Copula Theory and Applications: Quo Vadis?

Paul Embrechts, Marius Hofert

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

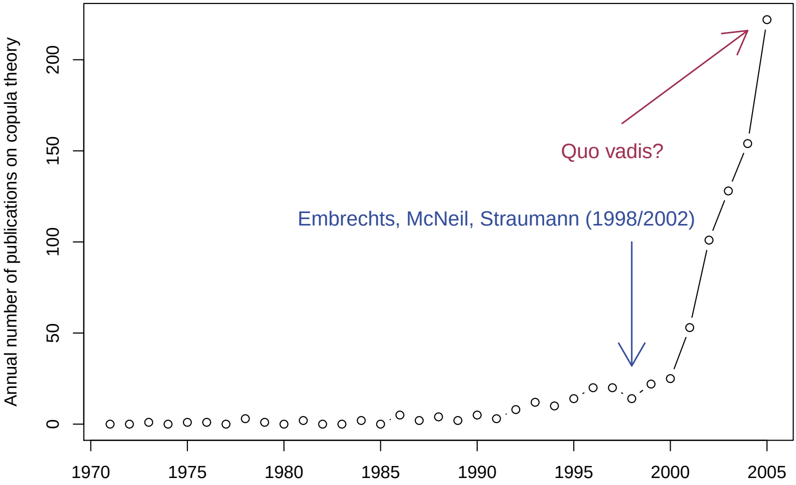
Outline

1 Historical remarks

- Hoeffding (1940): Research on standardized distribution functions ($[-1/2, 1/2]^2$); Féron (1956): on $[0, 1]^3$
- Sklar (1959): term copula $=$ link (linguistics: term linking a subject with a predicate)
- Until 1981: virtually all results obtained in the context of probabilistic metric spaces
- Schweizer and Wolff (1981): Analyze dependence between two or more random variables; invariance principle
- Mid-90s, Embrechts et al. (2002): copulas meet financial and insurance mathematics

A picture is worth a thousand words. . .

Genest et al. (2009):



2 Copulas: An introduction

2.1 Definition

Definition

A copula C is a distribution function with $\mathrm{U}[0,1]$ margins.

Characterization:

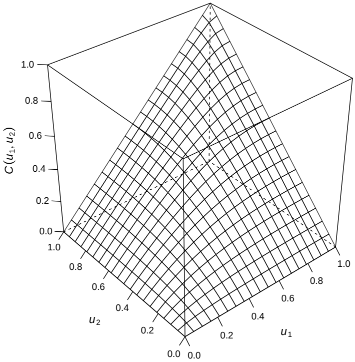
(1) C is grounded;

(2) C has $\mathrm{U}[0,1]$ margins;

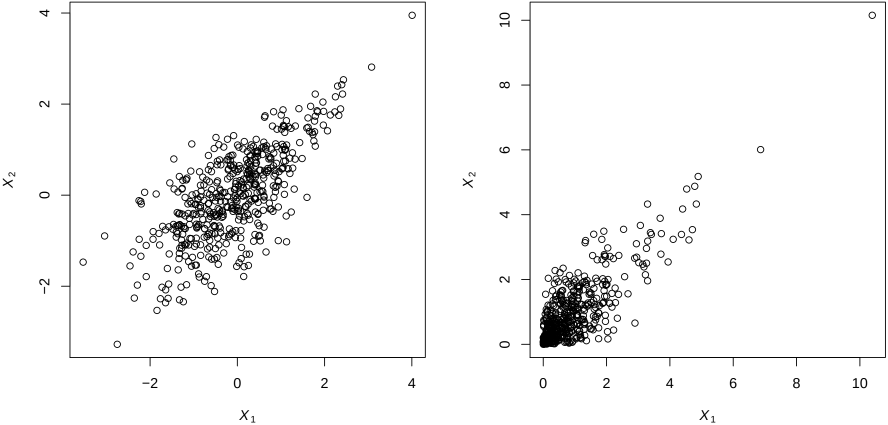
(3) C is d-increasing,

$$\mathbb{P}(\boldsymbol{U} \in (\boldsymbol{a}, \boldsymbol{b}]) = \Delta_{(\boldsymbol{a},\boldsymbol{b}]} C \geq 0.$$

Equivalently (if exists), $c(\boldsymbol{u}) \geq 0$.



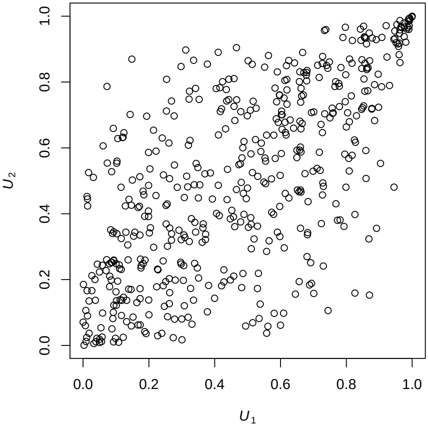
Why standardizing the margins?

