mathcal{K}} (-1)^{|s_1|+1}\{1 - C_{s_1}(1-u_j u, j\in s_1)\}.\end{aligned}$$

Theorem (Lower Tail Dependency)

Assume that the limits $\lim_{u\to 0^+} u^{-1} C_i(uu_{k_{i-1}+1}, \dots, uu_{k_i}) = \lambda_{L,i}(u_{k_{i-1}+1}, .., u_{k_i})$ *exist for all* $0 < u_{k_{i-1}+1}, .., u_{k_i} < 1$, $i = 1, \dots, m$. *Suppose that* $m + k - k_m \geq 2$. *If* ϕ_0^{-1} *is regularly varying at infinity with index* $-\lambda_0 \in [-\infty, 0]$, *then it holds for all* $0 < u_i < 1$, $i = 1, .., m$ *that*

$$\lim_{u\to 0^+} \frac{C(uu_1, \dots, uu_k)}{u} = \begin{cases} \min\{\lambda_{L,1}(u_1, .., u_{k_1}), \dots, \lambda_{L,m}(u_{k_{m-1}+1}, .., u_{k_m}), u_{k_m+1}, \dots, u_k\} \\ \quad \text{if} \quad \lambda_0 = \infty, \\ \left(\sum_{i=1}^m \lambda_{L,i}(u_{k_{i-1}+1}, .., u_{k_i})^{-\lambda_0} + \sum_{j=k_m+1}^k u_j^{-\lambda_0}\right)^{-1/\lambda_0} \\ \quad \text{if} \quad 0 < \lambda_0 < \infty, \\ 0 \text{ if} \quad \lambda_0 = 0. \end{cases}$$

In the following let

$$
\begin{aligned}
C_j^*(u) &= C_j(u_{k_{j-1}+1}u, .., u_{k_j}u) \,|_{u_{k_{j-1}+1}=..=u_{k_j}=1},\\
C^*(u) &= C(u_1 u, .., u_k u) \,|_{u_1=..=u_k=1},\\
\lambda_{L,j}^*(u, u_{k_{j-1}+1}, .., u_{k_j}) &= C_j(u_{k_{j-1}+1}u, \dots, u_{k_j}u)/C_j^*(u).
\end{aligned}
$$

Note that $0 \leq \lambda_{L,j}^*(u, u_{k_{j-1}+1}, .., u_{k_j}) \leq 1$. Moreover, if $\lim_{u\to 0^+} u^{-1}C_j(uu_{k_{j-1}+1}, \dots, uu_{k_j}) = \lambda_{L,j}(u_{k_{j-1}+1}, .., u_{k_j}) > 0$ for all $0 < u_{k_{j-1}+1}, .., u_{k_j} \leq 1$ then

$$
\begin{aligned}
\lambda_{L,j}^*(u_{k_{j-1}+1}, .., u_{k_j}) &= \lim_{u\to 0+} \frac{C_j(u_{k_{j-1}+1}u, .., u_{k_j}u)/u}{C_j^*(u)/u}\\
&= \frac{\lambda_{L,j}(u_{k_{j-1}+1}, .., u_{k_j})}{\lambda_{L,j}(1, .., 1)}
\end{aligned}
$$

Theorem (Lower Tail Dependency 2)

Assume that the limits

$$\lim_{u\to 0^+} \frac{C_i(uu_{k_{i-1}+1},\dots,uu_{k_i})}{C_i^*(u)} = \lambda^*_{L,i}(u_{k_{i-1}+1},..,u_{k_i})$$

exist for all $0 < u_{k_{i-1}+1},..,u_{k_i} \leq 1$, $i = 1,\dots,m$. *Let* $\phi_0^{-1} \in RV_0(0)$ *and let* $\psi(v) = -\phi_0(v)/\phi_0'(v)$ *be regularly varying at infinity with finite tail index* $\varkappa$ *then* $\varkappa \leq 1$ *and it holds for all* $0 < u_i < 1$, $i = 1,..,m$ *that*

$$\lim_{u\to 0^+} \frac{C(uu_1,\dots,uu_k)}{C^*(u)} = \prod_{j=1}^{m}[\lambda^*_{L,j}(u_{k_{j-1}+1},\dots,u_{k_j})]^{(m+k-k_m)^{-\varkappa}} \cdot \prod_{j=k_m+1}^{k} u_j^{(m+k-k_m)^{-\varkappa}}.$$

Theorem (Upper Tail Dependency)

Assume that the limits
$\lim_{u\to 0^+} u^{-1}[1 - C_i(1 - uu_{k_{i-1}+1}, \dots, 1 - uu_{k_i})] = \lambda_{U,i}(u_{k_{i-1}+1}, .., u_{k_i})$
exist for all $0 < u_{k_{i-1}+1}, .., u_{k_i} < 1$, $i = 1, \dots, m$. *Suppose that* $m + k - k_m \geq 2$. *If* $\phi_0^{-1}(1 - w)$ *is regularly varying at zero with index* $-\gamma_0 \in [-\infty, -1]$, *then it holds for all* $0 < u_i < 1$, $i = 1, .., m$ *that*

$$\lim_{u\to 0^+} \frac{1 - C(1 - uu_1, \dots, 1 - uu_k)}{u} = \begin{cases} \min\{\lambda_{U,1}(u_1, .., u_{k_1}), \dots, \lambda_{U,m}(u_{k_{m-1}+1}, .., u_{k_m}), u_{k_m+1}, \dots, u_k\} \\ \qquad \textit{if} \quad \gamma_0 = \infty, \\ \left(\sum_{i=1}^m [\lambda_{U,i}(u_{k_{i-1}+1}, .., u_{k_i})]^{\gamma_0} + \sum_{j=k_m+1}^k u_j^{\gamma_0}\right)^{1/\gamma_0} \\ \qquad \textit{if} \quad 1 \leq \gamma_0 < \infty, \end{cases}$$

Time Varying Copulae

Papers:

Härdle, W.K., Okhrin, O., and Wang, W., HMM in dynamic HAC models, *Econometric Theory* 31(5), 2015, pp 981-1015

Härdle, W.K., Okhrin, O. and Okhrin, Y., Dynamic Structured Copula Models, *Statistics and Risk Modeling*, 30(4), 2013, pp.361-388

Giacomini, E., Härdle, W. K. and Spokoiny, V. (2009). Inhomogeneous dependence modeling with time-varying copulae, *Journal of Business and Economic Statistics* 27(2): 224-234.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models, *Annals of Statistics* 32(2): 577-602.

Local Change Point Detection

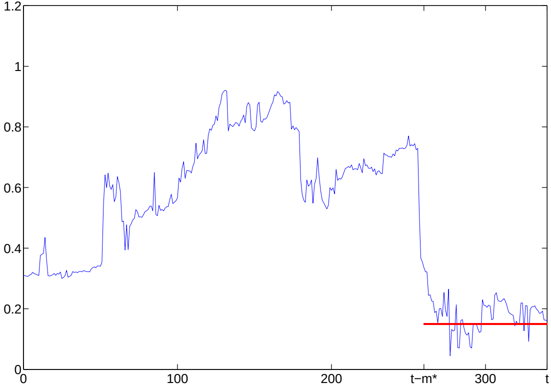


Figure 27: Dependence over time for DaimlerChrysler, Volkswagen, Bayer, BASF, Allianz and Münchener Rückversicherung, 20000101-20041231. Giacomini et. al (2009)

Adaptive Copula Estimation

- adaptively estimate largest interval where homogeneity hypothesis is accepted
- *Local Change Point* detection (LCP): sequentially test $\theta_t,\ s_t$ are constants (i.e. $\theta_t = \theta,\ s_t = s$) within some interval I (local parametric assumption).

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

$$\triangle_I(s, \boldsymbol{\theta}) = \sum_{t \in I} \mathcal{K}\{C(\cdot; s_t, \boldsymbol{\theta}_t), C(\cdot; s, \boldsymbol{\theta})\} \leq \triangle.$$

holds for some $\triangle \geq 0$. m_{k^*} is the ideal scale, $(s,\ \theta)^\top$ is ideally estimated from $I = [t_0 - m_{k^*}, t_0]$ and $\mathcal{K}(\cdot, \cdot)$ is the *Kullback-Leibler* divergence

$$\mathcal{K}\{C(\cdot; s_t, \boldsymbol{\theta}_t), C(\cdot; s, \boldsymbol{\theta})\} = \boldsymbol{E}_{s_t, \boldsymbol{\theta}_t} \log \frac{c(\cdot; s_t, \boldsymbol{\theta}_t)}{c(\cdot; s, \boldsymbol{\theta})}$$

Under the SMB condition on I_{k^*} and assuming that $\max_{k \le k^*} \boldsymbol{E}_{s,\boldsymbol{\theta}} |\mathcal{L}(\widetilde{s}_k, \widetilde{\boldsymbol{\theta}}_k) - \mathcal{L}(s, \boldsymbol{\theta})|^r \le \mathcal{R}_r(s, \boldsymbol{\theta})$, we obtain

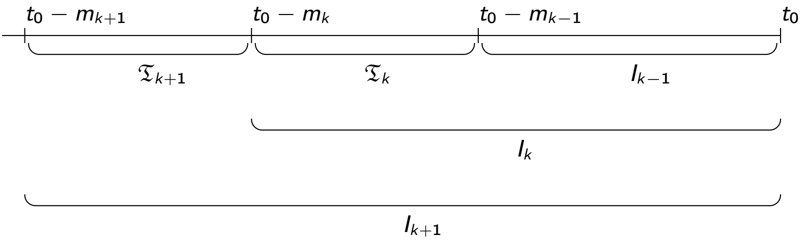
$$\boldsymbol{E}_{s_t, \boldsymbol{\theta}_t} \log \left\{ 1 + \frac{|\mathcal{L}(\widetilde{s}_{\widehat{k}}, \widetilde{\boldsymbol{\theta}}_{\widehat{k}}) - \mathcal{L}(s, \boldsymbol{\theta})|^r}{\mathcal{R}_r(s, \boldsymbol{\theta})} \right\} \le 1 + \triangle,$$

$$\boldsymbol{E}_{s_t, \boldsymbol{\theta}_t} \log \left\{ 1 + \frac{|\mathcal{L}(\widetilde{s}_{\widehat{k}}, \widetilde{\boldsymbol{\theta}}_{\widehat{k}}) - \mathcal{L}(\widehat{s}_{\widehat{k}}, \widehat{\boldsymbol{\theta}}_{\widehat{k}})|^r}{\mathcal{R}_r(s, \boldsymbol{\theta})} \right\} \le \rho + \triangle,$$

where $\widehat{a}_I$ is an adaptive estimator on I and $\widetilde{a}_I$ is any other parametric estimator on I.

