Copulas ? What copulas ?

R. Chicheportiche & J.P. Bouchaud, CFM

Multivariate linear correlations

- Standard tool in risk management/portfolio optimisation: the covariance matrix $R_{ij} = \langle r_i r_j \rangle$

- Find the portfolio with maximum expected return for a given risk or equivalently, minimum risk for a given return ($\mathcal{G}$)

- In matrix notation:

$$\mathbf{w} = \mathcal{G} \frac{\mathbf{R}^{-1}\mathbf{g}}{\mathbf{g}^T \mathbf{R}^{-1}\mathbf{g}}$$

  where all gains are measured with respect to the risk-free rate and $\sigma_i = 1$ (absorbed in g_i).

- More explicitely:

$$\mathbf{w} \propto \sum_\alpha \lambda_\alpha^{-1} \left(\mathbf{\Psi}_\alpha \cdot \mathbf{g}\right) \mathbf{\Psi}_\alpha = \mathbf{g} + \sum_\alpha (\lambda_\alpha^{-1} - 1) \left(\mathbf{\Psi}_\alpha \cdot \mathbf{g}\right) \mathbf{\Psi}_\alpha$$

Multivariate non-linear correlations

- Many situations in finance in fact require knowledge of higher order correlations
  - Gamma-risk of option portfolios: $\langle r_i^2 r_j^2 \rangle - \langle r_i^2 \rangle \langle r_j^2 \rangle$
  - Stress test of complex porfolios: correlations in extreme market conditions
  - Correlated default probabilities – Credit Derivatives (CDOs, basket of CDSs)

    "The Formula That Killed Wall Street" (Felix Salmon)

Different correlation coefficients

- Correlation coefficient: $\rho_{ij} = \text{cov}(r_i, r_j)/\sqrt{V(r_i)V(r_j)}$

- Correlation of squares or absolute values:

$$\rho_{ij}^{(2)} = \frac{\text{cov}(r_i^2, r_j^2)}{\sqrt{V(r_i^2)V(r_j^2)}} \qquad \rho_{ij}^{(a)} = \frac{\text{cov}(|r_i|, |r_j|)}{\sqrt{V(|r_i|)V(|r_j|)}}$$

- Tail correlation:

$$\tau_{ij}^{UU}(p) = \frac{1}{p}\text{Prob.}\left[r_i > \mathcal{P}_{>,i}^{-1}(p) \bigcap r_j > \mathcal{P}_{>,j}^{-1}(p)\right]$$

(Similar defs. for $\tau^{LL}, \tau^{UL}, \tau^{LU}$)

Copulas

- Sklar's theorem: any multivariate distribution can be "factorized" into

  - its marginals $\mathcal{P}_i \to u_i = \mathcal{P}_i(r_i)$ are $U[0,1]$

  - a "copula", that describes the correlation structure between N $U[0,1]$ standardized random variables: $c(u_1, u_2, ...u_N)$

- All correlations, linear and non linear, can be computed from the copula and the marginals

- For bivariate distributions:

$$C_{ij}(u,v) = \mathbb{P}\left[\mathcal{P}_{<,i}(X_i) \leq u \text{ and } \mathcal{P}_{<,j}(X_j) \leq v\right]$$

Copulas – Examples

- Examples: $(N = 2)$
  - The Gaussian copula: r_1, r_2 bivariate Gaussian $\rightarrow$ defines the Gaussian copula $c_G(u, v|\rho)$
  - The Student copula: r_1, r_2 bivariate Student with tail ν $\rightarrow$ defines the Student copula $c_S(u, v|\rho, \nu)$
  - Archimedean copulas: $\phi(u) : [0, 1] \rightarrow [0, 1]$, $\phi(1) = 0$, ϕ^{-1} decreasing, completely monotone

$$C_A(u, v) = \phi^{-1}\left[\phi(u) + \phi(v)\right]$$

Ex: Frank copulas, $\phi(u) = \ln[e^{\theta} - 1] - \ln[e^{\theta u} - 1]$;
Gumbel copulas, $\phi(u) = (-\ln u)^{\theta}$, $\theta < 1$.

The Copula red-herring

- Sklar's theorem: a nearly empty shell – almost any $c(u_1, u_2, ...u_N)$ with required properties is allowed.

- The usual financial mathematics syndrom: choose a class of copulas – sometimes absurd – with convenient mathematical properties and brute force calibrate to data

- If something fits it can't be bad (??) Statistical tests are not enough – intuition & plausible interpretation are required

- *But he does not wear any clothes!* – see related comments by Thomas Mikosch

The Copula red-herring

- Example 1: why on earth choose the Gaussian copula to describe correlation between (positive) default times???

- Example 2: Archimedean copulas: take two $U[0,1]$ random variables s, w. Set $t = K^{-1}(w)$ with $K(t) = t - \phi(t)/\phi'(t)$.

  $$u = \phi^{-1}\left[s\phi(t)\right]; \qquad v = \phi^{-1}\left[(1-s)\phi(t)\right]; \qquad \longrightarrow r_1, r_2$$

  Financial interpretation ???

- Models should reflect some plausible underlying structure or mechanism

Copulas ? What copulas ?