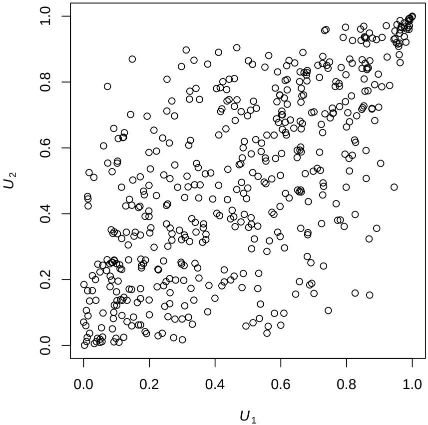


Question: What do these data sets have in common?

Difficult to determine: left: $\mathrm{N}(0,1)$ margins; right: $\mathrm{Exp}(1)$ margins.

Answer: The dependence structure between X_1 and X_2!





Note: Not only equal in distribution, but even the same realizations!
(from a Gumbel copula)

2.2 Sklar's Theorem

Theorem (Sklar (1959))
$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$, $\boldsymbol{x} \in \mathbb{R}^d$

Interpretation in two directions:

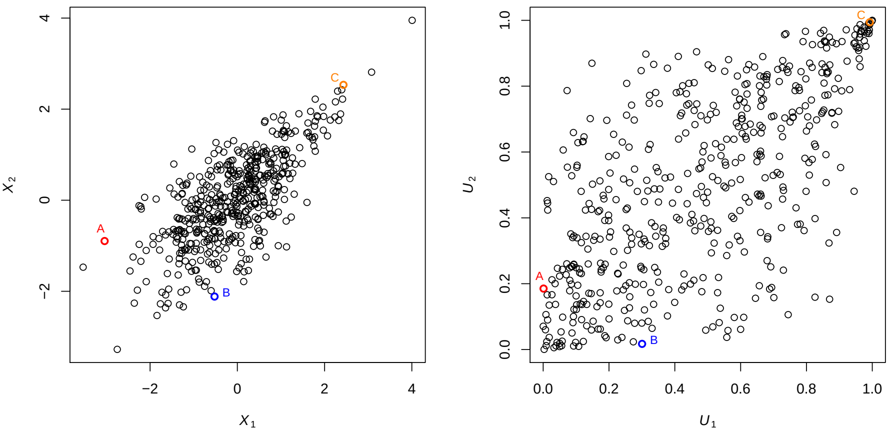
"$\Rightarrow$" Investigate dependencies (estimation, gof)

- $+$ Invariance principle: $(F_1(X_1), \dots, F_d(X_d))^\top \sim C$
- $+$ Numerics: reduce parameter space

"$\Leftarrow$" Construction of distributions (sampling)

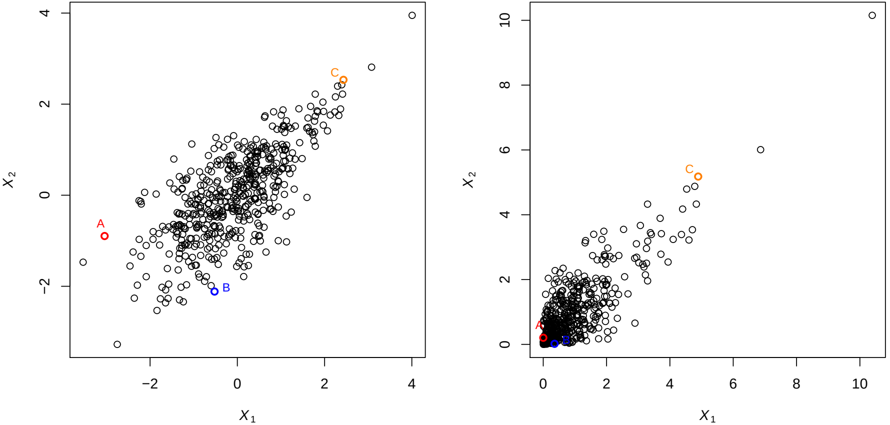
- $+$ Theory: Unifying framework
- $+$ Financial and Insurance Mathematics: Realistic models

Sklar's Theorem (explained graphically)



Note: The ranks of the points remain the same by transforming the margins with increasing transformations!

Sklar's Theorem (explained graphically)



Note: We can adjust the margins as we like while keeping the dependence structure the same!