Local Change Point Detection

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



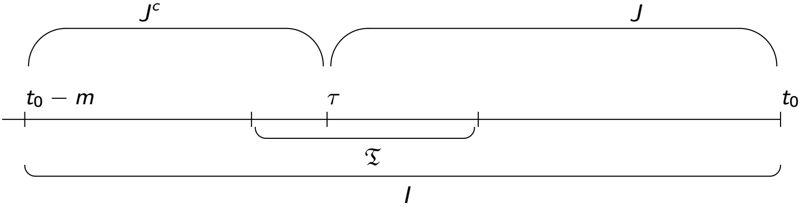
Local Change Point Detection

1. test homogeneity $H_{0,k}$ against the change point alternative in $\mathfrak{T}_k$ using I_{k+1}
2. if no change points in $\mathfrak{T}_k$, accept I_k. Take $\mathfrak{T}_{k+1}$ and repeat previous step until $H_{0,k}$ is rejected or largest possible interval I_K is accepted
3. if $H_{0,k}$ is rejected in $\mathfrak{T}_k$, homogeneity interval is the last accepted $\widehat{I} = I_{k-1}$
4. if largest possible interval I_K is accepted $\widehat{I} = I_K$

Test of homogeneity

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

$$
\begin{aligned}
H_0 \quad &: \quad \forall \tau \in \mathfrak{T},\ \theta_t = \theta,\ s_t = s, \\
& \quad\ \forall t \in J = [\tau, t_0], \forall t \in J^c = I \setminus J \\
H_1 \quad &: \quad \exists \tau \in \mathfrak{T},\ \theta_t = \theta_1,\ s_t = s_1;\ \forall t \in J, \\
& \quad\ \theta_t = \theta_2 \neq \theta_1;\ s_t = s_2 \neq s_1, \forall t \in J^c
\end{aligned}
$$



Test of homogeneity

Likelihood ratio test statistic for fixed change point location:

$$\begin{aligned} T_{I,\tau} &= \max_{\theta_1,\theta_2}\{L_J(\theta_1) + L_{J^c}(\theta_2)\} - \max_{\theta} L_I(\theta) \\ &= ML_J + ML_{J^c} - ML_I \end{aligned}$$

Test statistic for unknown change point location:

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

Reject H_0 if for a critical value ζ_I

$$T_I > \zeta_I$$

Selection of I_k and ζ_k

- set of numbers m_k defining the length of I_k and $\mathfrak{T}_k$ are in the form of a geometric grid
- $m_k = [m_0 c^k]$ for $k = 1, 2, \dots, K$, $m_0 \in \{20,\ 40\}$, $c = 1.25$ and $K = 10$, where $[x]$ means the integer part of x
- $I_k = [t_0 - m_k, t_0]$ and $\mathfrak{T}_k = [t_0 - m_k, t_0 - m_{k-1}]$ for $k = 1, 2, \dots, K$

(Mystery Parameters)

Sequential choice of ζ_k

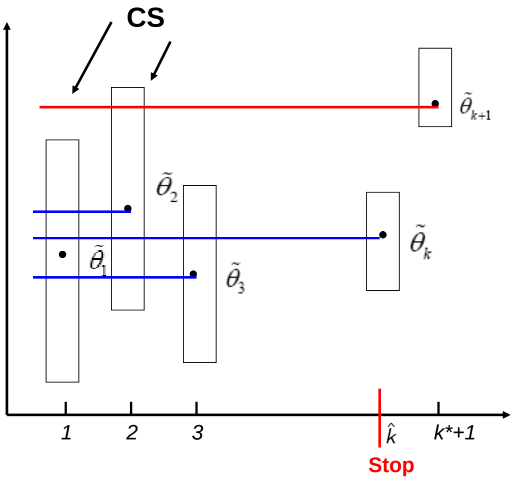
- after k steps are two cases: change point at some step $\ell \leq k$ and no change points.
- let $\mathcal{B}_\ell$ be the event meaning the rejection at step ℓ

$$\mathcal{B}_\ell = \{T_1 \leq \zeta_1, \dots, T_{\ell-1} \leq \zeta_{\ell-1}, T_\ell > \zeta_\ell\},$$

and $(\widehat{s}_k, \widehat{\boldsymbol{\theta}}_k) = (\widetilde{s}_{\ell-1}, \widetilde{\boldsymbol{\theta}}_{\ell-1})$ on $\mathcal{B}_\ell$ for $\ell = 1, \dots, k$.
- we find sequentially such a minimal value of ζ_ℓ that ensures following inequality

$$\max_{k=l,\dots,K} \boldsymbol{E}_{s^*,\theta^*} |\mathcal{L}(\widetilde{s}_k, \widetilde{\boldsymbol{\theta}}_k) - \mathcal{L}(\widetilde{s}_{\ell-1}, \widetilde{\boldsymbol{\theta}}_{\ell-1})|^r \mathbf{I}(\mathcal{B}_\ell) \leq \rho \mathcal{R}_r(s^*, \boldsymbol{\theta}^*) \frac{k}{K-1}$$

Illustration



Sequential choice of ζ_k

1. pairs of Kendall's τ: $\forall\{\tau_1, \tau_2\} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}^2$, $\tau_1 \geq \tau_2$
2. simul. from $C_{\theta_i,\theta_j}(u_1, u_2, u_3) = C\{C(u_1, u_2; \theta_1), u_3; \theta_2\}$, $\theta = \theta(\tau)$
3. run sequential algorithm for each sample

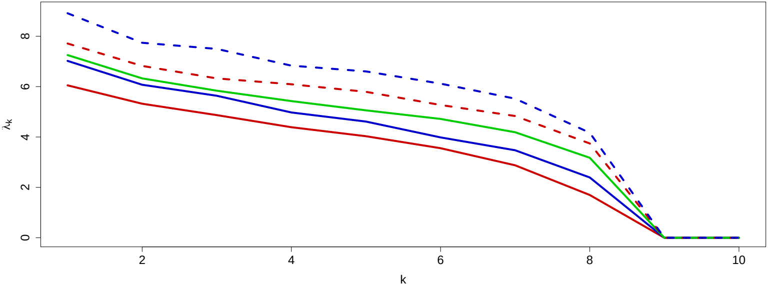


Figure 28: ζ_k of the 3-dimensional HAC as a function of k with the fixed $m_0 = 40$, $\rho = 0.5$, $r = 0.5$, $\tau_1 = 0.1$ and for different τ_2. $\tau_2 = 0.1$ (solid), $\tau_2 = 0.3$ (solid), $\tau_2 = 0.5$ (solid), $\tau_2 = 0.7$ (dashed), $\tau_2 = 0.9$ (dashed).

Simulation: Change in θ_1, I

$$C_t(u_1, u_2, u_3; s, \boldsymbol{\theta}) = \begin{cases} C\{u_1, C(u_2, u_3; \theta_1 = 3.33); \theta_2 = 1.43\} & \text{for } 1 \leq t \leq 200 \\ C\{u_1, C(u_2, u_3; \theta_1 = 2.00); \theta_2 = 1.43\} & \text{for } 200 < t \leq 400 \end{cases}$$

1. $N = 400$ and 100 runs
2. LCP based on the same critical values

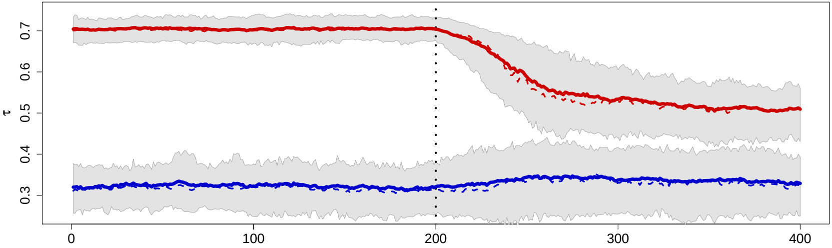


Figure 29: θ_1 and θ_2 on the intervals of homogeneity (median - dashed line, mean - solid line).