- Aim of this work
  - Develop intuition around copulas
  - Identify empirical stylized facts about multivariate correlations that copulas should reproduce
  - Discuss “self-copulas” as a tool to study empirical temporal dependences
  - Propose an intuitively motivated, versatile model to generate a wide class of non-linear correlations

Copulas

- Restricted information on copula: diagonal $C(p,p)$ and anti-diagonal $C(p,1-p)$. Note: $C(\frac{1}{2},\frac{1}{2})$ is the probability that both variables are simultaneously below their medians

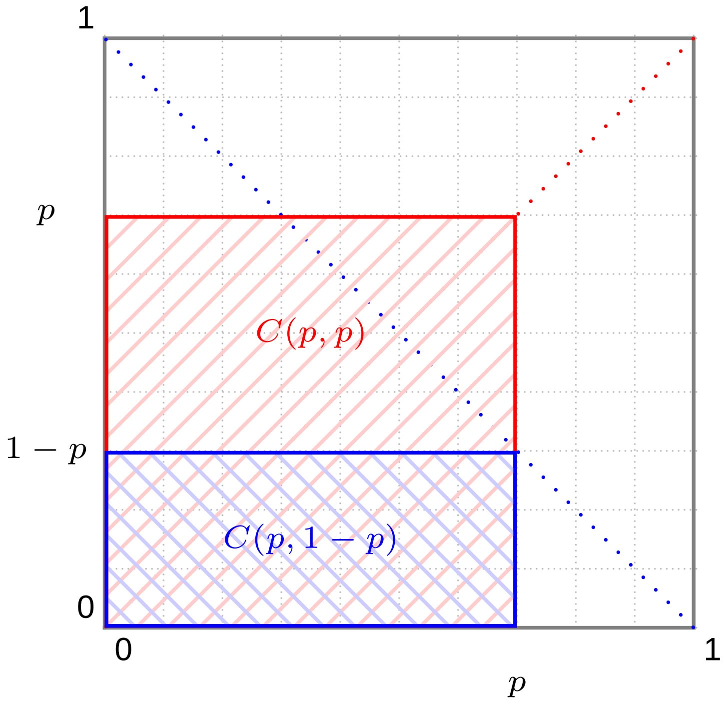
- Tail dependence:

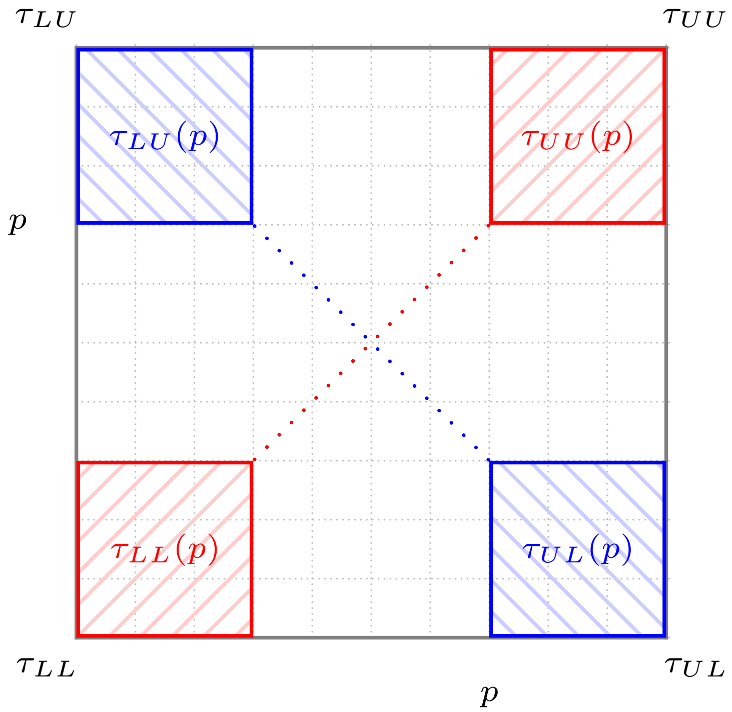
$$\tau^{UU}(p) = \frac{1-2p+C(p,p)}{1-p}, \qquad \text{etc.}$$

- Relative difference with respect to independence or to Gaussian:

$$\frac{C(p,p)-p^2}{p(1-p)} = \tau^{UU}(p)+\tau^{LL}(1-p)-1, \quad \text{or} \quad \frac{C(p,p)-C_G(p,p)}{p(1-p)}$$

Copulas





Student Copulas

- Intuition: $r_1 = \sigma\epsilon_1$, $r_2 = \sigma\epsilon_2$ with:
  - $\epsilon_{1,2}$ bivariate Gaussian with correlation ρ
  - σ is a *common* random volatility with distribution $P(\sigma) = \mathcal{N} e^{-\sigma_0^2/\sigma^2}/\sigma^{1+\nu}$
- The monovariate distributions of $r_{1,2}$ are Student with a power law tail exponent ν ($\in [3, 5]$ for daily data)
- The multivariate Student is a model of correlated Gaussian variables with a common random volatility:

$$r_i = \sigma\epsilon_i \qquad \rho_{ij} = \mathrm{cov}(\epsilon_i, \epsilon_j)$$

Student Copulas

- In this model, all higher-order correlations can be expressed in terms of ρ

- Explicit formulas: $(f_n = \langle \sigma^{2n} \rangle / \langle \sigma^n \rangle^2)$

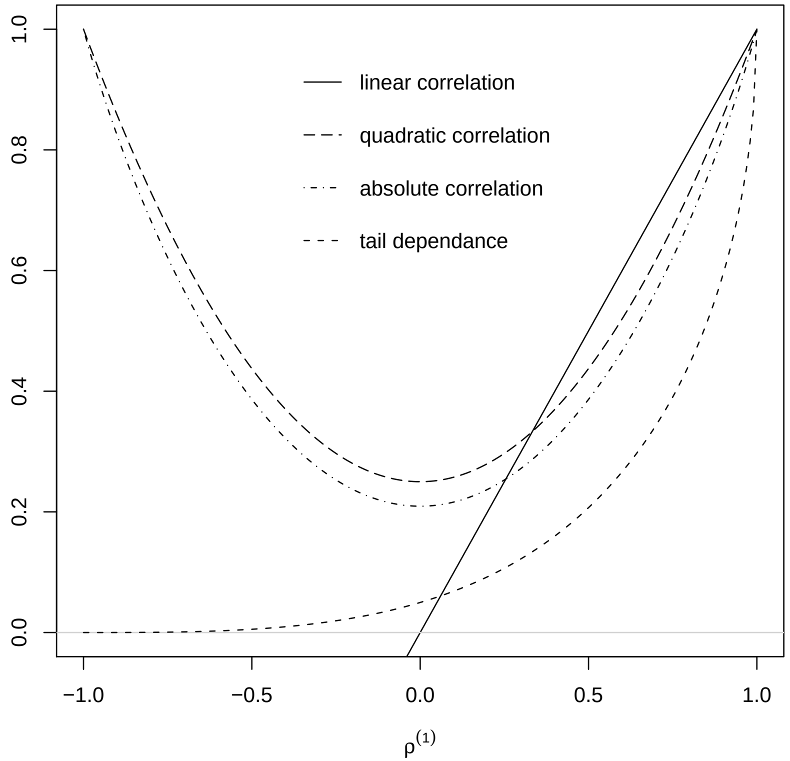
$$\rho^{(2)} = \frac{f_2(1 + 2\rho^2) - 1}{3 f_2 - 1}; \qquad \rho^a = \frac{f_1(\sqrt{1-\rho^2} + \rho \arcsin \rho) - 1}{\frac{\pi}{2} f_1 - 1}.$$