2.3 Praise and Criticism

Mikosch (2006):

- There is no particular advantage in using copulas – just use any suited model that can be treated statistically;
- Separation of marginals and dependence structure leads to a biased view of stochastic dependence, e.g., when one fits a model to data;
- Various copula models are mostly chosen because of mathematical convenience;
- The class of copulas is too big to be useful;
- Copulas do not contribute to a better understanding of multivariate extremes;
- Copulas do not fit into the existing framework of stochastic processes and time series analysis; they are essentially static models.

Embrechts (2006):

- It is up to us mathematicians to also point at the limitations;
- Queries on two-stage models and on truly dynamic models are pertinent;
- Real finance: two stages – add dependence to marginal models (Sklar); This can help to understand the model risk present;
- Often, there is no hope to obtain a global dynamic model; the data information available does not allow for risk measures (i.e., VaR) to be estimated precisely;
- There are three reasons why copulas are important: pedagogic, pedagogic, and stress testing;
- Stress testing: what range of possibilities exist for risk measures?
- I personally hope that mathematical finance and insurance will turn to truly (high-dimensional) multivariate modelling in QRM.

2.4 Stress testing

(1) "$\Leftarrow$" in Sklar's Theorem:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

(2) Given: $\boldsymbol{X} = (X_1, \dots, X_d)^\top$: risk factors

$\Psi(\boldsymbol{X})$: financial instrument

R: risk (or pricing) measure

Task: Calculate $R(\Psi(\boldsymbol{X}))$ under some assumptions on $\boldsymbol{X}$, Ψ, R.

Typical solution:

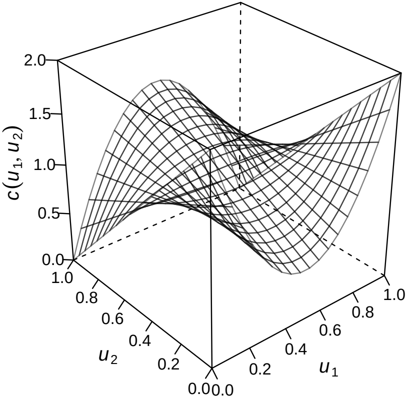
$$R_L \le R(\Psi(\boldsymbol{X})) \le R_U$$

Determine R_L and R_U and prove sharpness!

2.5 Correlation misunderstandings

Misunderstanding 1: F_1, F_2, and ρ determine H

Counter-example: $C(u_1, u_2) = u_1 u_2 (1 - 2\theta(u_1 - \frac{1}{2})(u_1 - 1)(u_2 - 1))$



Density for $\theta = 1$

Properties:

(1) F_1, F_2: $U[0,1]$ (extends to margins with $\mathbb{E}[X^2] < \infty$ and F_1 symmetric about 0)

(2) $\rho = 0$ for all $\theta \in [-1, 1]$

(3) Clearly, $C \neq \Pi$

In particular, $\rho = 0 \not\Rightarrow$ independence!

Reasoning:

Hoeffding's identity

$$\rho = 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (C(F_1(u_1), F_2(u_2)) - F_1(u_1)F_2(u_2))\, du_1 du_2$$

Now consider $\mathrm{U}[0,1]$ margins and

$$C(u_1, u_2) = u_1 u_2 + f_1(u_1) f_2(u_2)$$

with $f_j(0) = f_j(1) = 0$ and $f_1'(u_1) f_2'(u_2) \geq -1$. Then

$$\rho = 12 \int_0^1 f_1(u_1)\, du_1 \int_0^1 f_2(u_2)\, du_2$$

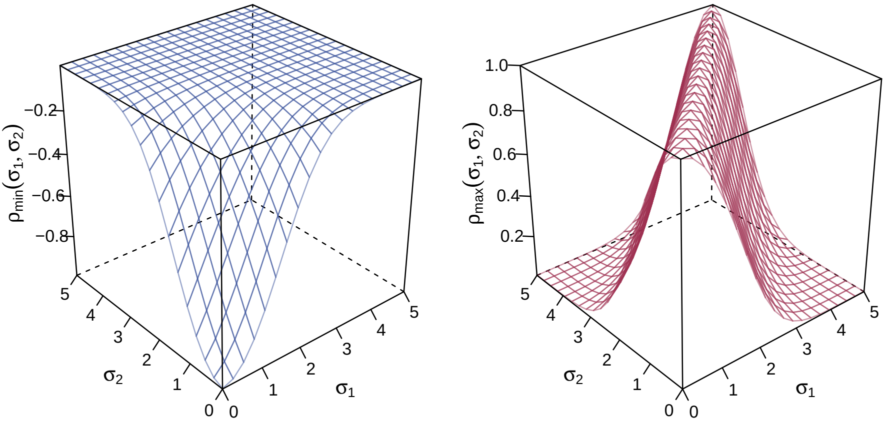
$\Rightarrow$ If f_1 is point symmetric about $1/2$ (as above), then $\rho = 0$.

