Copula methods in Finance

Wolfgang K. Härdle
Ostap Okhrin

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

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 Latinate 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%

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

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?

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

Outline

1. Motivation ✓
2. Univariate Distributions and their Estimation
3. Multivariate Distributions and their Estimation
4. Copulae
5. Classical Approach
6. Our Findings

Univariate Case

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

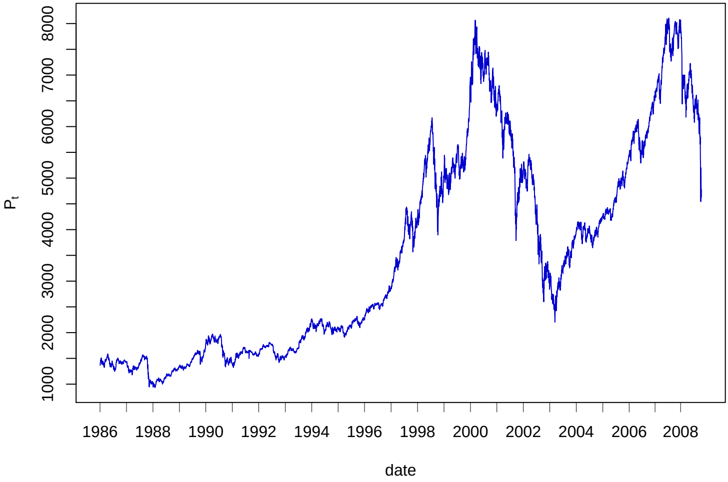
Example 1

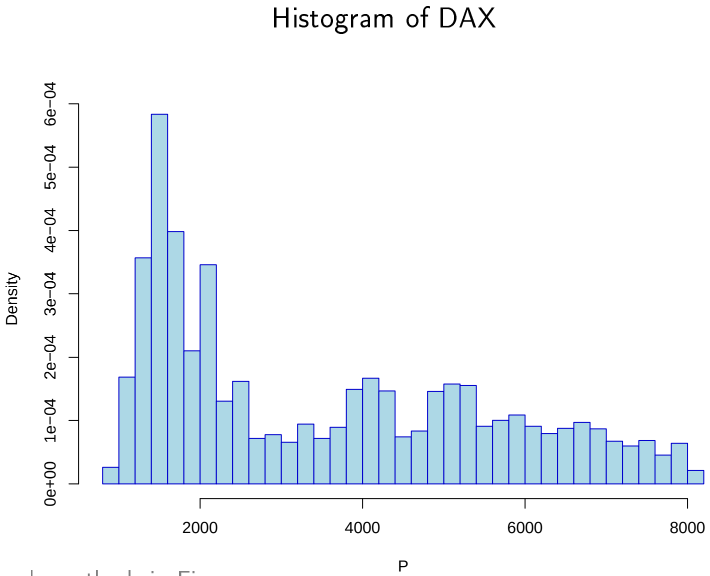
- x_i are returns of the asset for one firm at the day t_i
- x_i are numbers of sold albums *The Man Who Sold the World* by *David Bowie* at day t_i

What is a good approximation of F ?

