Fitting copula to Data

Ostap Okhrin

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

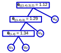
Outline

1. Motivation ✓
2. Hierarchical Archimedean copulae
3. Recovering the Structure
4. GoF
5. Bibliography

Gaussian Copula

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

- 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



Archimedean Copula

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

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

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

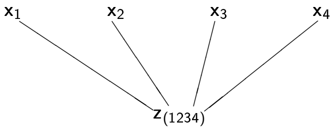
$$\begin{aligned} \phi_{Gumbel}(u,\theta) &= \exp\{-u^{1/\theta}\}, \text{ where } 1 \le \theta < \infty \\ \phi_{Clayton}(u,\theta) &= (\theta u + 1)^{-1/\theta}, \text{ where } \theta \in [-1,\infty) \backslash \{0\} \end{aligned}$$

Disadvantages: too restrictive, single parameter, exchangeable

Hierarchical Archimedean Copulae

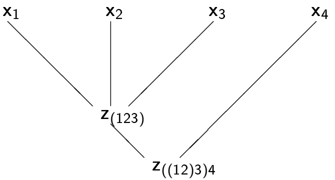
Simple AC with s=(1234)

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



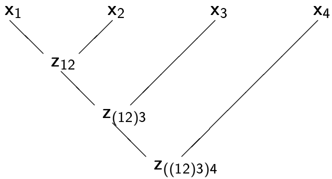
AC with s=((123)4)

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



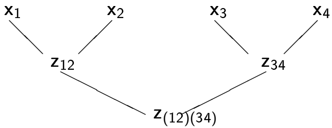
Fully nested AC with s=(((12)3)4)

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



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

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



Theoretical motivation

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

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

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

$$\int_0^\infty G_1^\alpha(x_1) \cdot \cdots \cdot G_d^\alpha(x_d) dM(\alpha) = \phi\left\{-\sum_{i=1}^d \ln G_i(x_i)\right\} = \phi\left[\sum_{i=1}^d \phi^{-1}\{F_i(x_i)\}\right].$$

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

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

Recovering the structure (theory)

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

Estimation Issues

$$
\begin{aligned}
F_j(x;\widehat{\alpha}_j) &= F_j(x; arg\max_\alpha \sum_{i=1}^{n} \log\ f_j(X_{ji},\alpha)),\\
\widehat{F}_j(x) &= \frac{1}{n+1}\sum_{i=1}^{n} \mathsf{I}(X_{ji} \leq x),\\
\widetilde{F}_j(x) &= \frac{1}{n+1}\sum_{i=1}^{n} K\left(\frac{x - X_{ji}}{h}\right)
\end{aligned}
$$

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.

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

Estimation Issues

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

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

Estimation Issues

Nonparametric Estimation

$$\widehat{C}(u_1, \ldots, u_d) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \mathrm{I}\{\breve{F}_j(X_{ji}) \leq u_j\}$$

$$\widetilde{C}(u_1, \ldots, u_d) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} K_j \left\{ \frac{u_j - \breve{F}_j(X_{ji})}{h_j} \right\}$$

where $\breve{F}_j(x) = \{\widehat{F}_j(x),\ \widetilde{F}_j(x),\ F_j(x, \widehat{\alpha}),\ F_j(x)\}$

K-distribution

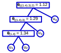
Let $V = C\{F_1(X_1), \dots, F_d(X_d)\}$ and let $K(t)$ denote the distribution function (K-distribution) of the random variable V.

We consider a HAC of the form $C_1\{u_1, C_2(u_2, \dots, u_d)\}$. Let $U_i \sim U[0,1]$ and let $V_2 = C_2(U_2, \dots, U_d) \sim K_2$.

Theorem

Let $U_1 \sim U[0,1]$, $V_2 \sim K_2$ and let $P(U_1 \le x, V_2 \le y) = C_1\{x, K_2(y)\}$ 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

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

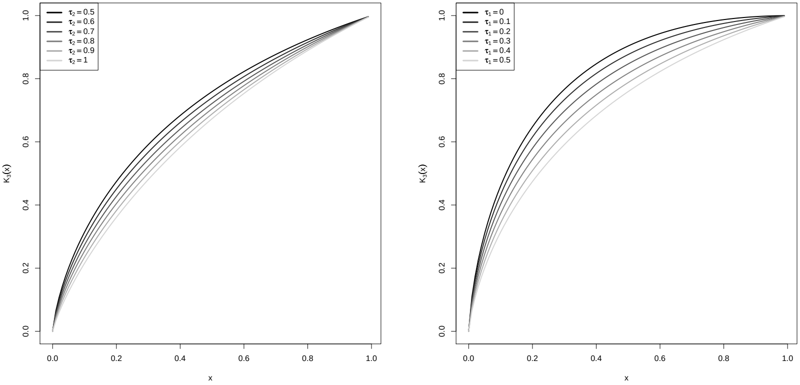


Figure 1: K distribution for three-dimensional HAC with Gumbel generators

Goodness-of-Fit Tests

$$H_0: \quad C \in \mathcal{C}_0, \quad \text{against} \quad H_1: \quad C \notin \mathcal{C}_0,$$

where $\mathcal{C}_0 = \{C_\theta : \theta \in \Theta\}$ is a known parametric family of copulae.