Hoeffding's identity and the Fréchet-Hoeffding bounds also imply:

$\rho_{\text{min}} \leq \rho \leq \rho_{\text{max}}$ attained for W and M

Misunderstanding 2: Given F_1, F_2, any $\rho \in [-1, 1]$ is attainable

Let $X_j \sim \mathrm{LN}(0, \sigma_j^2)$, $j \in \{1, 2\}$. Then the Hoeffding bounds on ρ are:



Example: For $\sigma_1^2 = 1$, $\sigma_2^2 = 16$: $\rho \in [-0.0003, 0.0137]$!

Further misunderstandings

- $\rho = \rho(X_1, X_2)$ exists for every pair (X_1, X_2) of random variables.
- $\rho(X_1, X_2)$ is invariant under strictly increasing transformations on X_1 or X_2.

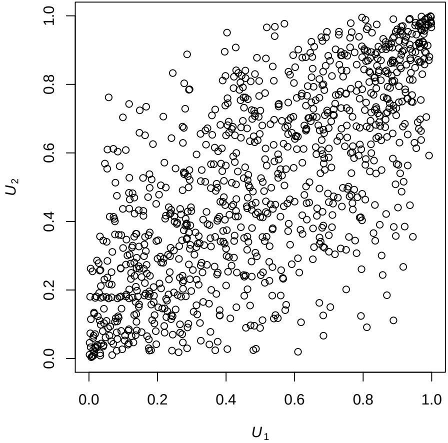
Counter-example: $X_1, X_2 \overset{\text{iid}}{\sim} \mathrm{Par}(3) \;\Rightarrow\; \rho(X_1, X_2) = 0$

but $\rho(X_1^2, X_2)$ does not exist!

Note: Copula-based measures of concordance (e.g., Kendall's tau, Spearman's rho) still cannot solve Misunderstanding 1. In other words, one cannot summarize dependence in one number!

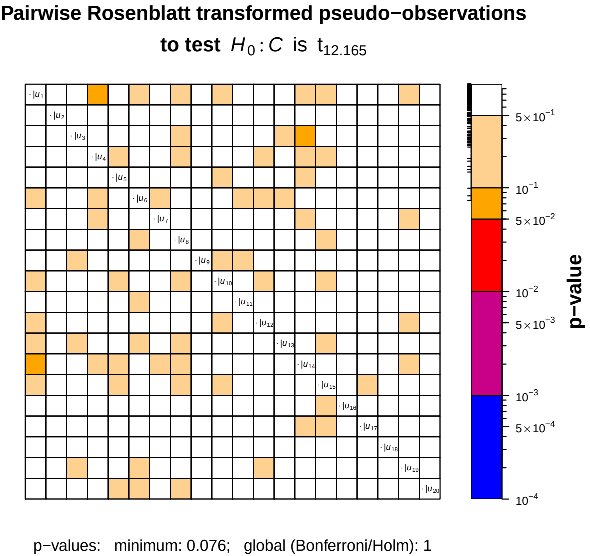
3 Classical copula models

3.1 Elliptical copulas

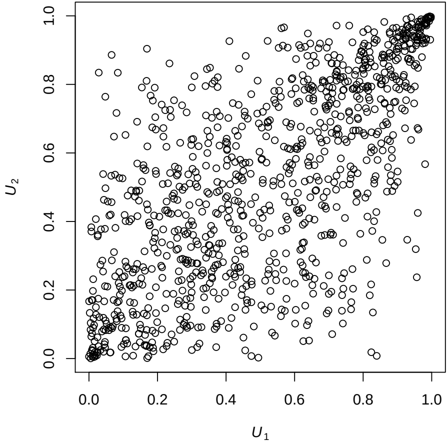


- constructed from Sklar:
  $\boldsymbol{X} = \mu + RA\boldsymbol{U}$
- radially symmetric $\Rightarrow \lambda_L = \lambda_U$
- typically, C not explicit
- density c available
- sampling often simple
- widely used in practice (pairwise thinking in terms of correlation)
- examples: Gaussian, t_ν

SMI daily log-returns from 2011-09-09 to 2012-03-28.