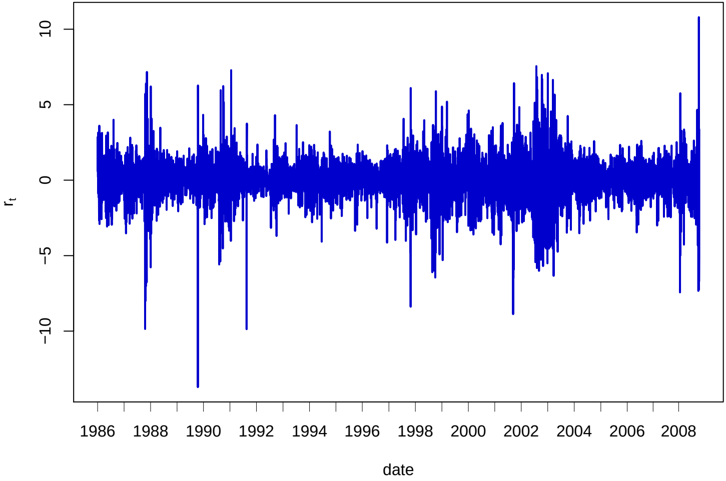
`traditional` or `modern` approach

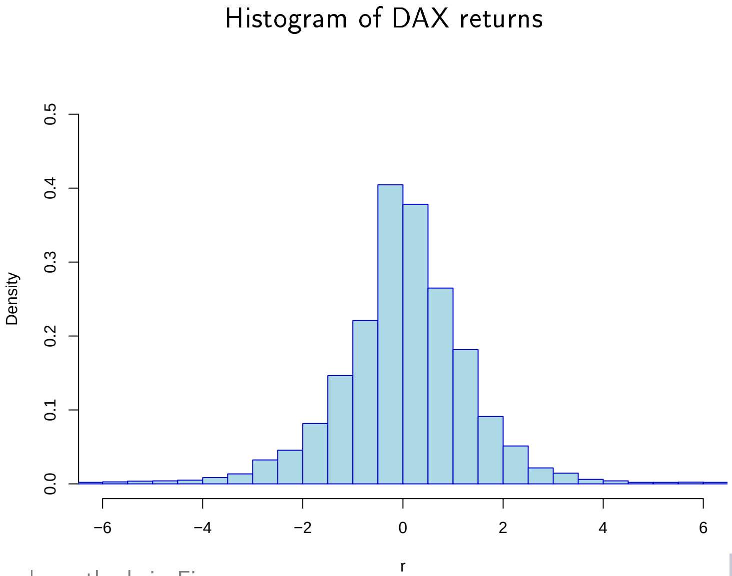
DAX (P_t)


8000
7000
6000
5000
4000
3000
2000
1000
P_t
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
date


Histogram of DAX
Density
0e+00
1e−04
2e−04
3e−04
4e−04
5e−04
6e−04
2000
4000
6000
8000
P

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




Histogram of DAX returns
Density
0.0
0.1
0.2
0.3
0.4
0.5
−6
−4
−2
0
2
4
6
r

Traditional approach:

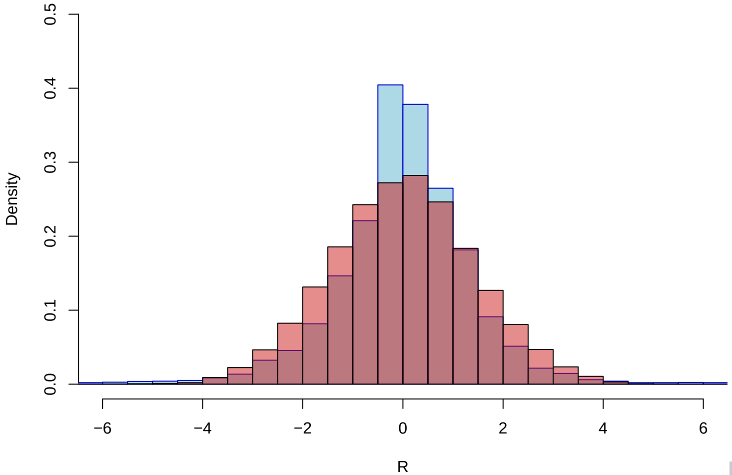
F_0 – known distribution

- 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

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

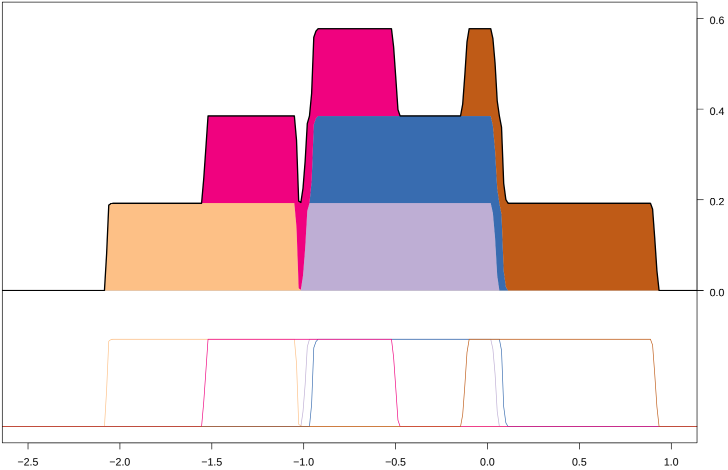
or the nonparametric kernel smoother

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

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

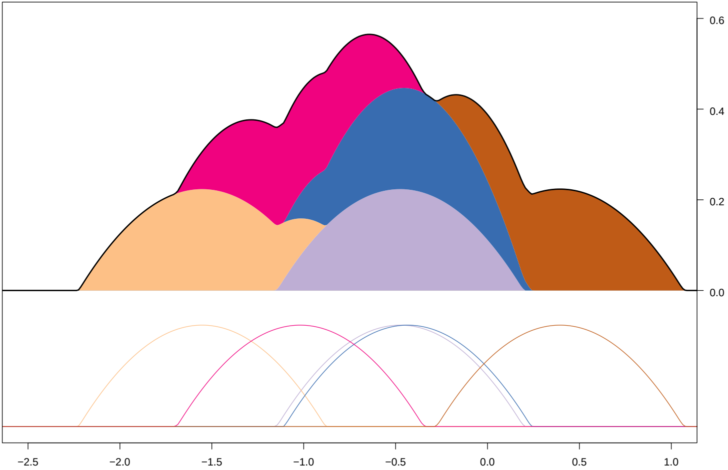
Kernel smoothing with UNI kernel

$x = (-0.475, -1.553, -0.434, -1.019, 0.395)$



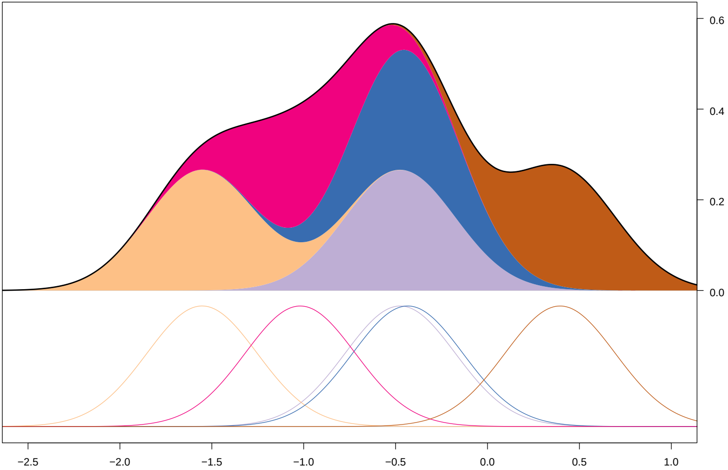
Kernel smoothing with EPA kernel

$$x = (-0.475, -1.553, -0.434, -1.019, 0.395)$$

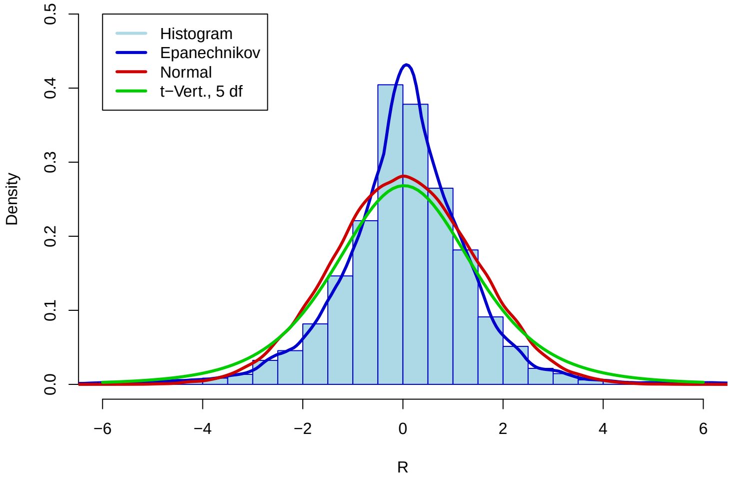


Kernel smoothing with GAU kernel

$$x = (-0.475, -1.553, -0.434, -1.019, 0.395)$$



The estimated density of DAX returns



Multivariate Case

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

Example 2

- $\{x_{1i}, \ldots, x_{di}\}_{i=1,\ldots,n}$ 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