Simulation: Change in θ_1, II

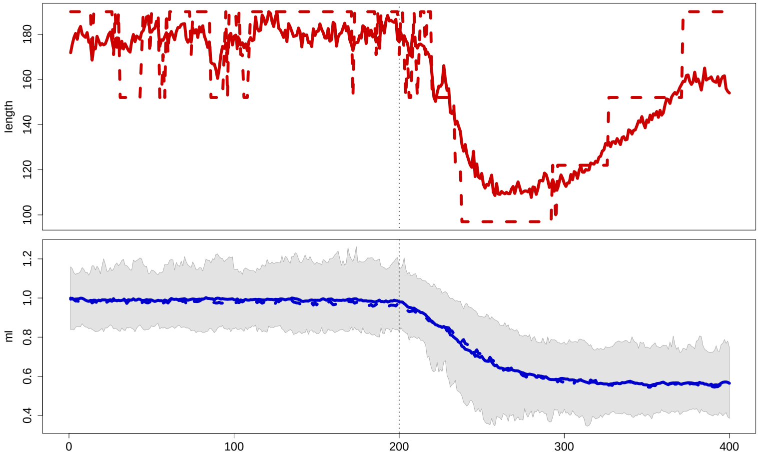


Figure 30: Intervals of homogeneity and ML on these intervals (median - dashed line, mean - solid line)

Simulation: Change in θ_2, I

$$C_t(u_1, u_2, u_3; s, \boldsymbol{\theta}) = \begin{cases} C\{u_1, C(u_2, u_3; \theta_1 = 3.33); \theta_2 = 1.43\} & \text{for } 1 \leq t \leq 200 \\ C\{u_1, C(u_2, u_3; \theta_1 = 3.33); \theta_2 = 2.00\} & \text{for } 200 < t \leq 400 \end{cases}$$

1. $N = 400$ and 100 runs
2. LCP based on the same critical values

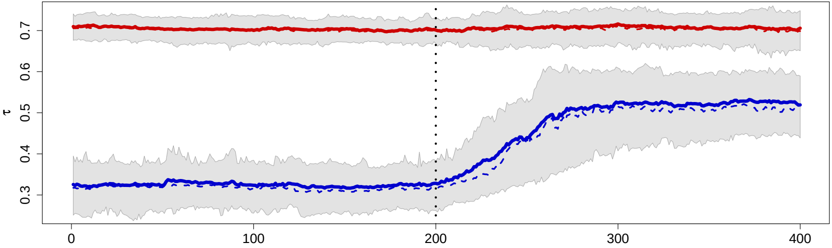


Figure 31: θ_1 and θ_2 on the intervals of homogeneity (median - dashed line, mean - solid line).

Simulation: Change in θ_2, II

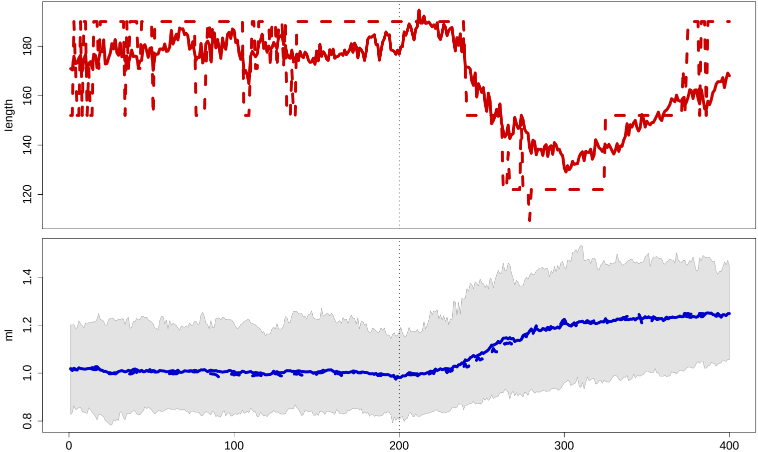


Figure 32: Intervals of homogeneity and ML on these intervals (median - dashed line, mean - solid line)

Simulation: Change in the Structure, I

$$C_t(u_1, u_2, u_3; s, \boldsymbol{\theta}) = \begin{cases} C\{u_1, C(u_2, u_3; \theta_1 = 3.33); \theta_2 = 1.43\} & \text{for } 1 \le t \le 200 \\ C\{C(u_1, u_2; \theta_1 = 3.33), u_3; \theta_2 = 1.43\} & \text{for } 200 < t \le 400 \end{cases}$$

1. $N = 400$ and 100 runs
2. LCP based on the same critical values

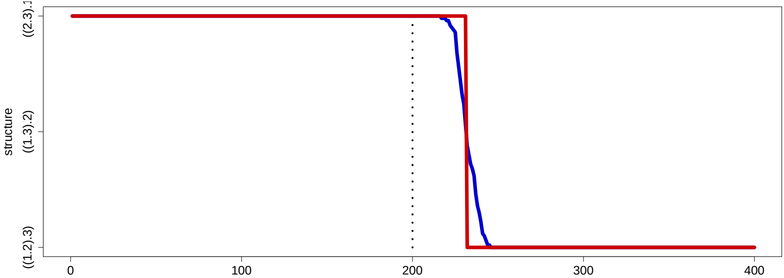


Figure 33: The structure of the est. HAC on the intervals of homogeneity (median - dashed line, mean - solid line)

Simulation: Change in the Structure, II

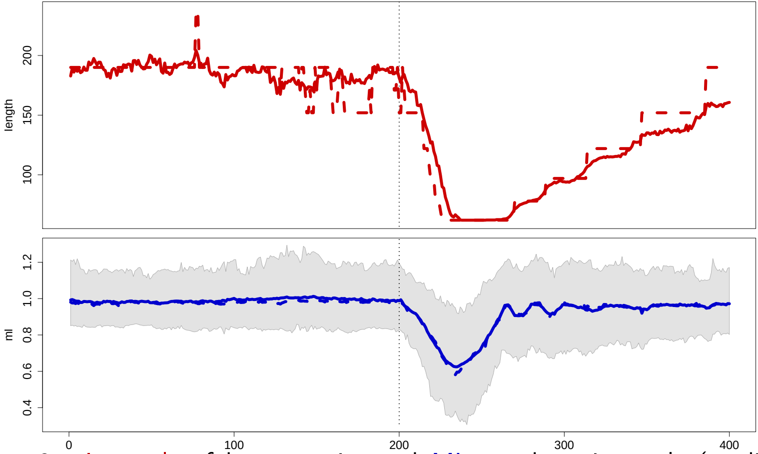


Figure 34: Intervals of homogeneity and ML on these intervals (median - dashed line, mean - solid line)

Application III

- daily values for the exchange rates
  JPN/USD, GBP/USD and EUR/USD
- timespan = [4.1.1999; 14.8.2009] ($n = 2771$)
- Gumbel and Clayton generators generators
- a univariate GARCH(1,1) process on log-returns

$$
\begin{aligned}
X_{j,t} &= \mu_{j,t} + \sigma_{j,t}\varepsilon_{j,t} \text{ with } \sigma_{j,t}^2 = \omega_j + \alpha_j \sigma_{j,t-1}^2 + \beta_j (X_{j,t-1} - \mu_{j,t-1})^2 \\
\varepsilon_t &\sim C\{F_1(x_1), \dots, F_d(x_d); \theta_t\}
\end{aligned}
$$

HAC for whole sample

Generator	Structure	ML
Clayton	$((\text{JPY.USD})_{0.808(0.042)}.\text{GBP})_{0.401(0.025)}$	617.268
Gumbel	$((\text{JPY.USD})_{1.521(0.025)}.\text{GBP})_{1.303(0.016)}$	736.341

Table 15: Estimation results for HAC with Clayton and Gumbel generators for indices and exchange rates using full samples. The standard deviations of the parameters are given in the parenthesis.