3.2 Archimedean copulas

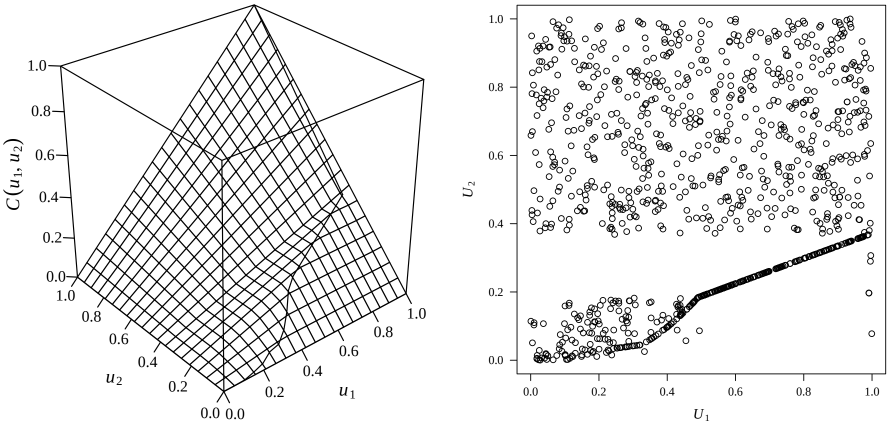


- constructed explicitly via
$$\psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$$
(or with a stoch. representation)
- relevant quantities expressible in the one-dimensional function ψ
- $\lambda_L \neq \lambda_U$ possible
- symmetric
- sampling often simple (MO)
- used in practice
- examples: A, C, F, G, J, opC,...

Recent findings: Densities for A, C, F, G, J, opC, see Hofert et al. (2012)

3.3 Exotic animals in the zoo

Copulas can appear in totally different stochastic contexts, e.g.,. . .



. . . as dependencies of default times in complicated credit default models.

4 Hierarchical models: From $d = 2$ to $d \gg 2$

The popular term “hierarchical” is overloaded! Some use it for

(1) density-based approaches;

(2) copula-based approaches;

(3) approaches based on stochastic representations;

(4) simulated dependencies.

Such dependencies should rather be called...

> *Dependence structures that extend in a more (but not too) flexible way to higher dimensions than their corresponding low-dimensional special cases.*

... but that’s not very practical ☺

4.1 Density-based approach: Pair-copula constructions

- This approach typically works with non-uniform margins;
- It is based on a decomposition of a multivariate density f into conditional densities of lower dimension:

$$f(x_1,\dots,x_d) = f(x_1)\prod_{j=1}^{d} f(x_j \,|\, x_1,\dots,x_{j-1})$$

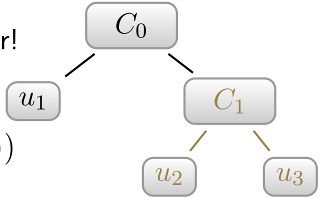
Further decompose the $f(x_j \,|\, x_1,\dots,x_{j-1})$'s via Sklar's Theorem:

$$\begin{aligned} f(x_j \,|\, \boldsymbol{x}_I) &= \frac{f(x_j, x_k \,|\, \boldsymbol{x}_{I\setminus\{k\}})}{f(x_j \,|\, \boldsymbol{x}_{I\setminus\{k\}}) f(x_k \,|\, \boldsymbol{x}_{I\setminus\{k\}})} f(x_j \,|\, \boldsymbol{x}_{I\setminus\{k\}}) \\ &= c_{j,k|I\setminus\{k\}}(F(x_j \,|\, \boldsymbol{x}_{I\setminus\{k\}}), F(x_k \,|\, \boldsymbol{x}_{I\setminus\{k\}})) f(x_j \,|\, \boldsymbol{x}_{I\setminus\{k\}}) \end{aligned}$$

$\Rightarrow$ One obtains a density decomposition into bivariate pieces

- Flexible model, likelihood tractable
- Not all bivariate margins (e.g.) are given explicitly (λ's etc.); error propagation when estimating the model step-wise

4.2 Copula-based approach: Nested Archimedean copulas



Idea: Plug Archimedean copulas into each other!

$$C(\boldsymbol{u}) = C_0(u_1, C_1(u_2, u_3))$$

$$= \psi_0(\psi_0^{-1}(u_1) + \psi_0^{-1}(\psi_1(\psi_1^{-1}(u_2) + \psi_1^{-1}(u_3))))$$

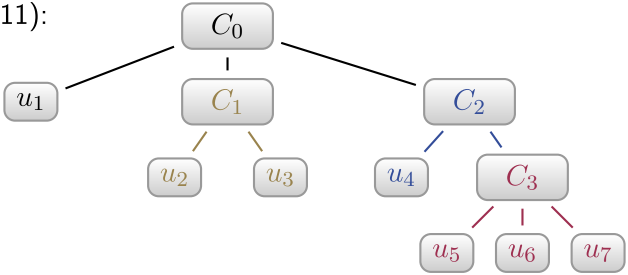
$\Rightarrow$ Asymmetries; not too many parameters;
All lower-dimensional margins known

Question: When is it a copula? Under an assumption on the nodes, e.g.:

Theorem (Joe (1997), McNeil (2008))
$(\psi_0^{-1} \circ \psi_1)'$ completely monotone $\Rightarrow C$ is a copula

Stochastic representation (and sampling)