- The tail correlations τ have a finite limit when $p \to 0$ because of the common volatility

- The central point of the copula:

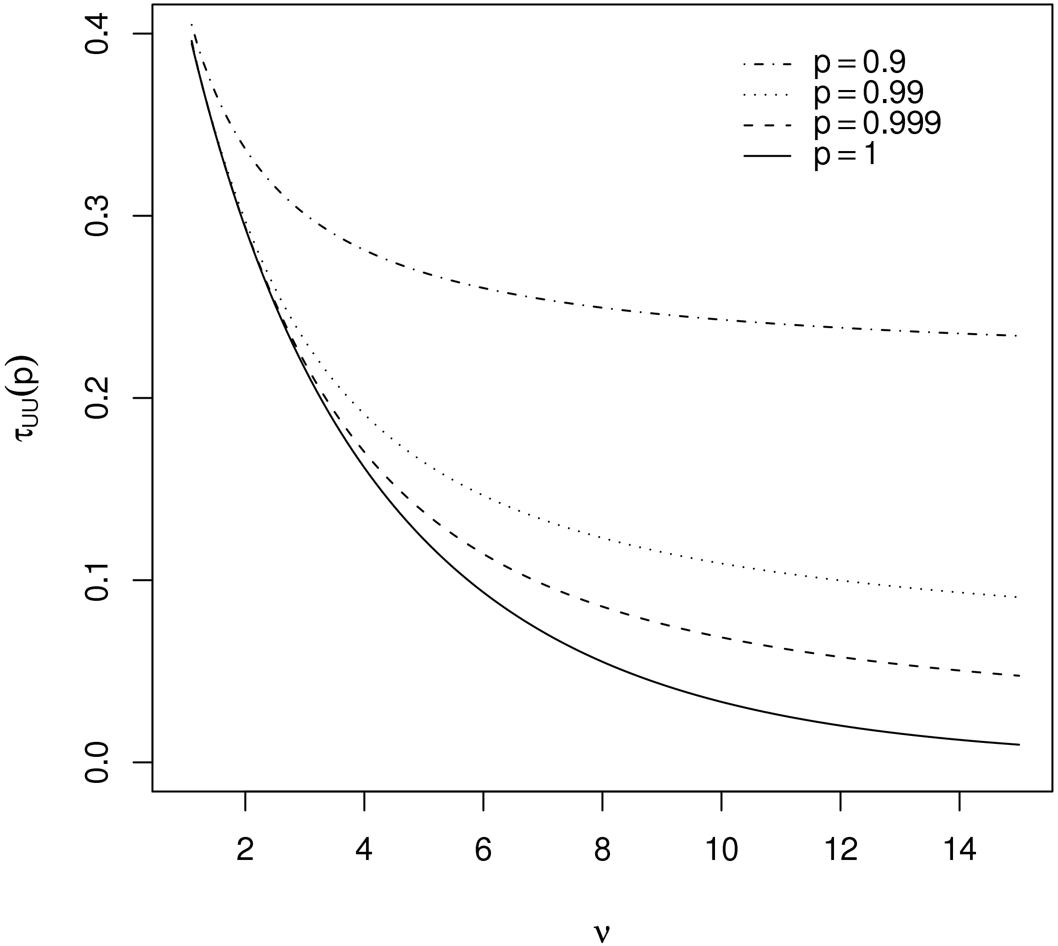
$$C(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho$$

Student Copulas



$$\nu = 5$$

Student Copulas



$\rho = 0.3$ – Note: corrections are of order $(1 - p)^{2/\nu}$

Elliptic Copulas

- A straight-forward generalisation: elliptic copulas $r_1 = \sigma\epsilon_1$, $r_2 = \sigma\epsilon_2$, $P(\sigma)$ arbitrary

- The above formulas remain valid for arbitrary $P(\sigma)$ in particular:

$$C(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho$$

- The tail correlations τ have a finite limit whenever $P(\sigma)$ decays as a power-law

- A relevant example: the log-normal model $\sigma = \sigma_0 e^{\xi}$, $\xi = N(0, \lambda^2)$ – very similar to Student with $\nu \sim \lambda^{-2}$ *

*Although the true asymptotic value of $\tau(p = 0)$ is zero.