LCP for HAC to real Data

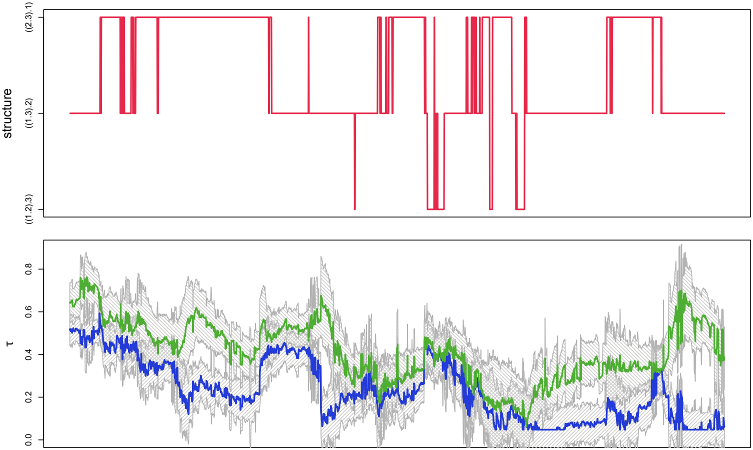


Figure 35: Structure, τ_1 and τ_2 of the HAC on the intervals of homogeneity

LCP for HAC to real Data

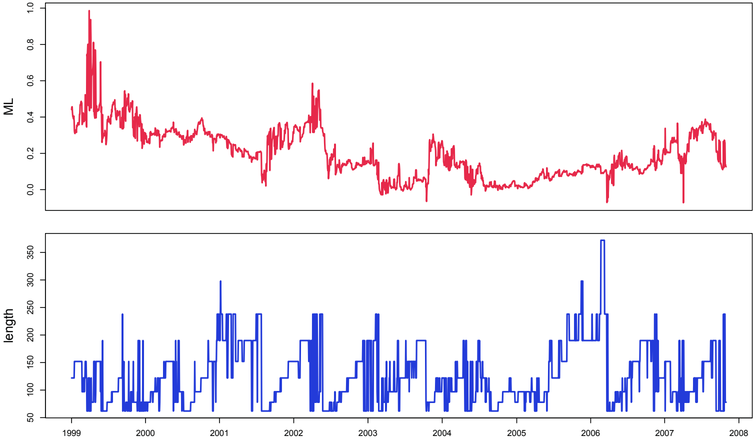


Figure 36: Intervals of homogeneity and ML on these intervals

VaR

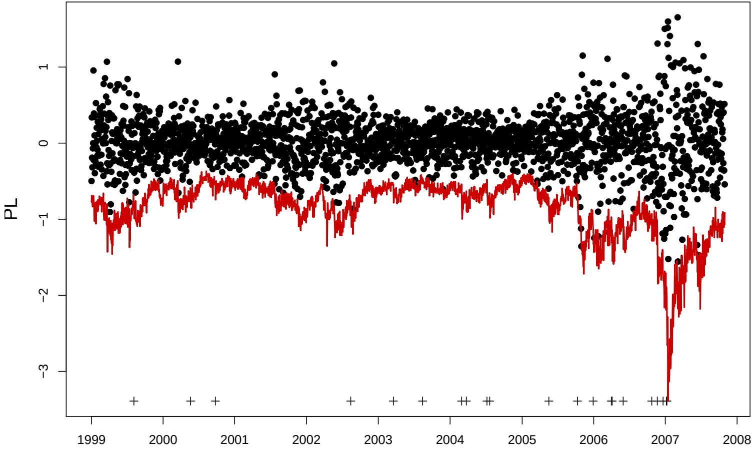


Figure 37: Profit and Loss function

VaR

	Clayton			Gumbel			DCC		
α	0.0100	0.0500	0.1000	0.0100	0.0500	0.1000	0.0100	0.0500	0.1000
$\widehat{\alpha}_{w^*}$	0.0100	0.0487	0.0951	0.0091	0.0474	0.0977	0.0156	0.0413	0.0817
$\widehat{\alpha}_{w_1}$	0.0083	0.0460	0.0912	0.0087	0.0447	0.0925	0.0152	0.0408	0.0812
$\widehat{\alpha}_{w_2}$	0.0100	0.0487	0.0951	0.0096	0.0487	0.0977	0.0156	0.0413	0.0812
$\widehat{\alpha}_{w_3}$	0.0100	0.0487	0.0951	0.0091	0.0482	0.0973	0.0156	0.0413	0.0812
$\widehat{\alpha}_{w_4}$	0.0100	0.0487	0.0951	0.0091	0.0469	0.0973	0.0156	0.0417	0.0817
$\widehat{\alpha}_{w_5}$	0.0100	0.0487	0.0947	0.0091	0.0478	0.0973	0.0156	0.0417	0.0817
A_W	-0.0217	-0.0328	-0.0557	-0.0895	-0.0526	-0.0341	0.5482	-0.1652	-0.1852
D_W	0.0649	0.0186	0.0125	0.0632	0.0406	0.0272	0.0335	0.0091	0.0042

Table 16: Exceedance ratios for portfolios of exchange rates with w^*, $w_i, i = 1, \dots, 5$, the average exceedance A_W over all portfolios and its standard deviation D_W.

Data and Copula

- daily returns values for Dow Jones (DJ), DAX and NIKKEI
- timespan = [01.01.1985; 23.12.2010] ($n = 6778$)
- Gumbel and Clayton generators
- APARCH(1,1) model with the residuals following the skewed-t distribution

$$
\begin{aligned}
X_{j,t} &= \mu_j + \sigma_{j,t}\varepsilon_{j,t} \\
\text{with} \quad & \sigma_{j,t}^{\delta_j} = \omega_j + \alpha_j(|X_{j,t-1} - \mu_j| - \gamma(X_{j,t-1} - \mu_j))^{\delta_j} + \beta_j \sigma_{j,}^{\delta}
\end{aligned}
$$

where $\varepsilon_{j,t} \sim t_{\text{skewed}}(\varkappa; \nu)$, $j = 1, \ldots, 3$. The parameters $\varkappa$ and ν stand for the skew and shape (degrees of freedom) of the distribution.

HAC for whole sample

Generator	Structure	ML
Clayton	$((DAX.DJ)_{0.459(0.021)}.NIKKEI)_{0.155(0.012)}$	545.399
Gumbel	$((DAX.DJ)_{1.272(0.012)}.NIKKEI)_{1.103(0.007)}$	542.736

Table 17: Estimation results for HAC with Clayton and Gumbel generators for indices and exchange rates using full samples. The standard deviations of the parameters are given in the parenthesis.

LCP for HAC to real Data

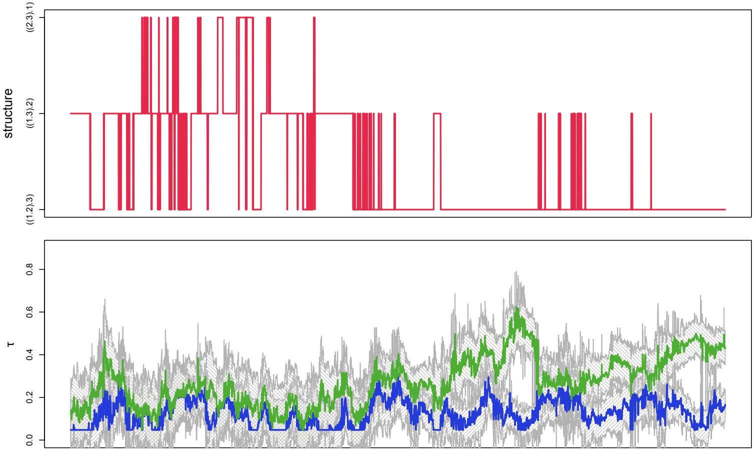


Figure 38: Structure, τ_1 and τ_2 of the HAC on the intervals of homogeneity

LCP for HAC to real Data

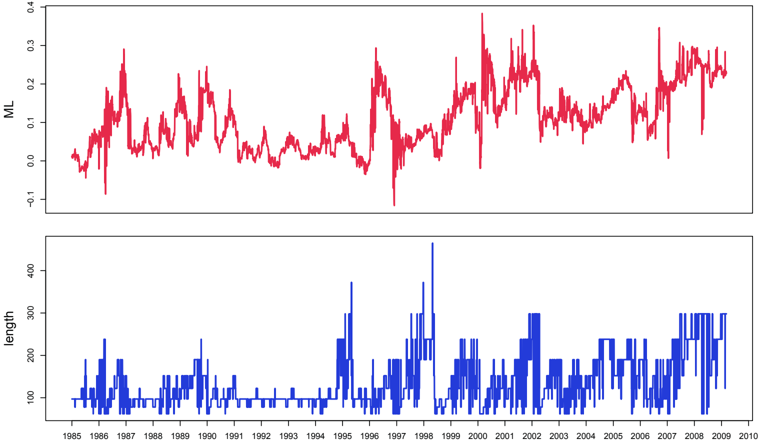


Figure 39: Intervals of homogeneity and ML on these intervals

VaR

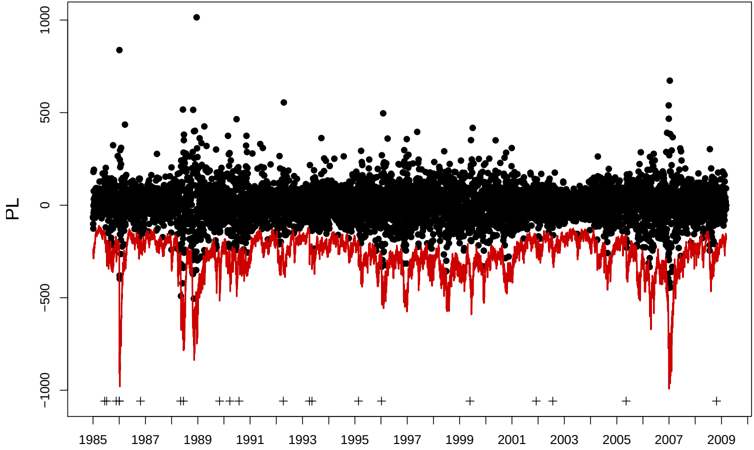


Figure 40: Profit and Loss function

VaR

	Clayton			Gumbel			DCC		
α	0.0100	0.0500	0.1000	0.0100	0.0500	0.1000	0.0100	0.0500	0.1000
$\widehat{\alpha}_{w^*}$	0.0054	0.0390	0.0935	0.0033	0.0249	0.0683	0.0155	0.0460	0.0830
$\widehat{\alpha}_{w_1}$	0.0055	0.0372	0.0916	0.0040	0.0239	0.0705	0.0162	0.0453	0.0864
$\widehat{\alpha}_{w_2}$	0.0073	0.0458	0.0994	0.0044	0.0303	0.0788	0.0152	0.0471	0.0830
$\widehat{\alpha}_{w_3}$	0.0055	0.0412	0.0940	0.0030	0.0254	0.0718	0.0160	0.0480	0.0808
$\widehat{\alpha}_{w_4}$	0.0052	0.0399	0.0943	0.0035	0.0225	0.0681	0.0157	0.0431	0.0818
$\widehat{\alpha}_{w_5}$	0.0062	0.0422	0.0976	0.0043	0.0290	0.0765	0.0160	0.0507	0.0887
A_W	-0.3902	-0.1781	-0.0497	-0.6187	-0.4496	-0.2686	0.5979	-0.0687	-0.1739
D_W	0.0930	0.0508	0.0286	0.0953	0.0932	0.0638	0.0959	0.0829	0.0609

Table 18: Exceedance ratios for portfolios of indices with w^*, $w_i, i = 1, \dots, 5$, the average exceedance A_W over all portfolios and its standard deviation D_W.