Hofert (2011):



$$\left(\psi_0\left(\frac{E_1}{V_0}\right), \psi_1\left(\frac{E_2}{V_{01}}\right), \psi_1\left(\frac{E_3}{V_{01}}\right), \psi_2\left(\frac{E_4}{V_{02}}\right), \psi_3\left(\frac{E_5}{V_{23}}\right), \psi_3\left(\frac{E_6}{V_{23}}\right), \psi_3\left(\frac{E_7}{V_{23}}\right)\right)^\top$$

where $V_0 \sim \mathcal{LS}^{-1}[\psi_0]$ $\quad V_{01}|V_0 \sim \mathcal{LS}^{-1}[\psi_{01}(\cdot\,;V_0)]$

$V_{02}|V_0 \sim \mathcal{LS}^{-1}[\psi_{02}(\cdot\,;V_0)]$ $\quad V_{23}|V_{02} \sim \mathcal{LS}^{-1}[\psi_{23}(\cdot\,;V_{02})]$

$\Rightarrow$ R package `copula`

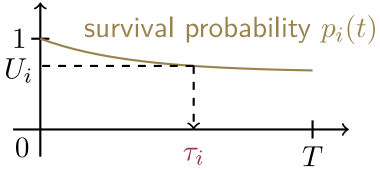
5 Application to Finance: CDO pricing

Goal: Pricing derivatives on large credit portfolios

Intensity-based default model:

$$p_i(t) = \exp\Big(-\int_0^t \lambda_i(s)\,ds\Big)$$

$$\tau_i = \inf\{t \geq 0 : p_i(t) \leq U_i\}$$



Note: $\lambda_U = 0 \Rightarrow$ No joint defaults within short time!

Copulas for the triggers $\boldsymbol{U}$:

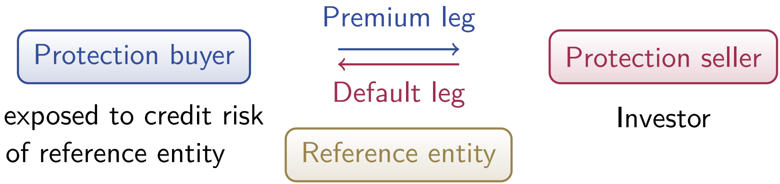
(1) Li (2000): Gaussian ($\lambda_U = 0$)

(2) Schönbucher and Schubert (2001): Archimedean ($\lambda_U > 0$ possible)

(3) Hofert and Scherer (2011): nested Archimedean ($\lambda_U > 0$, hierarchies)

5.1 Towards CDOs: CDS

- CDS = Credit default swap
- Contract of the form:



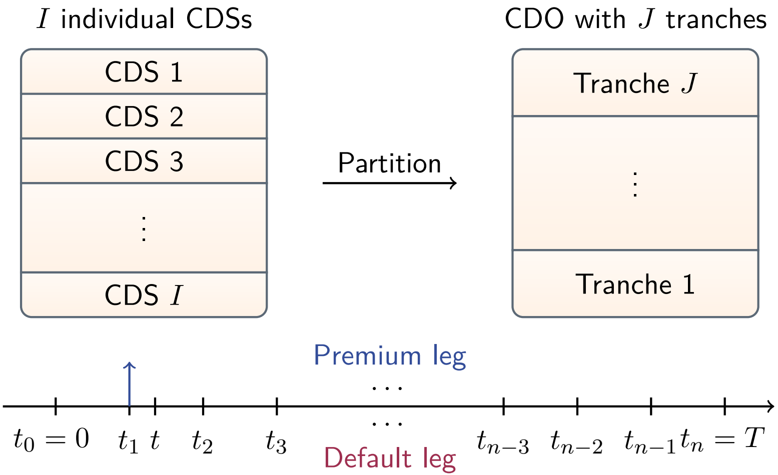
- Pricing problem: Determine the fair premium (“spread”)

Now consider a portfolio of I such contracts.

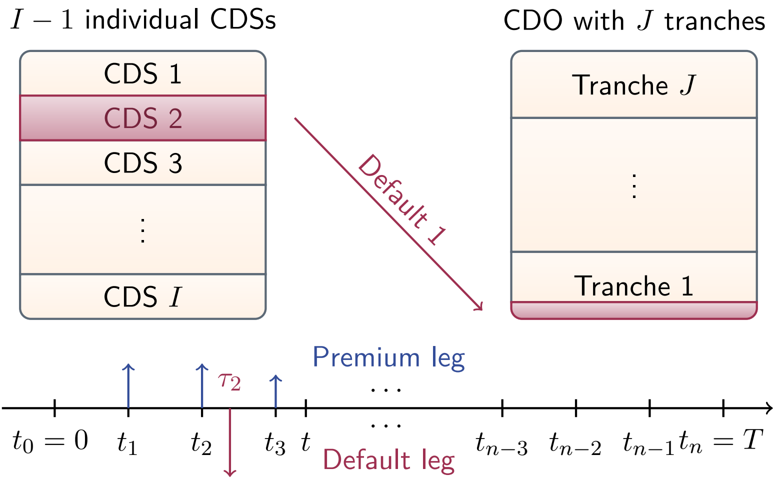
Main idea of a CDO: Partition it into tranches of different seniorities. See also Donnelly and Embrechts (2010).