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

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

Correlation

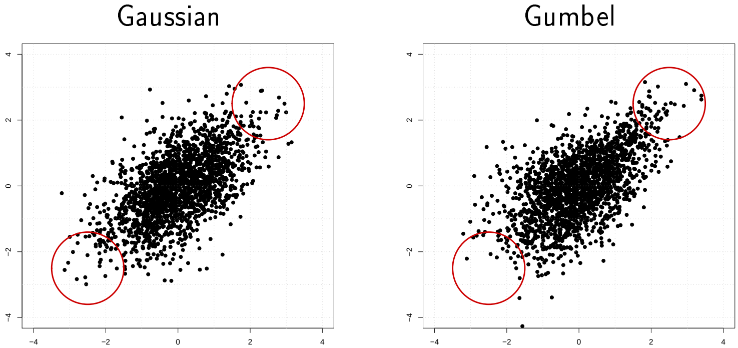


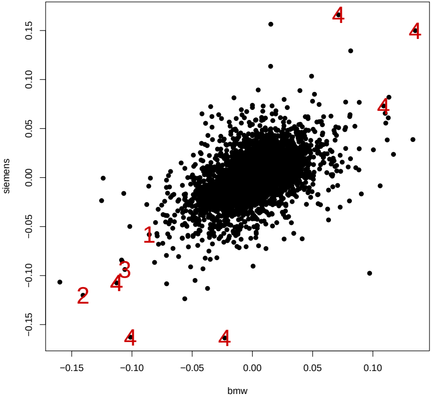
Figure 1: Scatterplots for two distribution with $\rho = 0.4$

- same marginal distributions
- same linear correlation coefficient

"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

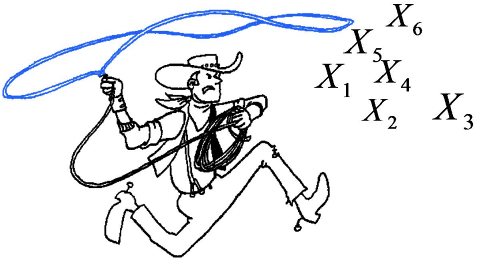


1. 19.10.1987
   Black Monday
2. 16.10.1989
   Berlin Wall
3. 19.08.1991
   Kremlin
4. 17.03.2008, 19.09.2008, 10.10.2008, 13.10.2008, 15.10.2008, 29.10.2008
   Krise

Copula

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

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



A little bit of history

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

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

Wassilij Hoeffding on BBI

A little bit of history

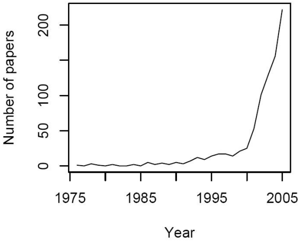
- 1940s: *Wassilij Hoeffding* studies properties of multivariate distributions
- 1959: The word **copula** appears for the first time (*Abe Sklar*)
- 1999: Introduced to financial applications (*Paul Embrechts, Alexander McNeil, Daniel Straumann* in RISK Magazine)
- 2000: Paper by *David Li* in *Journal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

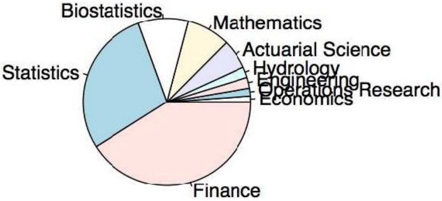
Applications

Practical Use:

1. medicine (Vandenhende (2003))
2. hydrology (Genest and Favre (2006))
3. biometrics (Wang and Wells (2000, JASA), Chen and Fan (2006, CanJoS))
4. economics
   - portfolio selection (Patton (2004, JoFE), Xu (2004, PhD thesis), 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))

Applications





Bourdeau-Brien (2007) covers 871 publications

Special Copulas

Theorem

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

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

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

The simplest copula is **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 copula

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

- for $\theta > 1$ allows to generate dependence in the upper tail
- for $\theta = 1$ reduces to the product copula
- for $\theta \to \infty$ obtain Frèchet-Hoeffding upper bound

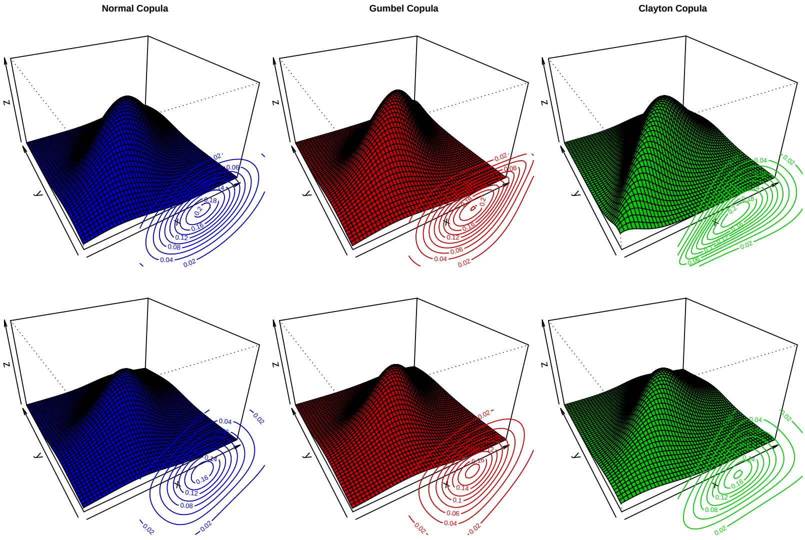
$$C_\theta(u_1, u_2) \stackrel{\theta\to\infty}{\to} \min(u_1, u_2)$$

Copula Examples 3

Clayton copula

$$C_\theta^{Cl}(u_1, u_2) = \left[\max\left(u_1^{-\theta} + u_2^{-\theta} - 1, 0\right)\right]^{-\frac{1}{\theta}}$$

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


Normal Copula
Gumbel Copula
Clayton Copula

Dependencies, Linear Correlation

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

- 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

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

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

- 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: $\delta(X_1, X_2) = \sin\left(\frac{\pi}{2}\tau\right)$

Dependencies, Spearmans's rho

Definition

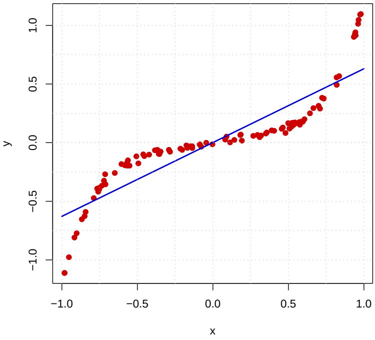
If F is a continuous bivariate cumulative distribution function with marginal F_1 and F_2 and let $(X_1, X_2) \sim F$. Then **Spearmans's rho** is a correlation between $F_1(X_1)$ and $F_2(X_2)$

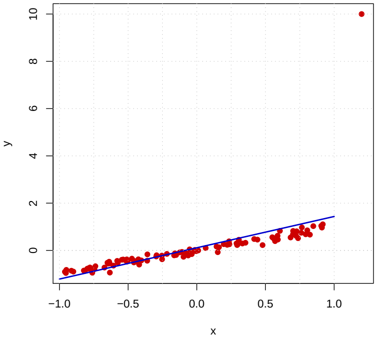
$$\rho = \frac{Cov\{F_1(X_1), F_2(X_2)\}}{\sqrt{Var\{F_1(X_1)\}Var\{F_2(X_2)\}}}.$$

- 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: $\delta(X_1, X_2) = 2\sin\left(\frac{\pi}{6}\rho\right)$

$$\delta = 0.892, \quad \tau = 0.956, \quad \rho = 0.996$$

$$\delta = 0.659, \quad \tau = 0.888, \quad \rho = 0.982$$





Dependencies, Examples

Gaussian copula

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

where δ is a linear correlation coefficient.
Gumbel copula

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

Multivariate Copula Definition

Definition

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

Theorem (Sklar, 1959)

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

$$F(x_1, \ldots, x_k) = C\{F_1(x_1), \ldots, F_k(x_k)\} \tag{1}$$

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

Copula Density

Several theorems provides existence of derivatives of copulas, having them copula density is defined as

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

Joint density function based on copula

$$_cf(x_1, \ldots, x_k) = c\{F_1(x_1), \ldots, F_k(x_k)\} \cdot f_1(x_1) \ldots f_k(x_k),$$

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

Special Copulas

Theorem

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

$$\max\left(\sum_{i=1}^{k} u_i + 1 - k, 0\right) \leq C(u_1, \ldots, u_k) \leq \min(u_1, \ldots, u_k),$$

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

The simplest copula is **product copula**

$$\Pi(u_1, \ldots, u_k) = \prod_{i=1}^{k} u_i$$

characterize the case of independence.

Simulation

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

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

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

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

Estimations: Empirical Copula

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

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

is an empirical copula

Estimation: bivariate case

- based on Kendall's τ estimator

$$\tau_n = \frac{4}{n(n-1)} P_n - 1,$$

where P_n is the number of concordant pairs.
For Gumbel copula $\widehat{\theta}_n = \frac{1}{1-\tau_n}$

- based on Spearman's ρ estimator

$$\rho_n = \frac{\sum_{i=1}^{n} (R_i - \overline{R})^2 (S_i - \overline{S})^2}{\sqrt{\sum_{i=1}^{n} (R_i - \overline{R})^2 \sum_{i=1}^{n} (S_i - \overline{S})^2}},$$

where $(R_i, S_i)\ \forall i = \overline{1, n}$ are pairs of ranks.
For Gaussian Copula $\delta_n = 2 \sin \frac{\pi \rho_n}{6}$

Copula Estimation

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

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

and its density is given by

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

Copula Estimation

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

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

and the corresponding log-likelihood function

$$\begin{aligned}\ell(\vartheta; x_1, \ldots, x_T) &= \sum_{t=1}^{T} \log c\{F_{X_1}(x_{1,t}, \delta_1), \ldots, F_{X_d}(x_{d,t}, \delta_d); \theta\} \\ &+ \sum_{t=1}^{T} \sum_{j=1}^{d} \log f_j(x_{j,t}, \delta_j)\end{aligned}$$

Full Maximum Likelihood (FML)

- FML estimates vector of parameters ϑ in one step through

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

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

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

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

Inference for Margins (IFM)

1. estimate parameters δ_j from the marginal distributions:

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

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

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

3. the estimates $\widehat{\vartheta}_{IFM} = (\widehat{\delta}_1, \ldots, \widehat{\delta}_d, \widehat{\theta})'$ solve

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

4. Advantage: numerically stable

Canonical Maximum Likelihood (CML)

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

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

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

where

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

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

$(X_1, X_2) \sim C_\theta^{Gu}$, with $\theta = 1.5$ and
$F_1 = F_2 = \mathcal{N}(\mu_1, \sigma_1^2) = \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}(0, 1)$

	estimate	std. error
μ_1	0.00365	0.00998
σ_1^2	1.00553	0.00690
μ_2	-0.00106	0.00991
σ_2^2	0.99779	0.00684
θ	1.49632	0.01327

Attractive Features

- A copula describes how the marginals are tied together in the joint distribution
- The joint df is decomposed into the marginal dfs and a copula
- The marginal dfs and the copula can be modelled and estimated separately, independent of each other
- Given a copula, we can obtain many multivariate distributions by selecting different marginal dfs
- The copula is invariant under increasing and continuous transformations