Hidden Markov Models

Stochastic process driven by an underlying Markov process, Bickel, Ritov and Ryden (1998), Fuh (2003):

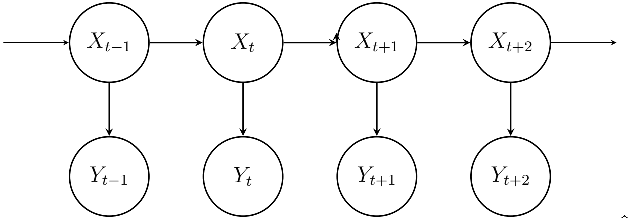


Figure 41: Graphical representation of the dependence structure of HMM

Hidden Markov Models

Observe i.i.d. $Y = (Y_1, Y_2, \dots, Y_T)^\top \in \mathbb{R}^d$, where $\{Y_t\}_{t\geq 0}$ is connected with an underlying Markov Chain $\{X_t\}_{t\geq 0}$, $t = 1, \dots, T$, X_t takes value on $1, \dots, M$.

$$\begin{aligned} \boldsymbol{P}(X_t|X_{1:(t-1)}, Y_{1:(t-1)}) &= \boldsymbol{P}(X_t|X_{t-1}) \qquad (3)\\ \boldsymbol{P}(Y_t|Y_{1:(t-1)}, X_{(1:t)}) &= \boldsymbol{P}(Y_t|X_t), \qquad (4) \end{aligned}$$

$\{X_t, Y_t\}$ follows an HMM.

Likelihood

- Define $p_{ij} = \boldsymbol{P}(X_t = j | X_{t-1} = i)$ the transition probability
- π_i the initial probability
- $f_j\{b; s^{(j)}, \boldsymbol{\theta}^{(j)}\}$ (abbreviated as $f_j(.)$) the HAC-based density
- $\mathfrak{g} \stackrel{\text{def}}{=} (\{\mathbf{s}, \theta\}, p_{ij})$ $(i = 1, \ldots, M,\ j = 1, \ldots, M)$.

Likelihood

For given d dimensional time series $y_1, \cdots, y_T \in \mathbb{R}^d$ $(y_t = (y_{1t}, y_{2t}, y_{3t}, \ldots, y_{dt})^\top)$ π_{x_t} as the π_i for $x_0 = i, i = 1, \ldots, M$, and $p_{x_{t-1}x_t} = p_{ji}$ for $x_{t-1} = j$ and $x_t = i$.
The likelihood of Y and X can be expressed as:

$$p_T(y_1, \cdots, y_T; x_1, \ldots, x_T) = \pi_{x_0} \prod_{t=1}^{T} p_{x_{t-1}x_t} f_{x_t}(y_t; \boldsymbol{\theta}^{(x_t)}, s^{(x_t)})$$

EM algorithm

Following Dempster, Laird and Rubin (1997)

(a) E-step : compute $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}^{(\nu)})$,

(b) M-step : choose the update parameters

$$\mathfrak{g}^{(\nu+1)} = \arg\max_{\mathfrak{g}} \mathcal{Q}(\mathfrak{g}; \mathfrak{g}^{(\nu)}),$$

where $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}^{(\nu)}) \stackrel{\text{def}}{=} \boldsymbol{E}_{\mathfrak{g}^{(\nu)}}\{\log L(Y, X, \theta, \mathsf{s}) | Y\}$.

EM algorithm – E-step

$$
\begin{aligned}
\mathcal{Q}(\mathfrak{g};\mathfrak{g}') &= \sum_{i=1}^{M} P_{(\mathfrak{g}')}(X_0 = i|Y) \log\{\pi_i f_i(Y_0)\} \\
&+ \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} P_{(\mathfrak{g}')}(X_{t-1} = i, X_t = j|Y) \log\{p_{ij}\} \\
&+ \sum_{t=1}^{T}\sum_{i=1}^{M} P_{(\mathfrak{g}')}(X_t = i|Y) \log f_i(Y_t)
\end{aligned}
$$

Likelihood with constraints:

$$
\mathfrak{L}(\mathfrak{g},\lambda;\mathfrak{g}') = \mathcal{Q}(\mathfrak{g};\mathfrak{g}') + \sum_{i=1}^{M} \lambda_i (1 - \sum_{j=1}^{M} p_{ij}). \tag{5}
$$

EM algorithm – M-step

$$
\begin{aligned}
\{\widehat{\theta}^{(i)}_{(\nu)}, \widehat{s}^{(i)}_{(\nu)}\} &= \arg\max_{s^{(i)},\theta^{(i)}} \sum_{t=1}^{T} \boldsymbol{P}(X_t = i|Y)\mathfrak{L}(\mathfrak{g}_i, \lambda; \mathfrak{g}') \\
\{\widehat{\theta}_{ij}\} &= \arg\operatorname{zero}_{\theta_{ij}} \sum_{t=1}^{T} \boldsymbol{P}(X_t = i|Y)\partial \log f_i(y_t)/\partial\theta_{ij}, \\
& \quad i \in 1, \cdots, M
\end{aligned}
$$

Theoretical Results

Theorem

Under certain conditions, we can consistently find the corresponding structure:

$$\lim_{n\to\infty} \boldsymbol{P}(\widehat{s}^{(i)} = s^{*(i)}) = 1, \forall i, 1, \ldots, M \tag{6}$$

Theorem

Given the selected structures $\{\widehat{s}^{(i)}\}$*s, the estimator* $\widehat{\theta}^{(i)}$ *satisfies:*

$$\lim_{n\to\infty} \boldsymbol{P}(\widehat{\boldsymbol{\theta}}^{(i)} = \boldsymbol{\theta}^{*(i)}) = 1, \forall i. \tag{7}$$

Simulations

- Aim is to check if estimation performance of HMM HAC is affected by
  1. adopting nonparametric margins.
  2. assuming parametric margins.
  3. introducing a GARCH dependency in the marginal time series.
- Simulation I: Simulation of 3 dimensional time series model
- Simulation II: Simulation of 5 dimensional time series model
- Simulation III: Forcasting comparison of the DCC method with the HMM HAC approach

Simulation, I

- Simulation according to a GARCH$(1,1)$ model for studying effect of deGARCH,

$$Y_{tj} = \mu_{tj} + \sigma_{tj}\varepsilon_{tj}$$

  with

$$\sigma_{tj}^2 = \omega_j + \alpha_j \sigma_{t-1j}^2 + \beta_j (Y_{t-1j} - \mu_{t-1j})^2,$$

  with parameters $\omega_j = 10^{-6}, \alpha_j = 0.8, \beta_j = 0.1$, with standard normal residuals $\varepsilon_{t1}, \varepsilon_{t2}, \varepsilon_{t3} \sim \boldsymbol{N}(0,1)$.
- Dependence structure is modeled by HAC with Gumbel generators.

Simulation, I

Transition matrix: $\begin{pmatrix} 0.982 & 0.010 & 0.008 \\ 0.008 & 0.984 & 0.008 \\ 0.003 & 0.002 & 0.995 \end{pmatrix}$

$n = 2000$, $d = 3$, $M = 3$, fixed marginal distributions: $Y_{t1}, Y_{t2}, Y_{t3} \sim \boldsymbol{N}(0,1)$

$$C\{u_1, C(u_2, u_3; \theta_1^{(1)} = 1.3); \theta_2^{(1)} = 1.05\} \text{ for } i = 1$$
$$C\{u_2, C(u_3, u_1; \theta_1^{(2)} = 2.0); \theta_2^{(2)} = 1.35\} \text{ for } i = 2$$
$$C\{u_3, C(u_1, u_2; \theta_1^{(3)} = 4.5); \theta_2^{(3)} = 2.85\} \text{ for } i = 3$$

with i indicating the three states.

Simulation, I

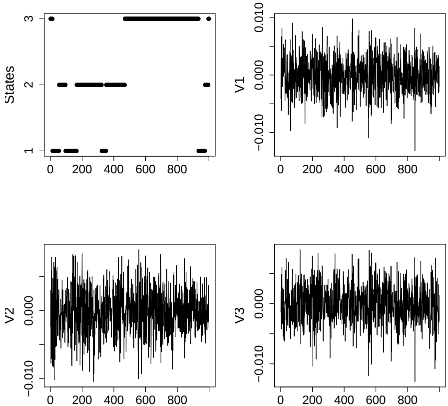


Figure 42: The underlying sequence x_t (upper left panel), marginal plots of $(y_{t1}, y_{t2}, y_{t3})(t = 0, \dots, 1000)$.

Simulation, I

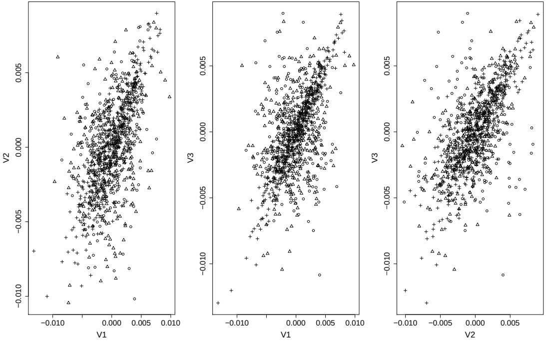


Figure 43: Snapshots of pairwise scatter plots of dependency structures ($t = 0, \ldots, 1000$), the (y_{t1}) vs. (y_{t2}) (left), the (y_{t1}) vs. (y_{t3}) (middle), and the (y_{t2}) vs. (y_{t3})(right). Circles, triangles, and crosses correspond to the observations from states $i = 1, 2, 3$ respectively.

With GARCH effects in the margins, nonparametrically estimated margins

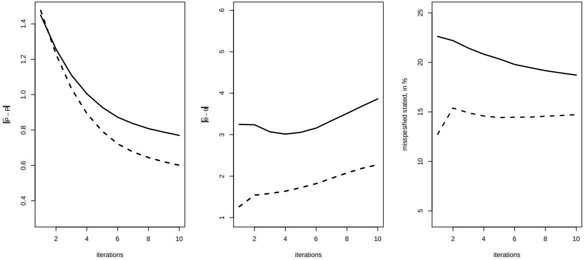


Figure 44: The averaged estimation errors for the transition matrix (left panel), parameters (middle panel), convergence of states (right panel). Estimation starts from near true value (dashed); starts from values obtained by rolling window (solid). x-axis represents represents iterations. Number of repetitions is 1000.

Without GARCH effects in the margins, nonparametrically estimated margins

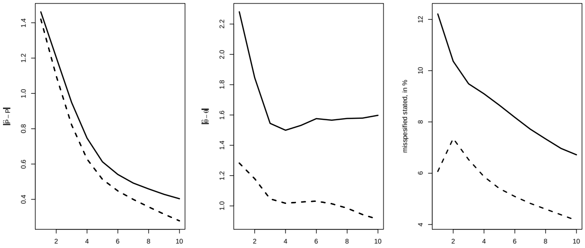


Figure 45: The averaged estimation errors for the transition matrix (left panel), parameters (middle panel), convergence of states (right panel). Estimation starts from near true value (dashed); starts from values obtained by rolling window (solid). x-axis represents represents iterations. Number of repetitions is 1000.

Without GARCH effects in the margins, parametric margins

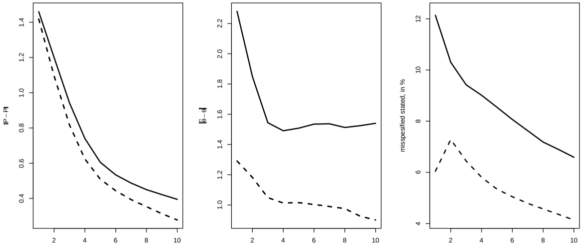


Figure 46: The averaged estimation errors for the transition matrix (left panel), parameters (middle panel), convergence of states (right panel). Estimation starts from near true value (dashed); starts from values obtained by rolling window (solid). x-axis represents represents iterations. Number of repetitions is 1000.

Simulation, I

				True DGP is ID		True DGP is GARCH(1.1)	
			True	Rol. Win.	True Str.	Rol. Win.	True Str.
Nonparametric Margins	C_1	$\theta_1^{(1)}$	1.05	1.030 (0.046, 0.003)	1.057 (0.068, 0.005)	1.100 (0.888, 0.791)	1.138 (0.080, 0.014)
		$\theta_2^{(1)}$	1.30	1.313 (0.156, 0.025)	1.308 (0.083, 0.007)	1.407 (0.888, 0.800)	1.246 (0.080, 0.009)
	C_2	$\theta_1^{(2)}$	1.35	1.366 (0.121, 0.015)	1.346 (0.182, 0.033)	1.403 (1.473, 2.173)	1.436 (2.608, 6.089)
		$\theta_2^{(2)}$	2.00	2.556 (1.052, 1.416)	3.212 (1.991, 5.433)	3.288 (1.473, 3.829)	5.106 (2.608, 16.449)
	C_3	$\theta_1^{(3)}$	2.85	2.854 (0.073, 0.005)	2.854 (0.073, 0.005)	2.772 (0.936, 0.882)	2.790 (0.941, 0.889)
		$\theta_2^{(3)}$	4.50	4.497 (0.133, 0.018)	4.496 (0.130, 0.017)	4.570 (0.936, 0.881)	4.606 (0.941, 0.897)
	rat. of correct states			0.958 (0.029)	0.933 (0.056)	0.853 (0.054)	0.813 (0.061)
	$\sum_{i,j=1}^{d} \lvert \widehat{p}_{ij} - p_{ij} \rvert$			0.278 (0.230)	0.404 (0.307)	0.601 (0.217)	0.770 (0.242)
	rat. of correct structures			0.949	0.918	0.853	0.757

Table 19: Simulation results for different DGPs, sample size $T = 2000$, 1000 repetitions, standard deviations and MSEs are provided in brackets.

Simulation, I

			True DGP is ID		True DGP is GARCH(1.1)	
		True	Rol. Win.	True Str.	Rol. Win.	True Str.
Parametric Margins C_1	$\theta_1^{(1)}$	1.05	1.030 (0.041, 0.002)	1.056 (0.066, 0.004)	1.205(1.261,1.614)	1.107(0.079,0.009)
	$\theta_2^{(1)}$	1.30	1.310 (0.154, 0.024)	1.306 (0.087, 0.008)	1.843 (1.261,1.885)	1.145 (0.079,0.030)
C_2	$\theta_1^{(2)}$	1.35	1.365 (0.130, 0.017)	1.344 (0.173, 0.030)	1.577(1.381, 1.959)	1.838(1.612,2.837)
	$\theta_2^{(2)}$	2.00	2.544 (0.962, 1.221)	3.157 (1.906, 4.971)	3.15(1.381,3.230)	3.480 (2.270, 7.343)
C_3	$\theta_1^{(3)}$	2.85	2.855 (0.074, 0.006)	2.854 (0.074, 0.005)	3.879(1.453, 3.170)	3.906(1.523, 3.435)
	$\theta_2^{(3)}$	4.50	4.513 (0.133, 0.018)	4.513 (0.132, 0.018)	6.39 (1.453,5.683)	6.592(1.523, 6.696)
rat. of correct states			0.959 (0.029)	0.934 (0.056)	0.732 (0.08)	0.747 (0.053)
$\sum_{i,j=1}^{d} \lvert \widehat{p}_{ij} - p_{ij} \rvert$			0.278 (0.232)	0.395 (0.297)	0.761 (0.179)	0.76 (0.156)
rat. of correct structures			0.955	0.921	0.358	0.323

Table 20: Simulation results for different DGPs, sample size $T = 2000$, 1000 repetitions, standard deviations and MSEs are provided in brackets.

Simulation, I

				True DGP is ID		True DGP is GARCH(1.1)	
			True	Rol. Win.	True Str.	Rol. Win.	True Str.
deGARCHing	C_1	$\theta_1^{(1)}$	1.05	1.030 (0.045, 0.002)	1.056 (0.067, 0.005)	1.030 (0.736, 0.542)	1.067 (0.141, 0.020)
		$\theta_2^{(1)}$	1.30	1.320 (0.264, 0.070)	1.307 (0.081, 0.007)	1.333(0.736, 0.543)	1.305 (0.141, 0.020)
	C_2	$\theta_1^{(2)}$	1.35	1.367 (0.123, 0.015)	1.345 (0.166, 0.028)	1.356 (1.059, 1.122)	1.333 (1.755, 3.080)
		$\theta_2^{(2)}$	2.00	2.577 (1.273, 1.953)	3.180 (1.976, 5.297)	2.579(1.059, 1.457)	3.351(1.755, 4.905)
	C_3	$\theta_1^{(3)}$	2.85	2.852 (0.074, 0.005)	2.852 (0.074, 0.005)	2.835(0.816, 0.666)	2.833(0.816, 0.666)
		$\theta_2^{(3)}$	4.50	4.489 (0.133, 0.018)	4.488 (0.130, 0.017)	4.452(0.816, 0.668)	4.451(0.816, 0.668)
	rat. of correct states			0.958 (0.029)	0.933 (0.056)	0.958 (0.028)	0.925(0.058)
	$\sum_{i,j=1}^{d} \lvert\widehat{p}_{ij} - p_{ij}\rvert$			0.280 (0.234)	0.399 (0.299)	0.299 (0.235)	0.46 (0.325)
	rat. of correct structures			0.950	0.919	0.938	0.916

Table 21: Simulation results for different DGPs, sample size $T = 2000$, 1000 repetitions, standard deviations and MSEs are provided in brackets.

Simulation, II

Marginal distributions: $Y_{t1}, Y_{t2}, Y_{t3}, Y_{t4}, Y_{t5} \sim \boldsymbol{N}(0,1)$, $M = 3$, $n = 2000$, $d = 3$

$$C(u_1, C[u_2, C\{u_3, C(u_5, u_4; \theta_1^{(1)} = 3.15); \theta_2^{(1)} = 2.45\}; \theta_3^{(1)} = 1.75]; \theta_4^{(1)} = 1.05) \text{ for } i = 1,$$
$$C(u_3, C[u_5, C\{u_2, C(u_1, u_4; \theta_1^{(2)} = 3.15); \theta_2^{(2)} = 2.45\}; \theta_3^{(2)} = 1.75]; \theta_4^{(2)} = 1.05) \text{ for } i = 2,$$
$$C(u_5, C[u_4, C\{u_3, C(u_1, u_2; \theta_1^{(3)} = 3.15); \theta_2^{(3)} = 2.45\}; \theta_3^{(3)} = 1.75]; \theta_4^{(3)} = 1.05) \text{ for } i = 3,$$

Transition matrix:

$$P = \begin{pmatrix} 0.82 & 0.10 & 0.08 \\ 0.08 & 0.84 & 0.08 \\ 0.03 & 0.02 & 0.95 \end{pmatrix}$$

Simulation, II

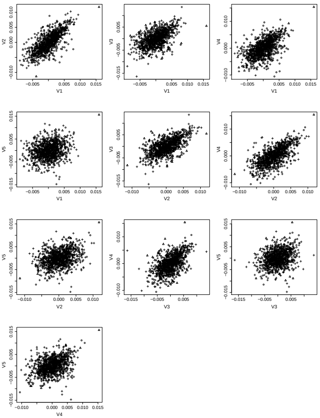


Figure 47: Snapshots of pairwise scatter plots of dependency structures ($t = 0, \ldots, 1000$).

Simulation, II

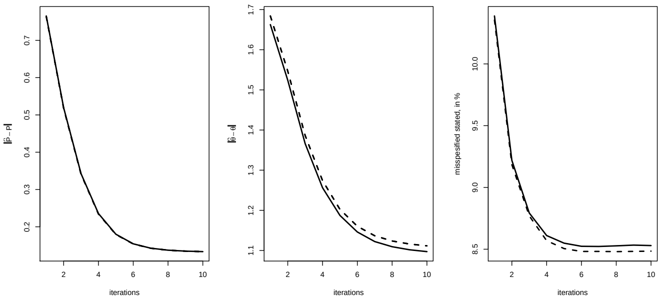


Figure 48: The averaged estimation errors for transition matrix (left panel), parameters (middle panel), convergence of states (right panel). Estimation starts from near true value (dashed); starts from values obtained by rolling window (solid). x-axis represents represents iterations. Number of repetitions is 1000.