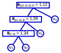
$$S = n \int_{[0,1]^d} \{\widehat{C}(u_1, \ldots, u_d) - C(u_1, \ldots, u_d, \widehat{\theta})\}^2 \, d\widehat{C}(u_1, \ldots, u_d),$$

$$T = \sup_{u_1, \ldots, u_d \in [0,1]} \sqrt{n} |\widehat{C}(u_1, \ldots, u_d) - C(u_1, \ldots, u_d, \widehat{\theta})|,$$

$$S_K = n \int_0^1 \{\widehat{K}(v) - K(v, \theta)\}^2 \, dv,$$

$$T_K = \sup_{v \in [0,1]} |\widehat{K}(v) - K(v, \theta)|.$$

where $\widehat{K}(v) = \frac{1}{n} \sum_{i=1}^n \mathrm{I}\{V_i \leq v\}$.



Simulation Study

1. F : two methods of estimation of margins (parametric and nonparametric);
2. C_0 : hypothesised copula models under H_0 (three models);
3. C : copula model from which the data were generated (three models with 3, 3 and 15 levels of dependence respectively);
4. n : size of each sample drawn from C (two possibilities, $n = 50$ and $n = 150$).

$\rightsquigarrow 2 \times 3 \times (3 + 3 + 15) \times 2 =$ 252 models with 100 repetitions

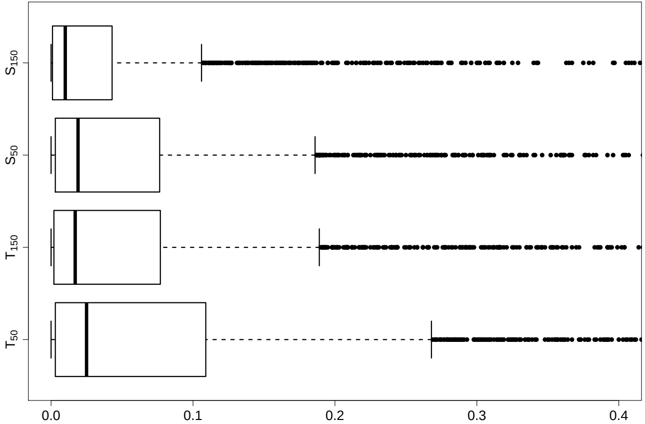


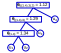
Figure 2: Levels of goodness-of-fit tests for different sample size, for parametric margins.

		AC							
		$n = 50$				$n = 150$			
θ		T		S		T		S	
		emp.	par.	emp.	par.	emp.	par.	emp.	par.
$\theta(0.25)$	HAC	0.88	0.51	0.83	0.38	0.93	0.36	0.90	0.35
	AC	0.88	0.51	0.89	0.50	0.95	0.32	0.90	0.34
	Gauss	0.71	0.29	0.56	0.22	0.69	0.11	0.43	0.08
$\theta(0.5)$	HAC	0.90	0.38	0.94	0.30	0.87	0.35	0.88	0.27
	AC	0.96	0.55	0.95	0.45	0.90	0.45	0.92	0.35
	Gauss	0.76	0.30	0.65	0.19	0.47	0.13	0.31	0.02
$\theta(0.75)$	HAC	0.93	0.29	0.93	0.15	0.89	0.27	0.89	0.10
	AC	0.93	0.29	0.93	0.22	0.90	0.25	0.91	0.13
	Gauss	0.77	0.19	0.65	0.10	0.57	0.11	0.24	0.05

Table 1: Non-rejection rate of the different models, where the sample is drawn from the simple AC

		HAC							
		$n = 50$				$n = 150$			
θ		T		S		T		S	
		emp.	par.	emp.	par.	emp.	par.	emp.	par.
$\theta(0.25, 0.5)$	HAC	0.88	0.29	0.90	0.24	0.96	0.31	0.92	0.26
	AC	0.91	0.26	0.93	0.36	0.54	0.13	0.53	0.07
	Gauss	0.82	0.20	0.69	0.19	0.57	0.14	0.37	0.04
$\theta(0.25, 0.75)$	HAC	0.93	0.21	0.92	0.13	0.88	0.18	0.88	0.09
	AC	0.46	0.14	0.54	0.07	0.00	0.00	0.00	0.00
	Gauss	0.84	0.19	0.71	0.13	0.52	0.10	0.42	0.01
$\theta(0.5, 0.75)$	HAC	0.86	0.31	0.87	0.18	0.91	0.20	0.94	0.08
	AC	0.89	0.36	0.92	0.28	0.44	0.04	0.47	0.02
	Gauss	0.70	0.19	0.55	0.12	0.50	0.11	0.30	0.05