Again, a picture is worth a thousand words...

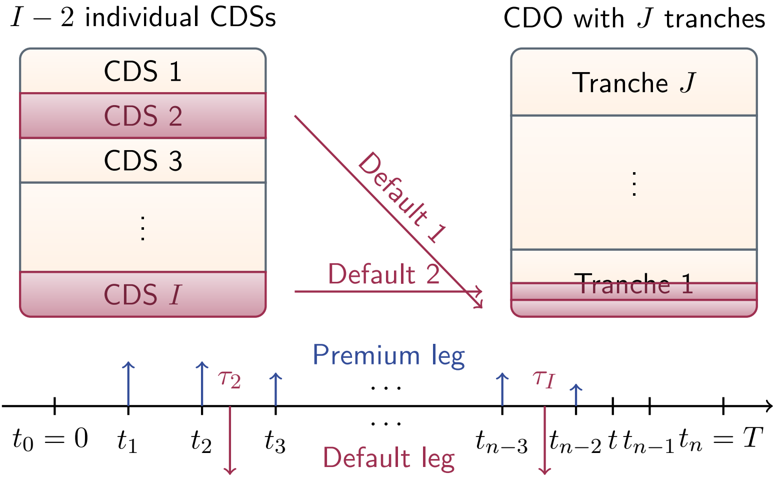
5.2 CDO: Main idea



5.2 CDO: Main idea



5.2 CDO: Main idea



Crucial observations

- Untranched portfolio: The overall loss process is (R recovery rate)

$$L_t = \frac{1-R}{I} \sum_{i=1}^{I} \mathbb{1}_{\{\tau_i \le t\}}$$

$\Rightarrow$ Expected loss $= \mathbb{E}[L_t] = \frac{1-R}{I} \sum_{i=1}^{I} \mathbb{P}(p_i(t) \le U_i)$

$\Rightarrow$ Independent of C. Calibrate to CDS quotes, the "marginals" here.

- Tranched portfolio: The loss affecting tranche j is

$$L_{t,j} = \min\{\max\{0, L_t - l_j\}, u_j - l_j\}$$

$\Rightarrow$ A non-linear functional in the overall loss L_t

$\Rightarrow$ Dependence on C! Calibrate C to CDO quotes (e.g., iTraxx).

Requires fast MC; one spread available for each tranche per day; motivation for nested Archimedean copulas, see Hofert and Scherer (2011).

6 Copulas and Statistics: Problems for $d \gg 2$

Statistics is mainly investigated in $d = 2$. For larger d (besides the theoretical difficulties), there are serious numerical problems.

In short: Copulas meet Numerics for d large!

Task: Evaluate the density of a Gumbel copula.

- General formula (5s): $C(\boldsymbol{u}) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$ implies

$$c(\boldsymbol{u}) = (-1)^d \psi^{(d)}\left(\sum_{j=1}^{d} \psi^{-1}(u_j)\right) \cdot \prod_{j=1}^{d} -(\psi^{-1})'(u_j).$$

  $\Rightarrow$ log-density

- Finding $(-1)^d \psi^{(d)}$ for $\psi(t) = \exp(-t^{1/\theta})$ theoretically (some hours):

$$(-1)^d \psi^{(d)}(t) = \frac{\psi(t)}{t^d} P(t^{1/\theta})$$

  where P is a polynomial with coefficients $a_{dk}(\theta)$ (again polynomials!)

- **Num. Problem 1:** $\log(-1)^d\psi^{(d)}(t) = \log\sum$. Since the sum is typically not in a range representable in computer arithmetic, we can't first compute the sum and then take the log! **Idea:** intelligent log:

$$\begin{aligned}\log\sum_{i=1}^n x_i &= \log\sum_{i=1}^n \exp(b_i), \quad b_i = \log x_i\\ &= \log\Big(\exp(b_{\mathsf{max}})\sum_{i=1}^n \exp(b_i - b_{\mathsf{max}})\Big)\\ &= b_{\mathsf{max}} + \log\sum_{i=1}^n \exp(b_i - b_{\mathsf{max}})\end{aligned}$$

  This can be adapted to our setup where $x_i \le 0$ for some i.

- Careful implementation of 8 methods for evaluation, checks,... (several weeks/months).

- **Num. Problem 2:** Checks are particularly difficult since CASs fail!

Example: $\psi^{(50)}(15) = ?$ for $\theta = 5/4$ (correct answer: 1056.93...)