Simulation, II

		True	Param. Margins	deGARCHing
C_1	$\theta_1^{(1)}$	1.05	1.019 (0.020, 0.001)	1.019 (0.020, 0.001)
	$\theta_2^{(1)}$	1.75	1.739 (0.077, 0.006)	1.741 (0.078, 0.006)
	$\theta_3^{(1)}$	2.45	2.584 (0.126, 0.034)	2.583 (0.126, 0.034)
	$\theta_4^{(1)}$	3.15	3.328 (0.194, 0.069)	3.318 (0.194, 0.066)
C_2	$\theta_1^{(2)}$	1.05	1.017 (0.021, 0.002)	1.017 (0.021, 0.002)
	$\theta_2^{(2)}$	1.75	1.795 (0.084, 0.009)	1.797 (0.084, 0.009)
	$\theta_3^{(2)}$	2.45	2.499 (0.120, 0.017)	2.499 (0.122, 0.017)
	$\theta_4^{(2)}$	3.15	3.381 (0.216, 0.100)	3.369 (0.215, 0.094)
C_3	$\theta_1^{(3)}$	1.05	1.044 (0.017, 0.000)	1.045 (0.018, 0.000)
	$\theta_2^{(3)}$	1.75	1.745 (0.041, 0.002)	1.747 (0.041, 0.002)
	$\theta_3^{(3)}$	2.45	2.492 (0.065, 0.006)	2.492 (0.065, 0.006)
	$\theta_4^{(3)}$	3.15	3.189 (0.094, 0.010)	3.185 (0.095, 0.010)
rat. of correct states			0.915 (0.011)	0.915 (0.011)
$\sum_{i,j=1}^{d} \lvert\widehat{p}_{ij} - p_{ij}\rvert$			0.133 (0.054)	0.133 (0.054)
rat. of correct structures			1	1

Table 22: The summary of estimation accuracy in five dimensional model, standard deviations and MSEs are provided in brackets. The case of deGARCHing is with nonparametrically estimated margins.

Simulation, III

- Simulation of data from three different true models: HMM GARCH, HMM ID and DCC
- Simulation of three dimensional time series from the simulated data
- Application of the models on the simulated data
- one-step ahead distribution forecast comparison of true and estimated model
- Comparison by Kolmogorov-Smirnov (KS) test statistics

Simulation, III

True \ Estimated	Sample size	HMMGARCH	HMM ID	DCC
HMM GARCH		**0.0899 (0.0353)**	0.1243 (0.0571)	0.1949 (0.1112)
DCC	250	**0.0607 (0.0241)**	0.0723 (0.0320)	0.0782 (0.0309)
HMM ID		0.0908 (0.0359)	**0.0867 (0.0345)**	0.1424 (0.0271)
HMMGARCH		**0.0889 (0.0338)**	0.1203 (0.0556)	0.2117 (0.0782)
DCC	500	**0.0541 (0.0194)**	0.0672 (0.0325)	0.0774 (0.0254)
HMM ID		**0.0936 (0.0331)**	0.0924 (0.0326)	0.1515 (0.0239)
HMM GARCH		**0.0869 (0.0321)**	0.1237 (0.0605)	0.3703 (0.1366)
DCC	1000	**0.0494 (0.0166)**	0.0659 (0.0320)	0.0823 (0.0392)
HMM ID		**0.0919 (0.0331)**	0.0907 (0.0322)	0.1509 (0.0213)

Table 23: The estimated mean KS test statistics (standard deviation) of the forecast distribution from the true model and the estimated model. Number of repetitions is 1000.

Application IV

- JPN/EUR, GBP/EUR and USD/EUR, from `DataStream`,
- [4.1.1999; 14.8.2009], 2771 obs.
- Fit to each marginal time series of log-returns a univariate GARCH(1,1) process:

$$X_{j,t} = \mu_{j,t} + \sigma_{j,t}\varepsilon_{j,t} \text{ with } \sigma^2_{j,t} = \omega_j + \alpha_j \sigma^2_{j,t-1} + \beta_j (X_{j,t-1} - \mu_{j,t-1})^2,$$

and $\omega > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$, $\alpha_j + \beta_j < 1$.

Application

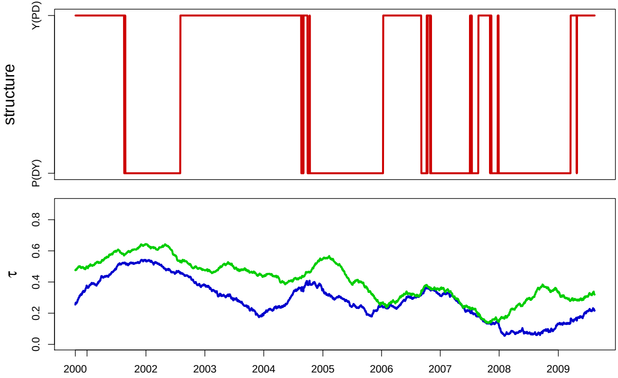


Figure 49: Rolling window for Exchange Rates: structure (upper) and parameters (lower, θ_1 and θ_2) for Gumbel HAC. $w = 250$.

Application

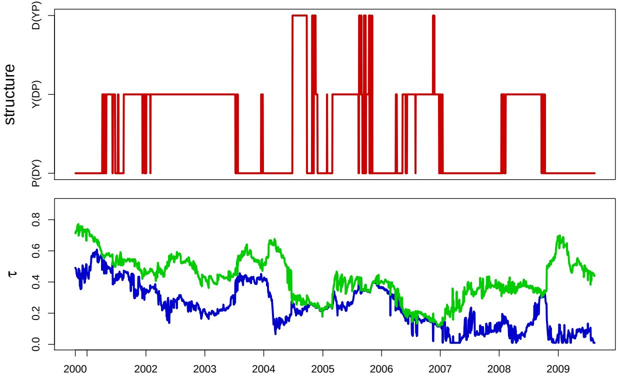


Figure 50: LCP for Exchange Rates: structure (upper) and parameters (lower, θ_1 and θ_2) for Gumbel HAC. $m_0 = 40$.

Application

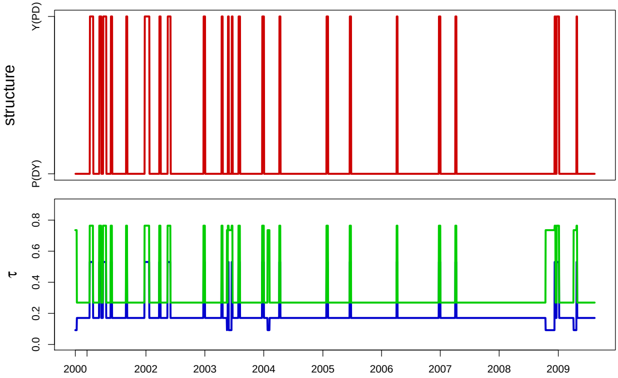


Figure 51: HMM for Exchange Rates: structure (upper) and parameters (lower, θ_1 and θ_2) for Gumbel HAC.

Application

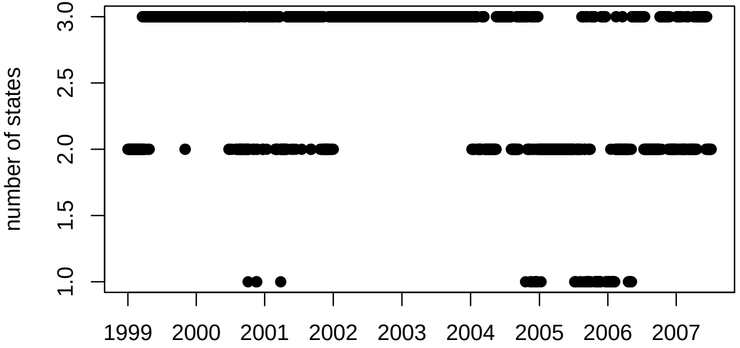


Figure 52: Plot of estimated number of states for each window

VaR

$T = 2219$, $N = 10^4$ is the sample size, $\omega = 1000$ portfolios.
The P&L function is $L_{t+1} = \sum_{i=1}^{3} w_i(y_{i,t+1} - y_{i,t})$. The VaR at level α is $VaR(\alpha) = F_L^{-1}(\alpha)$

$$\widehat{\alpha}_{\mathbf{w}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{I}\{L_t < \widehat{VaR}_t(\alpha)\}.$$

The distance between $\widehat{\alpha}$ and α

$$e_{\mathbf{w}} = (\widehat{\alpha}_{\mathbf{w}} - \alpha)/\alpha.$$

The performance of models is measured through

$$A_W = \frac{1}{|W|}\sum_{\mathbf{w}\in W} e_{\mathbf{w}}, \qquad D_W = \left\{\frac{1}{|W|}\sum_{\mathbf{w}\in W}(e_{\mathbf{w}} - A_W)^2\right\}^{1/2}.$$

Backtesting

	Window\\α	0.1	0.05	0.01
HMM, RGum	500	0.0980	**0.0507**	**0.0128**
HMM, Gum	500	**0.0981**	0.0512	0.0135
Rolwin, RGum	250	0.1037	0.0529	0.0151
Rolwin, Gum	250	0.1043	0.0539	0.0162
LCP, $m_0 = 40$	468	0.0973	0.0520	0.0146
LCP, $m_0 = 20$	235	0.1034	0.0537	0.0169
DCC	500	0.0743	0.0393	0.0163

Table 24: VaR backtesting results, $\overline{\widehat{\alpha}}$, where "Gum" denotes the Gumbel copula and "RGum" the rotated survival Gumbel one.

Backtesting

	Window\ α	0.1	0.05	0.01
HMM, RGum	500	-0.0204 (0.013)	**0.0147** (0.012)	**0.2827** (0.064)
HMM, Gum	500	-**0.0191** (0.008)	0.0233 (0.018)	0.3521 (0.029)
Rolwin, RGum	250	0.0375 (0.009)	0.0576 (0.012)	0.5076 (0.074)
Rolwin, Gum	250	0.0426 (0.009)	0.0772 (0.030)	0.6210 (0.043)
LCP, $m_0 = 40$	468	-0.0270 (0.010)	0.0391 (0.018)	0.4553 (0.037)
LCP, $m_0 = 20$	235	0.0344 (0.009)	0.0735 (0.026)	0.6888 (0.050)
DCC	500	-0.2573 (0.015)	-0.2140 (0.015)	0.6346 (0.091)

Table 25: Robustness relative to $A_W(D_W)$

Thank you for your attention !!!

Assumptions - Law of Large Numbers

Back LLN Back CLT Back Local Power Back Theorem

- Notation:
  - $\mathcal{N}(\theta^*)$ denote an open neighborhood of θ^*
  - $\ell_{\theta,j}(u_1,\ldots,u_d;\theta) = \frac{\partial \ell_\theta(u_1,\ldots,u_d;\theta)}{\partial u_j}$, $j = 1,\ldots,d$
  - $\ell_{\theta\theta,j}(u_1,\ldots,u_d;\theta) = \frac{\partial \ell_{\theta\theta}(u_1,\ldots,u_d;\theta)}{\partial u_j}$, $j = 1,\ldots,d$
- Assumptions:

A1: $\ell_\theta(u;\theta)$ and $\ell_{\theta\theta}(u;\theta)$ are continuous with respective to θ for any $u \in [0,1]^d$; there exist integrable functions $G_1(u)$ and $G_2(u)$ such that $\|\ell_\theta(u;\theta)\ell_\theta^\top(u;\theta)\| \le G_1(u)$, $\|\ell_{\theta\theta}(u;\theta)\| \le G_2(u)$ $\forall\theta \in \mathcal{N}(\theta^*)$

A2: Matrix $S(\theta^*) = -\operatorname{E}_0\left[\ell_{\theta\theta}\{F(X_1)\};\theta^*\right]$ is finite and nonsingular.

Assumptions - CLT I

Back CLT | Back Local Power | Back Theorem

B1: Denote $J_i(u) = const \times \prod_{k=1}^{d} \{u_k(1-u_k)\}^{-\xi_{ik}}$, where $\xi_{ik} \geq 0$, $i = 1, 2$, ξ_{ik} are some constants. Suppose that for all $\theta \in \mathbb{N}_{\theta^*}$, $\|\ell_\theta(u;\theta)\ell_\theta^\top(u;\theta)\| \leq J_1(u)$, $\|\ell_{\theta\theta}(u;\theta)\| \leq J_2(u)$, and $\mathrm{E}_0\left[J_i^2\{F(X_1)\}\right] < \infty$.

B2: Suppose that both $\ell_{\theta,k}(u;\theta)$ and $\ell_{\theta\theta,k}(u;\theta)$, $k = 1, 2, \ldots, d$ exist and are continuous. Denote $\widetilde{J}_i^k(u) = const \times \{u_k(1-u_k)\}^{-\widetilde{\xi}_{ik}} \prod_{j=1, j\neq k}^{d} \{u_j(1-u_j)\}^{-\xi_{ij}}$, where $\widetilde{\xi}_{ij} > \xi_{ij}$ are some constants, such that for all $\theta \in \mathbb{N}(\theta^*)$, $\|\ell_{\theta,k}(u;\theta)\| \leq \widetilde{J}_1^k(u)$ and $\|\ell_{\theta\theta,k}(u;\theta)\| \leq \widetilde{J}_2^k(u)$, and furthermore, $\mathrm{E}_0\left[\widetilde{J}_i\{F(X_1)\}\right] < \infty$, $i = 1, 2$ and $k = 1, 2, \ldots, d$.

Assumptions - CLT II

▸ Back CLT ▸ Back Theorem

B3: Suppose $\frac{\partial \ell_{\theta\theta}(u;\theta)}{\partial \theta_k}$, $k = 1, 2, \ldots, p$ exist and are continuous with $\theta \in \mathbb{N}(\theta^*)$, and there exists an integrable function $G_3(u)$ such that $\|\frac{\partial \ell_{\theta\theta}(u;\theta)}{\partial \theta_k}\| \leq G_3(u)$ for all $\theta \in \mathbb{N}(\theta^*)$, $k = 1, \ldots, d$.

C1: The block size m is of order $o(n^a)$ with $0 \leq a \leq \frac{1}{4}$.

Assumption - Local Power of Evaluation

Back Local Power

D1: Both the copula $C_0(\cdot;\theta_0)$ and $C_1(\cdot)$ in $P_n^{C_1,\delta}(x)$ are absolutely continuous with respective to square integrable densities $c_0(\cdot;\theta_0)$ and $c_1(\cdot)$. Moreover

$$\int_{u\in[0,1]^d}\left[\sqrt{n}\left\{\sqrt{p_n^{c_1,\delta}(u)}-\sqrt{p_0(u)}\right\}-\frac{1}{2}\delta g(u)\sqrt{p_0(u)}\right]^2du\to 0,$$

as $n\to\infty$, where $p_n^{c_1,\delta}(u)=(1-\frac{\delta}{\sqrt{n}})c_0(u;\theta_0)+\frac{\delta}{\sqrt{n}}c_1(u)$, $p_0(u)=c_0(u;\theta_0)$ and $g(u)=\frac{c_1(u)-c_0(u;\theta_0)}{c_0(u;\theta_0)}$.

Assumptions - Large sample properties I

Back Theorem

E1. $\left\{\left(Y_t^\top, Z_t^\top\right), t = 1, \ldots, n\right\}$ is stationary β-mixing with serial decay rate of order $O(t^{-\frac{\xi}{\xi-1}})$ for some $\xi > 1$

E2. $\widehat{\eta}$ is a root-n consistent estimator of η_0

E3. For all $t \geq 1$ and $j = 1, \ldots, d$, $\epsilon_{tj} = \Sigma_{tj}^{-1/2}(\eta^0)\left\{Y_{tj} - \mu_{tj}(\eta_1^0)\right\}$ is continuously differentiable in the neighbood of η^0, and
$\omega_1 = \mathrm{E}_0\left\{\Sigma_{tj}^{-1/2}(\eta^0)\dot{\mu}_{tj}(\eta_1^0)\right\} < \infty$ and
$\omega_2 = \mathrm{E}_0\left\{\Sigma_{tj}^{-1}(\eta^0)\dot{\Sigma}_{tj}(\eta^0)\right\} < \infty$, where $\dot{\mu}_{tj}(\eta_1^0) = \frac{\partial \mu_{tj}(\eta_1^0)}{\partial \eta_1}$ and
$\dot{\Sigma}_{tj}(\eta^0) = \frac{\partial \Sigma_{tj}(\eta^0)}{\partial \eta}$.

Assumptions - Large sample properties II

E4. The PMLE $\widehat{\theta}$ has the following asymptotic expansion

$$\widehat{\theta} - \theta^* = \frac{1}{n}\sum_{t=1}^{n} \varphi_\theta\left(U_t; \theta^*\right) + o_p(n^{-1/2}),$$

where $U_t = (U_{t1}, \dots, U_{td})^\top$, $U_{tj} = F_j(\epsilon_{tj})$, $j = 1, \dots, d$, $t = 1, \dots, n$ and

$$\varphi_\theta\left(U_t; \theta^*\right) = S(\theta^*)^{-1}(\ell_\theta\left(U_t; \theta^*\right) + \sum_{j=1}^{d} \mathsf{E}_0\left[\ell_{\theta,j}\left(U_s; \theta^*\right)\left\{\mathbf{I}(U_{tj} \le U_{sj}) - U_{sj}\right\} | U_{tj}\right]).$$

Assumptions - Penalization I

Define $\ell_i(\theta) = \log c(U_{i1}, \dots, U_{id_k}; \theta)$:

(1) Model is identifiable and $\theta_{k(\ell),0}$ is an interior point of the compact parameter space $\Theta_{k(\ell)}$. We assume that $\boldsymbol{E}_{\theta_{k(\ell)}}\{\ell_i'(\theta_{k(\ell)})\} = 0$ and information equality holds,

$$V(\theta_{k(\ell)}) \stackrel{\text{def}}{=} \boldsymbol{E}_{\theta_{k(\ell)}}\left\{\ell_i'(\theta_{k(\ell)})^2\right\} = -\boldsymbol{E}_{\theta_{k(\ell)}}\left\{\ell_i''(\theta_{k(\ell)})\right\}$$

for $i = 1, \dots, n$.

(2) Fisher information $V(\theta_{k(\ell)})$ is finite and strictly positive at $\theta_{k(\ell),0}$.

Assumptions - Penalization II

(3) There exists an open subset Ω of $\Theta_{k(\ell)}$ containing the true parameter $\theta_{k(\ell),0}$ such that for almost all U_i, $i = 1, \ldots, n$, the density $c(U_{i1}, \ldots, U_{id_k}; \theta_{k(\ell)})$ admits all third derivatives $c'''(\cdot; \theta_{k(\ell)})$ for all $\theta_{k(\ell)} \in \Omega$. Furthermore, there exist functions $M(\cdot)$ such that $\left|\ell_i'''(\theta_{k(\ell)})\right| \leq M(U_i)$, for all $\theta_{k(\ell)} \in \Omega$, with $\boldsymbol{E}\left\{M(U_i)\right\} < \infty$.

Penalized ML