Table 2: Non-rejection rate of the different models, where the sample is drawn from the HAC



Gauss									
		$n = 50$				$n = 150$			
Σ		T		S		T		S	
		emp.	par.	emp.	par.	emp.	par.	emp.	par.
$\Sigma(0.25, 0.25, 0.75)$	HAC	0.89	0.20	0.93	0.11	0.78	0.08	0.81	0.02
	AC	0.43	0.13	0.47	0.09	0.00	0.00	0.00	0.00
	Gauss	0.88	0.22	0.89	0.12	0.87	0.11	0.86	0.03
$\Sigma(0.25, 0.75, 0.25)$	HAC	0.92	0.20	0.91	0.14	0.76	0.07	0.69	0.04
	AC	0.39	0.12	0.39	0.04	0.00	0.00	0.00	0.00
	Gauss	0.90	0.18	0.87	0.13	0.92	0.12	0.94	0.10
$\Sigma(0.75, 0.25, 0.25)$	HAC	0.89	0.30	0.93	0.16	0.78	0.10	0.75	0.04
	AC	0.51	0.16	0.46	0.07	0.00	0.00	0.00	0.00
	Gauss	0.91	0.28	0.90	0.17	0.88	0.13	0.86	0.06

Table 3: Non-rejection rate of the different models, where the sample is drawn from the Gaussian copula

Data and Copula

- daily returns of Bank of America, Citigroup, Santander
- timespan = [29.09.2000 - 16.02.2001] ($n = 100$)
- ARMA(1,1)-GARCH(1,1)-residuals are conditionally distributed with estimated copula

$$
\begin{aligned}
R_{tj} &= \mu_j + \gamma_j R_{t-1,j} + \zeta_j \sigma_{t-1,j}\varepsilon_{t-1,j} + \sigma_{tj}\varepsilon_{tj}, \\
\sigma^2_{tj} &= \omega_j + \alpha_j \sigma^2_{t-1,j} + \beta_j \sigma^2_{t-1,j}\varepsilon^2_{t-1,j} \\
\varepsilon &\sim C\{F_1(x_1), \dots, F_d(x_d); \theta_t\}
\end{aligned}
$$

where $F_1, \dots, F_d$ are marginal distributions and θ_t are the copula parameters and $\omega > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$, $\alpha_j + \beta_j < 1$, $|\zeta| < 1$.

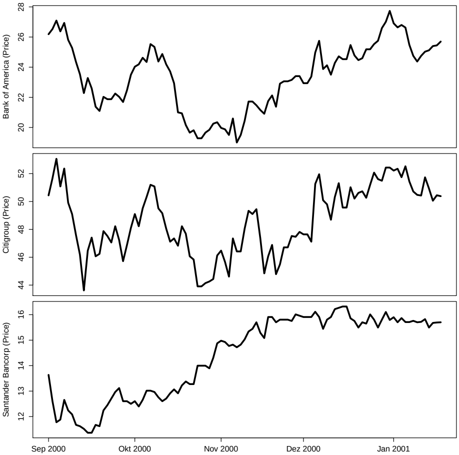


Figure 3: Stock prices for Bank of America, Citigroup and Santander (from top to bottom).

	$\widehat{\mu}_j$	$\widehat{\gamma}_j$	$\widehat{\zeta}_j$	$\widehat{\omega}_j$	$\widehat{\alpha}_j$	$\widehat{\beta}_j$
Bank of America	1.87e-3	0.22	-0.23	3.46e-4	0.55	0.17
(0.57, 0.83)	(2.59e-3)	(0.64)	(0.65)	(1.37e-04)	(0.28)	(0.16)
Citigroup	0.11e-3	0.31	-0.46	2.67e-4	0.09	0.47
(0.57, 0.79)	(1.48e-3)	(0.29)	(0.29)	(5.53e-04)	(0.17)	(1.01)
Santander	1.35e-3	0.43	-0.56	4.51e-10	0.01	0.98
(0.91, 0.78)	(0.91e-3)	(0.15)	(0.17)	(1.38e-05)	(0.02)	(0.05)

Table 4: Fitting of univariate ARMA(1,1)-GARCH(1,1) to asset returns. The standard deviation of the parameters, which are quiet big because of the small sample size, are given in parentheses. Each second row provides the p-values of the Box-Ljung test (BL) for autocorrelations and Kolmogorov-Smirnov test (KS) for testing of normality of the residuals.

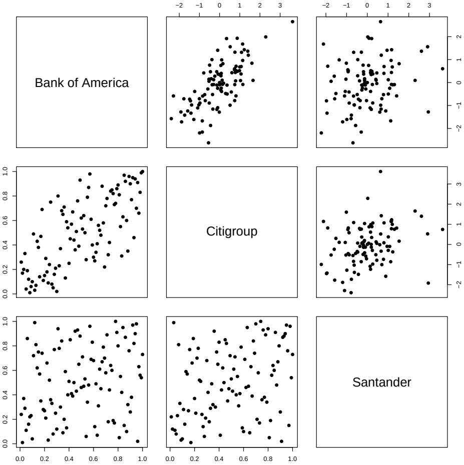


Figure 4: Scatterplots from ARMA-GARCH residuals (upper triangular) and from residuals mapped on unit square by the cdf (lower triangular).

	T_{100}	S_{100}	estimates
HAC	0.3191	0.1237	$C\{C(u_1, u_2; 1.996), u_3; 1.256\}$
AC	0.0012	0.0002	$C(u_1, u_2, u_3; 1.276)$
Gauss	0.0160	0.0078	$C_N\{u_1, u_2, u_3; \Sigma(0.697, 0.215, 0.312)\}$

Table 5: p-values of both GoFs and estimates of the models under different H_0 hypotheses.

The value V_t of the portfolio $w = \{w_1, \ldots, w_d\},\ w_i \in \mathbb{Z}$ is given by

$$V_t = \sum_{j=1}^{d} w_j X_{tj} \tag{2}$$

and the profit and loss (P&L) function of the portfolio

$$L_{t+1} = (V_{t+1} - V_t) = \sum_{j=1}^{d} w_j X_{tj} \left\{ \exp(R_{t+1,j}) - 1 \right\} \tag{3}$$

The distribution function of L is given by

$$F_L(x) = P(L \leq x). \tag{4}$$

As usual the Value-at-Risk at level α from a portfolio w:

$$\mathrm{VaR}(\alpha) = F_L^{-1}(\alpha). \tag{5}$$

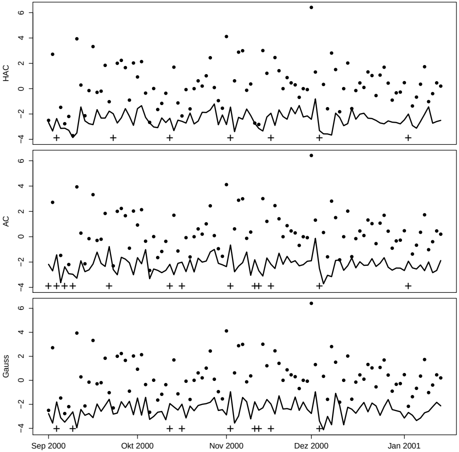


Figure 5: Profit and loss function and VaR based on different models.

α	$\widehat{\alpha}_{HAC}$	$\widehat{\alpha}_{AC}$	$\widehat{\alpha}_{Gauss}$
0.10	0.091	0.122	0.081
0.05	0.040	0.061	0.031
0.01	0.000	0.010	0.000

Table 6: Backtesting for the estimation of VaR under different alternatives.

A. Charpentier, J. Segers,
Tails of multivariate Archimedean copulae,
Journal of Multivariate Analysis 100 (2009) 1521-1537.

P. Barbe, C. Genest, K. Ghoudi, B. Rémillard,
On Kendall's Process,
Jurnal of Multivariate Analysis 58 (1996) 197-229.

A. Sklar,
Fonctions dé Repartition á n Dimension et Leurs Marges,
Publ. Inst. Stat. Univ. Paris 8 (1959) 299-231

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

O. Okhrin, Y. Okhrin, W. Schmid
Properties of Hierarchical Archimedean Copulas,
submitted for publication.

O. Okhrin, Y. Okhrin, W. Schmid
On the Structure and Estimation of Hierarchical Archimedean Copulas,
under revision in *Journal of Econometrics.*

M. Odening, O. Okhrin, W. Xu
Systemic Nature of Weather Risk,
to appear in *Agricultural Finance Review.*

B. Choroś, W. Härdle, O. Okhrin
HAC and CDO,
submitted for publication.

W. Härdle, O. Okhrin, Y. Okhrin
Time Varying HAC,
submitted for publication.

W. Härdle, O. Okhrin
De copulis non est disputandum,
Advances of Statistical Analysis - AStA: 94(1), 2010, pp. 1-31