- Maple 14: 10 628, -29 800,... (chaotic!; sign wrong; slow)
- Mathematica 8: – (aborted after 10 min)
- MATLAB 7.11.0: ✓ ($d = 100$: aborted after several min)
- Sage 4.7.1: – (aborted after 10 min)

Remark: Automatic differentation might provide a solution.

- **Note:** This is only one evaluation! It has to be done...
  - $n (= 100)$ times for computing the log-likelihood once
  - $m (= 10)$ times for computing MLEs
  - $N (= 1000)$ times within a bootstrap
  - $M (= 200)$ times to (num.) show bootstrap convergence
  - for various n, d, θ...

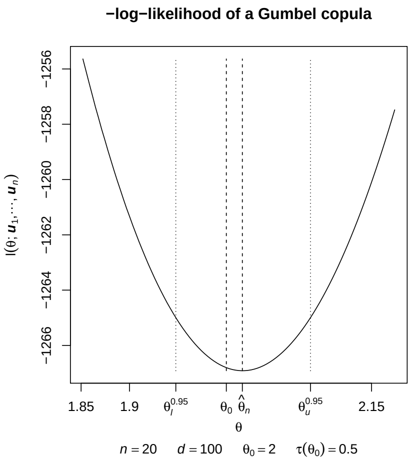
$\Rightarrow$ Parallel computing required; still (!) run time matters...

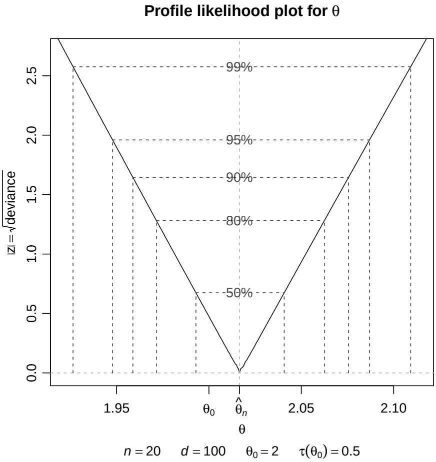
Will anyone care about this in 20 years?

Likely answer: Yes

- Useful formulas can not always be obtained from CASs;
- Even if you get a formula, it might not be in a numerically stable form;
- Computations go wrong every day, people do not seem to care about them too much $\Rightarrow$ operational risk!
- Careful checks have to be made (often not acknowledged);
- (Modern) mathematicians should be (more!) aware of these issues;
- There are many more of the above problems with serious consequences to statistics for copulas in large dimensions.

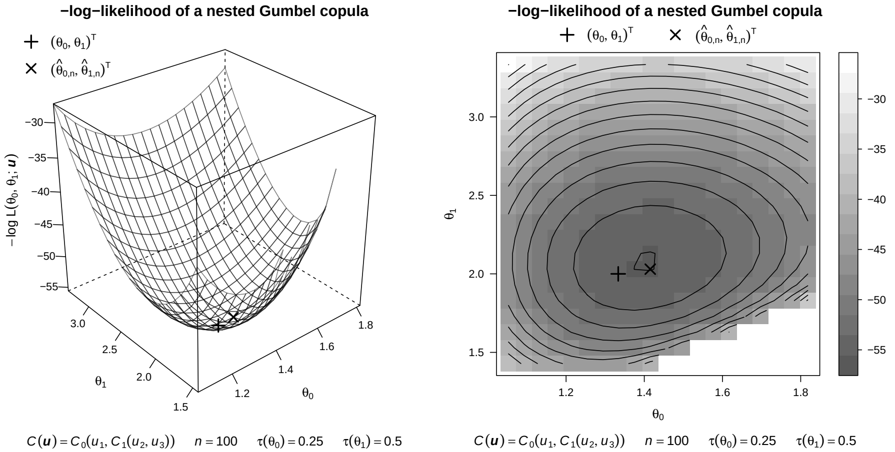
If you pay attention to the numerical issues...





Striking result: (Archimedean copulas; all families, all τ): MSE $\propto \frac{1}{nd}$

-log-likelihood of a nested Gumbel copula $C(\boldsymbol{u}) = C_0(u_1, C_1(u_2, u_3))$:



References

Donnelly, C. and Embrechts, P. (2010), The devil is in the tails: actuarial mathematics and the subprime mortgage crisis, *ASTIN Bulletin*, 40.1, 1–33.

Embrechts, P., McNeil, A. J., and Straumann, D. (2002), Correlation and Dependency in Risk Management: Properties and Pitfalls, *Risk Management: Value at Risk and Beyond*, ed. by Dempster, M., Cambridge University Press, 176–223.

Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009), The Advent of Copulas in Finance, *The European Journal of Finance*, 15, 609–618.

Thank you for your attention

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich