Dynamics of the Multivariate Copula-based Models

Wolfgang Härdle
Ostap Okhrin
Yarema Okhrin

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

Simple AC over time

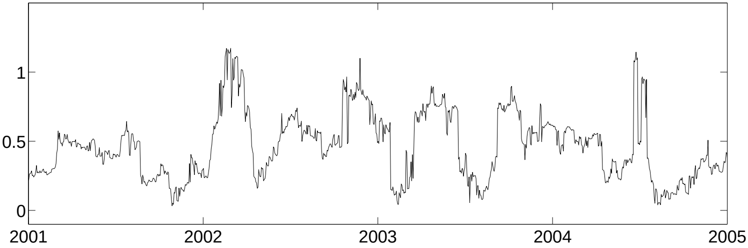


Figure 1: Estimated copula dependence parameter $\widehat{\theta}_t$ with the Local Change Point method for 6-dimensional data: DC, VW, Bayer, BASF, Allianz and Münchener Rückversicherung. Clayton Copula.
Giacomini et. al (2008)

Main Idea

- combine interpretability with flexibility without loosing statistical precision
- determine the optimal structure of HAC for a given data set
- find the intervals of the homogeneity of the dependency

Outline

Archimedean Copulæ

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

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

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.

Example

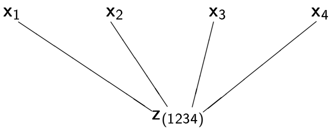
$$\begin{aligned} \phi_{Gumbel}(u, \theta) &= \exp\{-u^{1/\theta}\}, \text{ where } 1 \leq \theta < \infty \\ \phi_{Clayton}(u, \theta) &= (\theta u + 1)^{-1/\theta}, \text{ where } \theta \in [-1, \infty) \backslash \{0\} \end{aligned}$$

Disadvantages: too restrictive, single parameter, exchangeable

Hierarchical Archimedean Copulæ

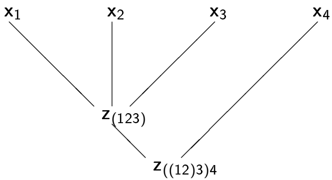
Simple AC with s=(1234)

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



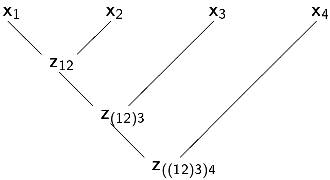
AC with s=((123)4)

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



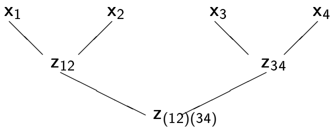
Fully nested AC with s=(((12)3)4)

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



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 Copulæ

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
- subcopulæ are also HAC

Hierarchical Archimedean Copulæ

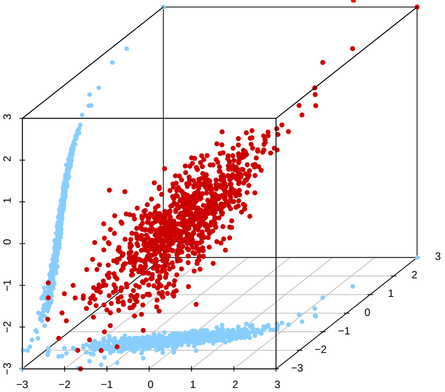


Figure 2: Scatterplot of the
$C_{Gumbel}[C_{Gumbel}\{\Phi(x_1), t_2(x_2); \theta_1 = 2\}, \Phi(x_3); \theta_2 = 10]$

Hierarchical Archimedean Copulæ

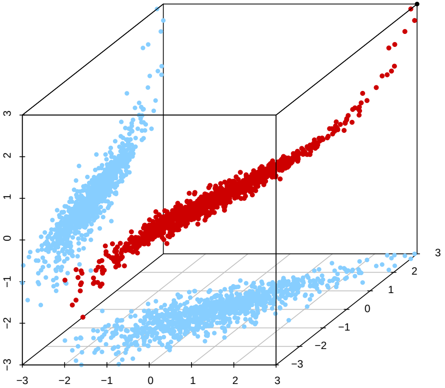


Figure 3: Scatterplot of the
$C_{Gumbel}[\Phi(x_2), C_{Gumbel}\{t_2(x_1), \Phi(x_3); \theta_1 = 2\}; \theta_2 = 10]$

Data and Copula

daily returns of four companies listed in DAX index

company: Commerzbank (CBK), Merck (MRK) and ThyssenKrupp (TKA)

timespan = [13.11.1998 - 18.10.2007] ($n = 2400$)

$\mathcal{M} = \{\phi = \exp(-u^{1/\theta})\}$ - Gumbel generator

Data and Copula

- GARCH-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 and θ_t are the copula parameters.
- in each model we assumed that margins are normal, however

Copulæ over time

window for 250 days

$\boldsymbol{\Theta}_t(d \times d)$ - matrix of the pairwise θ based on the 250 days before t

$$\begin{aligned}
||\widehat{\boldsymbol{\Theta}}_t - \widehat{\boldsymbol{\Theta}}_{t-1}||_1 &= \max_{1 \le i \le d} \sum_{j=1}^{d} |\widehat{\theta}_{ij,t} - \widehat{\theta}_{ij,t-1}|, \\
||\widehat{\boldsymbol{\Theta}}_t - \widehat{\boldsymbol{\Theta}}_{t-1}||_2 &= \sqrt{\lambda_{\max}\{(\widehat{\boldsymbol{\Theta}}_t - \widehat{\boldsymbol{\Theta}}_{t-1})(\widehat{\boldsymbol{\Theta}}_t - \widehat{\boldsymbol{\Theta}}_{t-1})^\top\}}
\end{aligned}$$

Copulæ over time

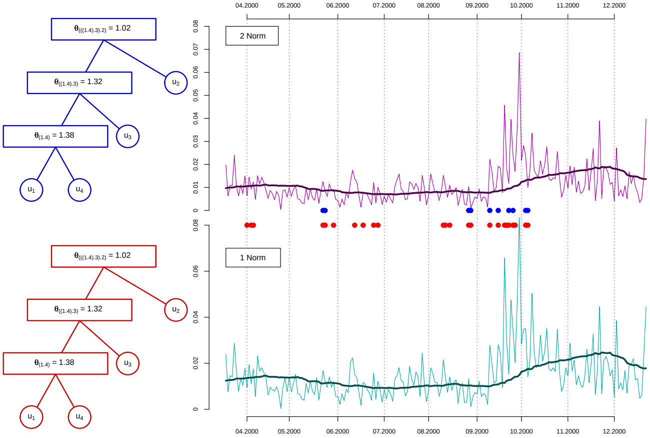
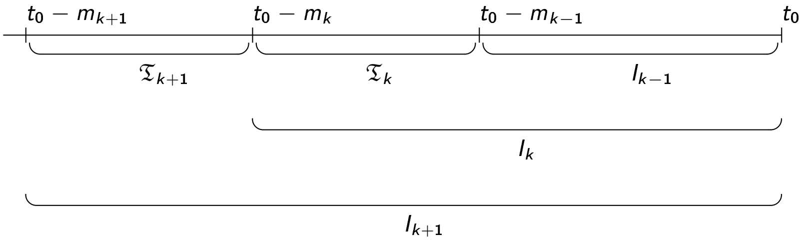


Figure 4: Film of time-varying HAC

Local Change Point Detection

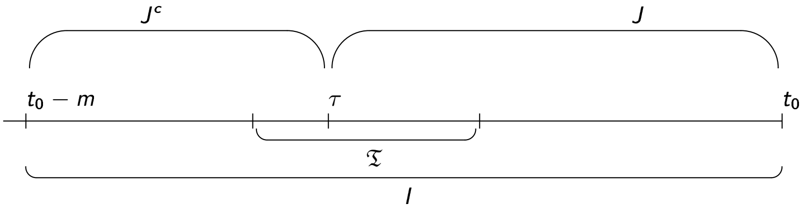
1. define family of nested intervals
   $I_0 \subset I_1 \subset I_2 \subset \ldots \subset I_K = I_{K+1}$ with length m_k as
   $I_k = [t_0 - m_k, t_0]$
2. define $\mathfrak{T}_k = [t_0 - m_k, t_0 - m_{k-1}]$



Test of homogeneity

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

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



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

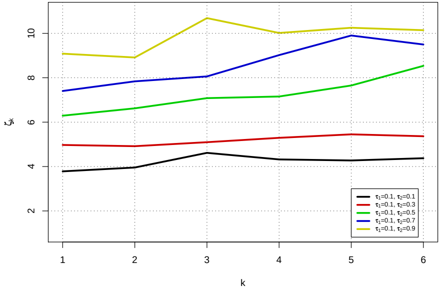
$$
T_I > \zeta_I
$$

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 $k = 1, 2, \ldots, K$, $m_0 = 40$ and $c = 1.25$, 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 $k = 1, 2, \ldots, K$
- estimated copula from the whole sample $s^* = ((1.3)_{1.331}.2)_{1.145}$
- ζ_I is selected by a simulation from the true structure s^*

Selection 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. $\theta = \theta(\tau)$
3. simulating from $C_{\theta_i, \theta_j}(u_1, u_2, u_3) = C\{C(u_1, u_2; \theta_1), u_3; \theta_2\}$
4. calculation of the 0.95 quantile



LCP for HAC

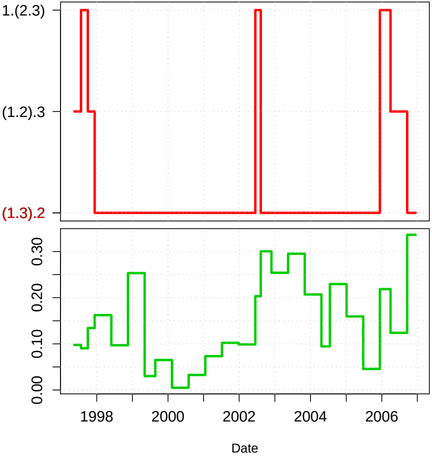


Figure 5: Structure and ML of the estimated HAC on the intervals of homogeneity

LCP for Simulated Data

1. $N = 1000$,
   1.1 on the sample $[1;500]$ the structure is $(1.(2.3)_{1.25})_{1.111}$
   1.2 on the sample $[501;1000]$ the structure is $((1.2)_{1.25}.3)_{1.111}$
2. 10 runs
3. LCP based on the same critical values

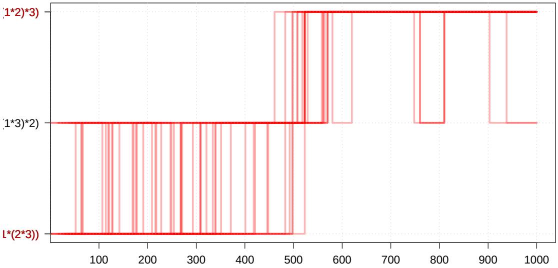


Figure 6: The structure of the est. HAC on the intervals of homogeneity

Kendall's τ for HAC

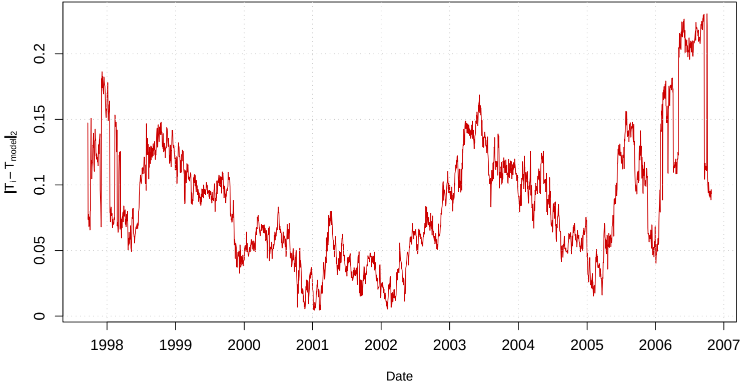


Figure 7: Norm of the difference between Kendall's τ matrices for the model and from the data

Programming Issues

- ADOL-C – C++/C package for automatic differentiation (gradients, Jacobians, tensors, hessians etc.)
- Numerical Intergation – for computation statistic, and for calculation inverse chi-square cdf.
- Inverse Functions – mainly for simulating (for calculating inverse conditional distribution functions)
- OOP – containers and classes for simplification the work
  1. `TTree` class for representation of HAC;
  2. `myMath` unit containing mathematical classes as `TVector`, `TMatrix` with statistical methods;
  3. `THACSimulation` and `THACSolution` classes for simulation the sample and estimation the HAC function respectively
- Kernel Estimation, Normal CDF, Maximization Problems, etc.

Simulation

Frees and Valdez, (1998), Whelan, (2004), Marshal and Olkin, (1988), Hofert (2007)

Conditional inversion method:

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

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

1. Generate i.r.v. $v_1, \dots, v_k \sim U(0,1)$
2. Set $u_1 = v_1$
3. $u_i = C_k^{-1}(v_i|u_1, \dots, u_{i-1})\ \forall i = \overline{2, k}$

Criteria for grouping using GOF

$$H_0: \quad C = C_0, \quad \text{against} \quad H_1: \quad C \neq C_0.$$

probability integral transform, Rosenblatt (1952, AMS)

$$\begin{aligned} Y_{1i} &= \widehat{F}_1(x_{1i}), \\ Y_{ji} &= C(\phi, \widehat{\theta}, s)(\widehat{F}_j(x_{ji}) | \widehat{F}_1(x_{1i}), \ldots, \widehat{F}_{j-1}(x_{j-1,i})). \end{aligned}$$

$$\begin{aligned} \widehat{W}_i &= \sum_{j=1}^{d} [\Phi^{-1}(Y_{ji})]^2, \\ \widehat{g}_W(w) &= \frac{1}{nh} \sum_{i=1}^{n} K_h(w, F_{\chi^2_d}(\widehat{W}_i)), \\ \widehat{J}_n &= \int_0^1 (\widehat{g}_W(w) - 1)^2 \, dw \end{aligned}$$

test statistic (Chen et al. 2004)

$$T_n = \frac{(n\sqrt{h}\widehat{J}_n - c_n)}{\sigma} \to \mathcal{N}(0, 1).$$

E. Giacomini, W. Härdle, V. Spokoiny
Inhomogeneous Dependency Modelling with Time Varying Copulæ
J. Bus. Econ. Statist., 2008

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

D. Mercurio, and V. Spokoiny
Estimation of Time Dependent Volatility via Local Change Point Analysis
Ann. Statist., **32**: 577-602, 2004

R. Nelsen
An intoduction to copulas
Springer, 1999

O. Okhrin, Y. Okhrin, W. Schmid
On the structure and estimation of hierarchical Archimedean copulas
under revision in JoE