Student Copulas and empirical data

- The empirical curves $\rho^a(\rho)$ or $\rho^{(2)}(\rho)$ cross the set of Student predictions, as if "more Gaussian" for small ρ's

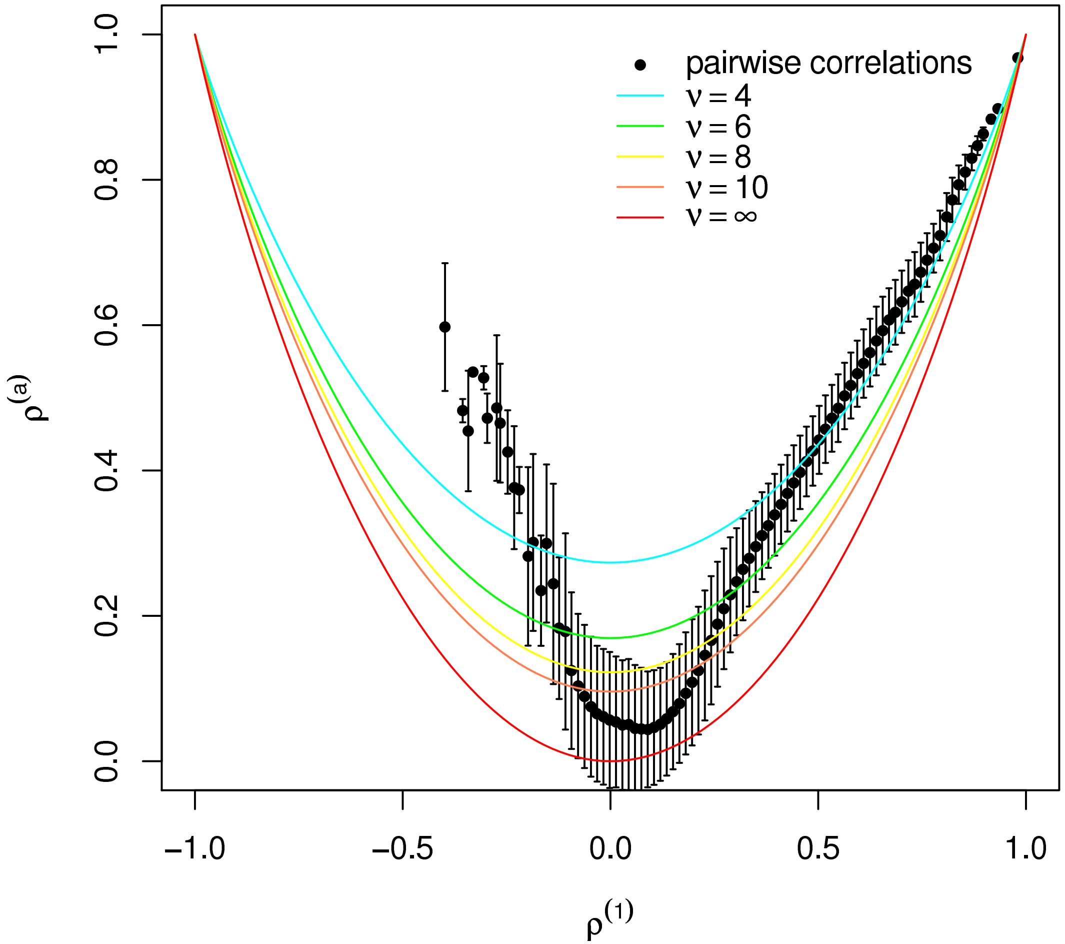
- Same with tail correlation coefficients (+ some level of assymetry)

- $C(\frac{1}{2}, \frac{1}{2})$ systematically different from Elliptic prediction $= \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho$ – in particular $C(\frac{1}{2}, \frac{1}{2} | \rho = 0) > \frac{1}{4}$

- $C(p,p) - C_G(p,p)$ incompatible with a Student model: concave for $\rho < 0.25$ becoming convex for $\rho > 0.25$

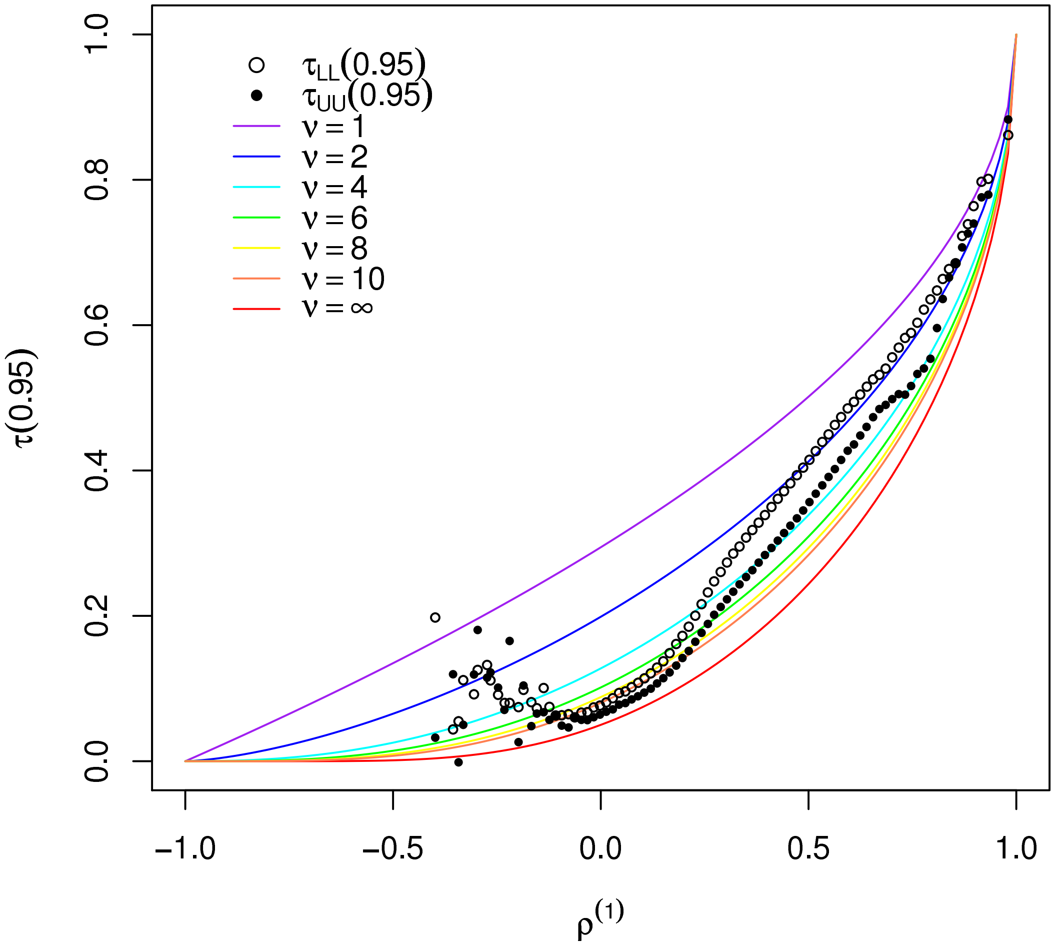
- To be sure: Archimedean copulas are even worse !

Absolute correlation



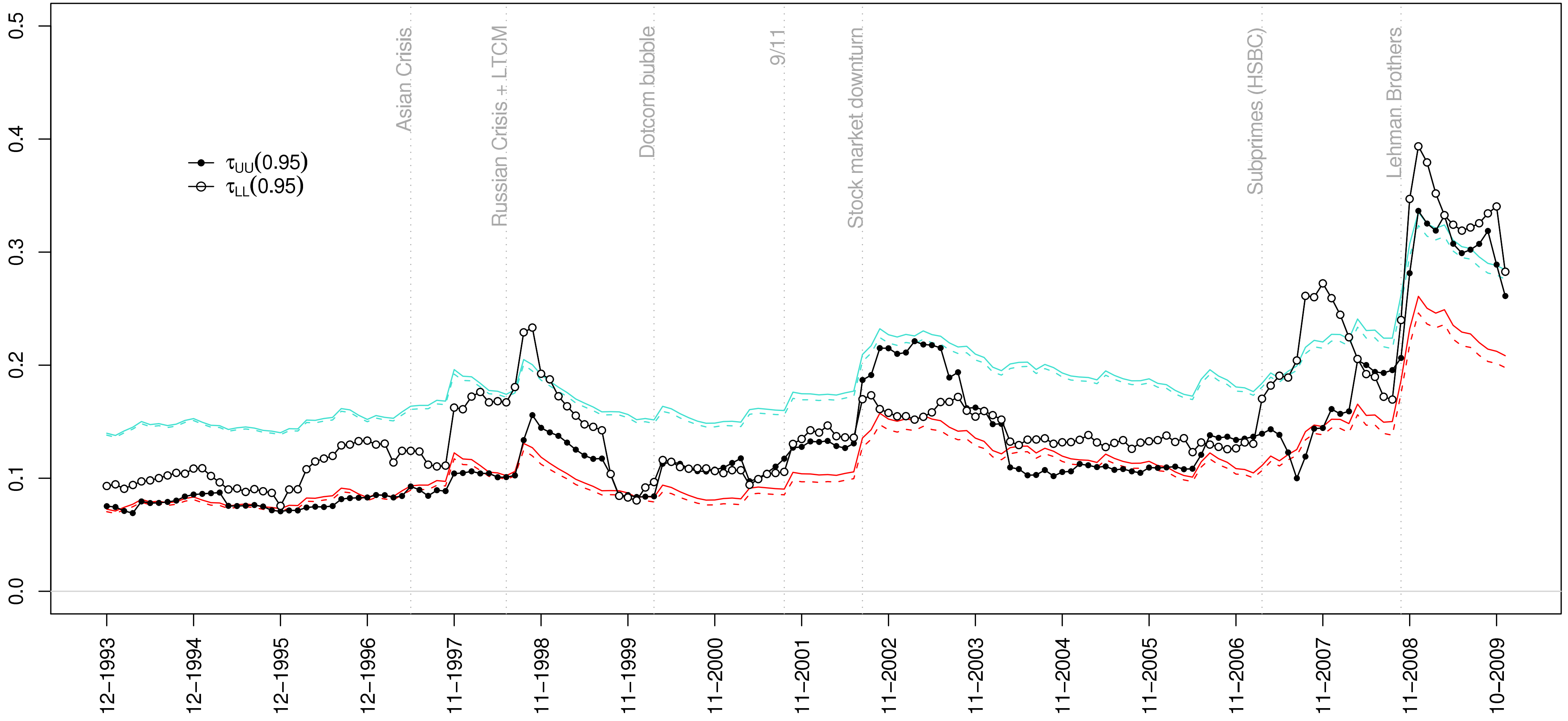
2005-2009

Tail correlation

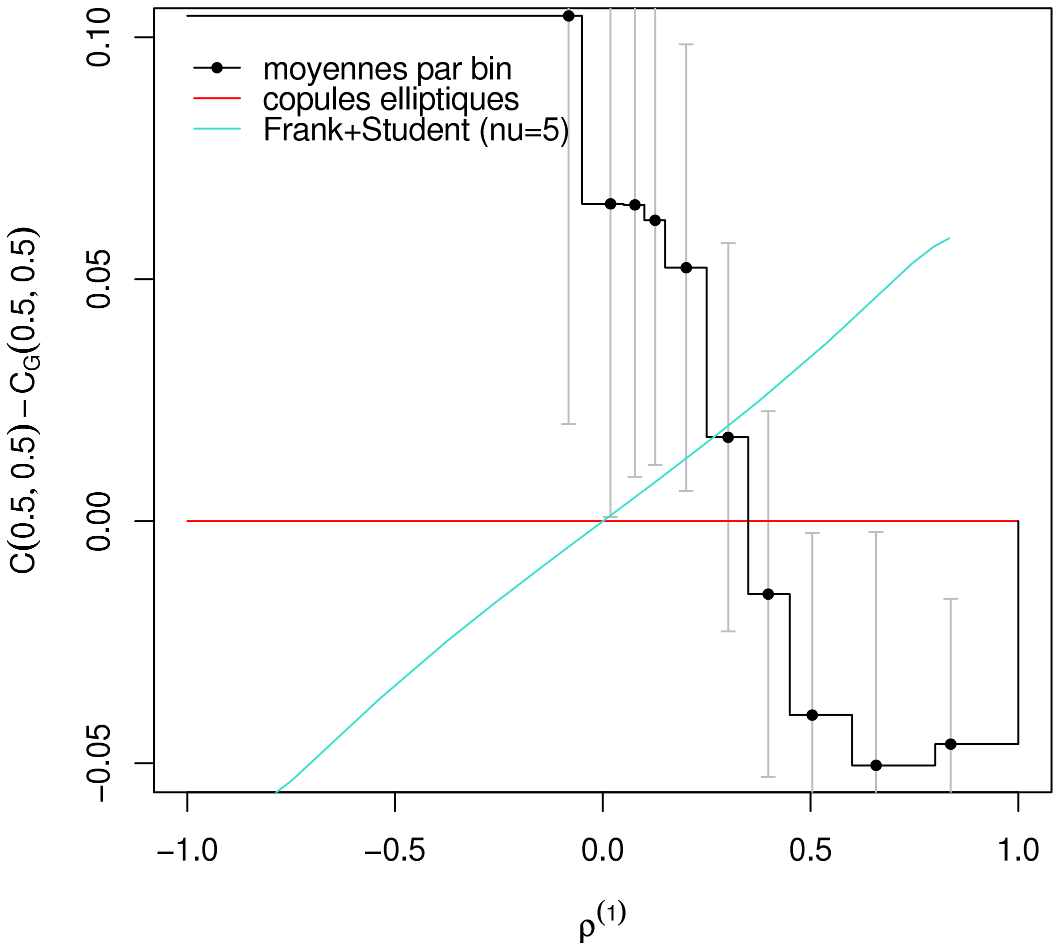


2005-2009

Tail correlation: time series

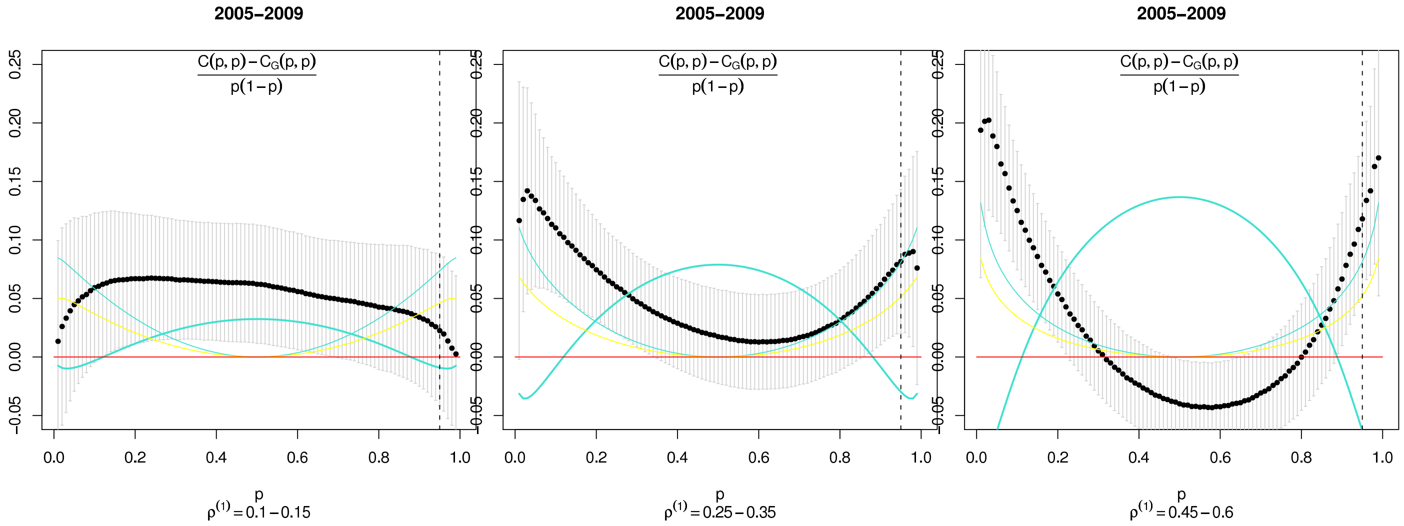


Centre point vs ρ



Difference between empirical results and Student (Frank) prediction for $C(\frac{1}{2}, \frac{1}{2})$

Diagonal



$$\rho = 0.1, 0.3, 0.5$$

Student Copulas: Conclusion

- Student (or even elliptic) copulas are not sufficient to describe the multivariate distribution of stocks!

- Obvious intuitive reason: one expects more than one volatility factor to affect stocks

- How to describe an entangled correlation between returns and volatilities?

- In particular, any model such that $r_i = \sigma_i \epsilon_i$ with correlated random σ's leads to $C(\frac{1}{2}, \frac{1}{2}) \equiv \frac{1}{4}$ for $\rho = 0$!

Constructing a realistic copula model

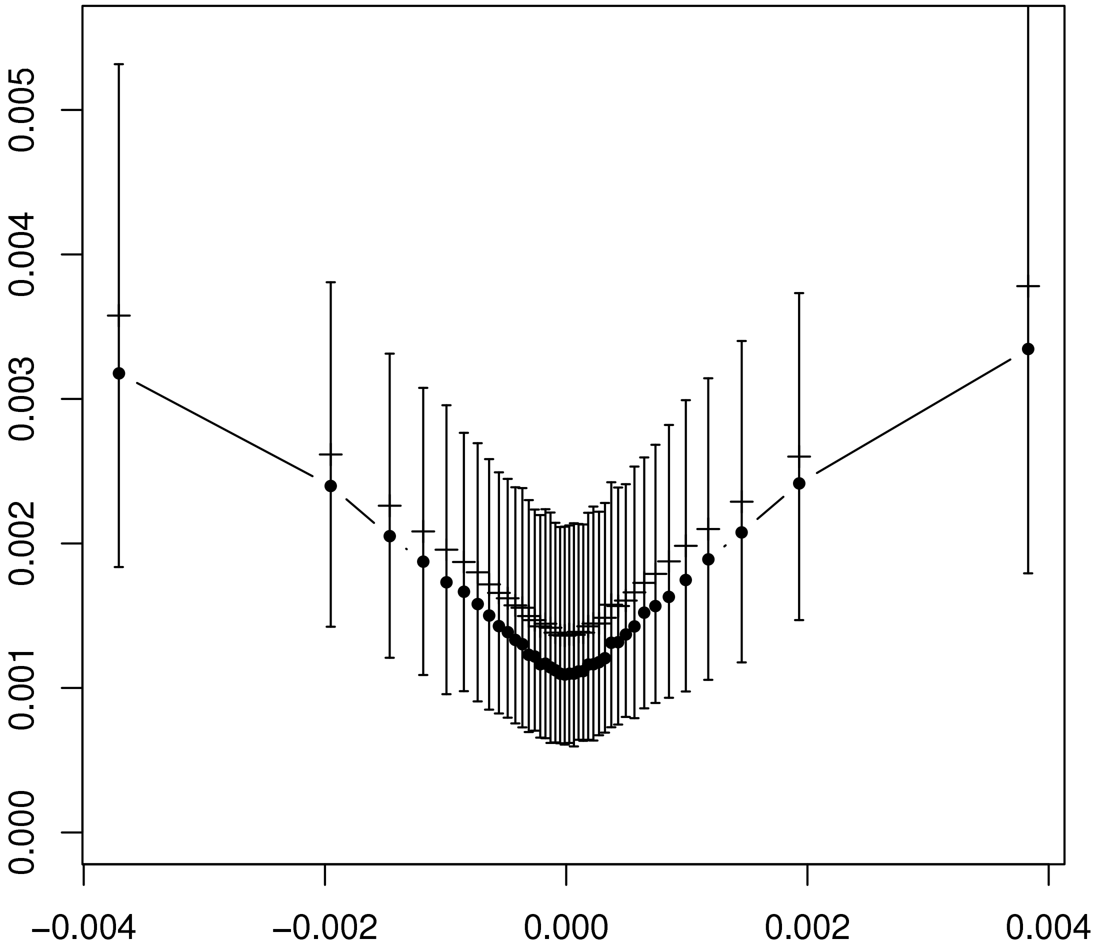
- How do we go about now (for stocks) ?
  - a) stocks are sensitive to "factors"
  - b) factors are hierarchical, in the sense that the vol of the market influences that of sectors, which in turn influence that of more idiosyncratic factors

- Empirical fact: within a one-factor model,

$$r_i = \beta_i \varepsilon_0 + \varepsilon_i$$

  volatility of residuals increases with that of the market ε_0

“Entangled” volatilities



with Romain Allez

Constructing a realistic copula model

- An entangled one-factor model

$$r_i = \beta_i \sigma_0 e^{\xi_0} \varepsilon_0 + \sigma_1 e^{\alpha \xi_0 + \xi_i} \varepsilon_i$$

with $\xi_0 \sim N(0, s_0^2)$, $\xi_i \sim N(0, s_1^2)$, IID,

- The volatility of the idiosyncratic factor is clearly affected by that of the market mode

- Kurtosis of the market factor and of the idiosyncratic factor:

$$\kappa_0 = e^{4s_0^2}\left[e^{4s_0^2} - 1\right]; \qquad \kappa_1 = e^{4(\alpha^2 s_0^2 + s_1^2)}\left[e^{4(\alpha^2 s_0^2 + s_1^2)} - 1\right]$$

Constructing a realistic copula model

- An interesting remark: take two stocks with opposite exposure to the second factor

$$r_{\pm} = \sigma_0 e^{\xi_0} \varepsilon_0 \pm \sigma_1 e^{\alpha \xi_0 + \xi_1} \varepsilon_1$$

- Choose parameters such that volatilities are equal

$$\sigma_0 e^{s_0^2} = \sigma_1 e^{\alpha^2 s_0^2 + s_1^2}$$

such that $\mathrm{cov}(r_+, r_-) = 0$

- Then:

$$C(\frac{1}{2}, \frac{1}{2} | \rho = 0) \approx \frac{1}{4} \left(1 + \frac{\kappa_1 - \kappa_0}{6\pi} \right)$$

A hierachical tree model

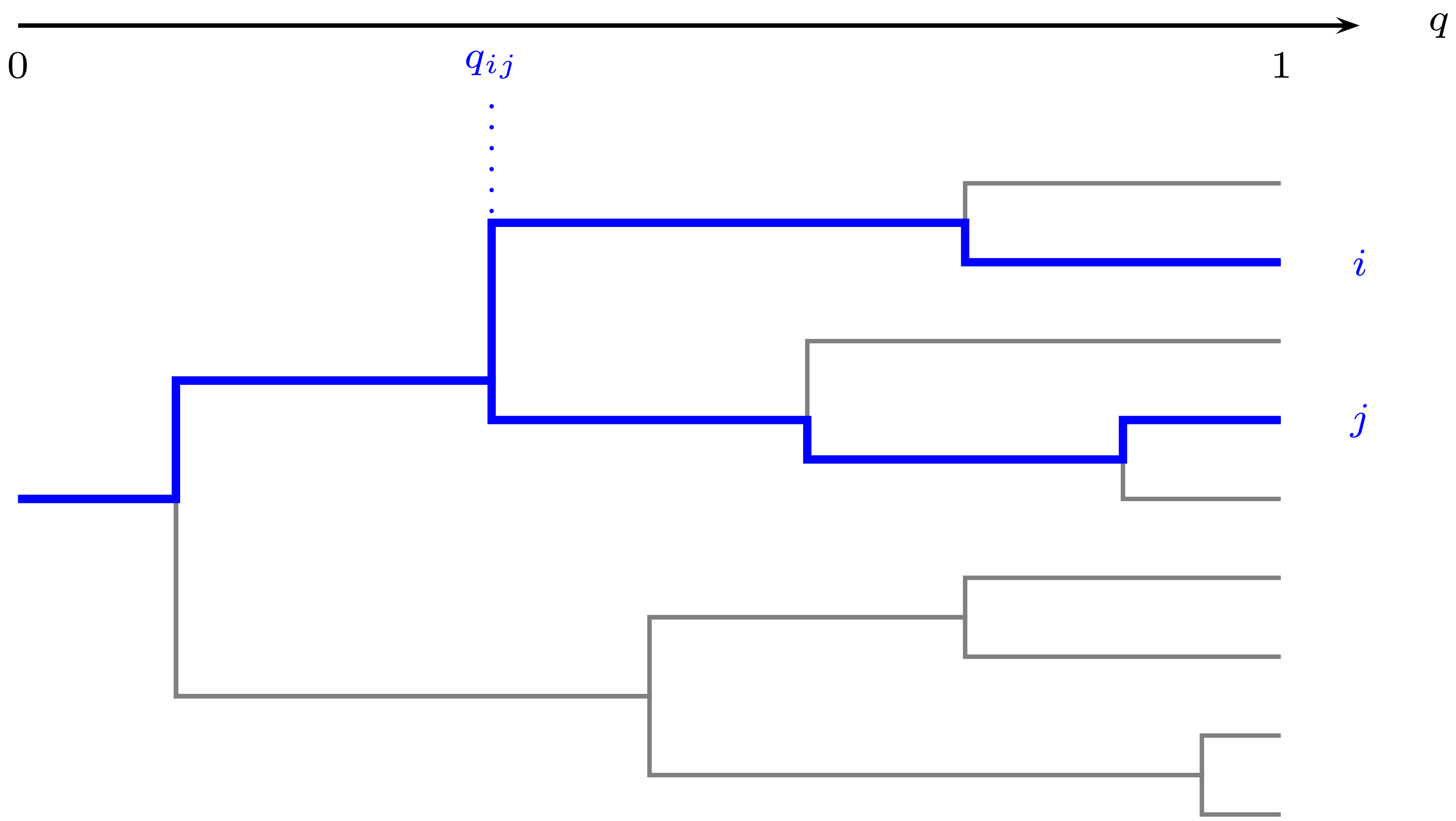
- Construct a tree such that the trunk is the market factor, and each link is a factor with entangled vol.

- The return of stock i is constructed by following a path $\mathcal{C}_i$ along the tree from trunk to leaves

$$r_i = \int_{\mathcal{C}_i, q \in [0,1]} \beta_i(q)\sigma(q)\mathrm{d}\varepsilon(q) \times \exp\left[\int_{\mathcal{C}_i, q' \in [0,q]} \alpha(q,q')\mathrm{d}\xi(q')\right]$$

- Parameters: Branching ratio of the tree $b(q)$, volatility function $\sigma(q)$, intrication function $\alpha(q,q')$

A hierachical tree model



A hierachical tree model

- Calibration on data: work in progress...

- Find simple, systematic ways to calibrate such a huge model $\Rightarrow$ stability of R_{ij}??

- Preliminary simulation results for reasonable choices: the model is able to reproduce all the empirical facts reported above, including $C(1/2, 1/2) > 1/4$ and the change of concavity of

$$\frac{C(p,p) - C_G(p,p)}{p(1-p)}$$

as ρ increases

Self-copulas

- One can also define the copula between a variable and itself, lagged:

$$C_\tau(u,v) = \mathbb{P}\left[\mathcal{P}_<(X_t) \leq u \text{ and } \mathcal{P}_{<,j}(X_{t+\tau}) \leq v\right]$$

- Example: log-normal copula

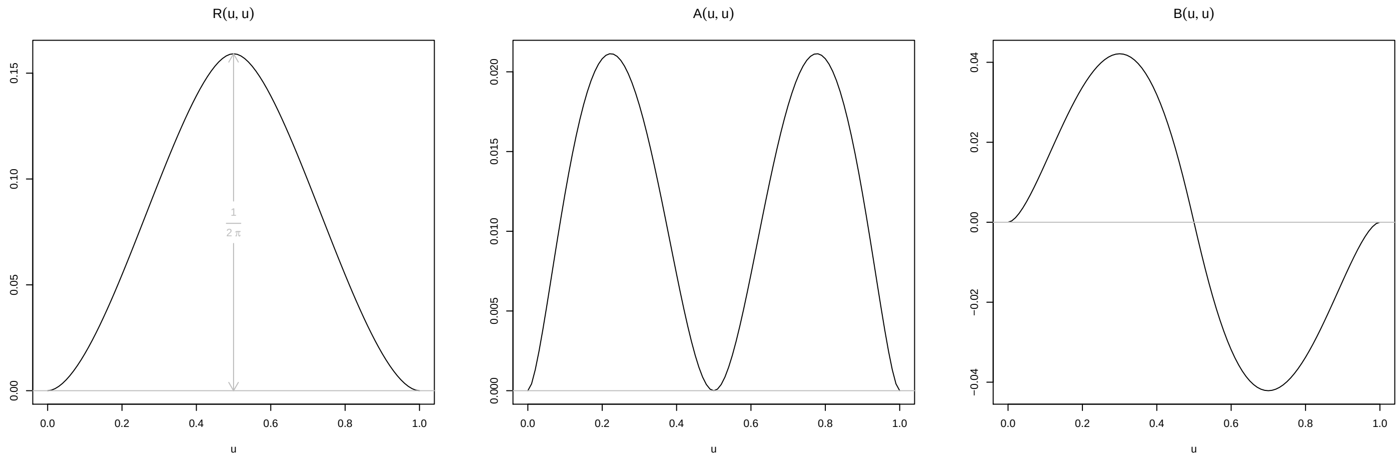
$$X_t = e^{\omega_t}\xi_t$$

with correlations between ξ's (linear), ω's (vol) and $\omega\xi$ (leverage)

- In the limit of weak correlations:

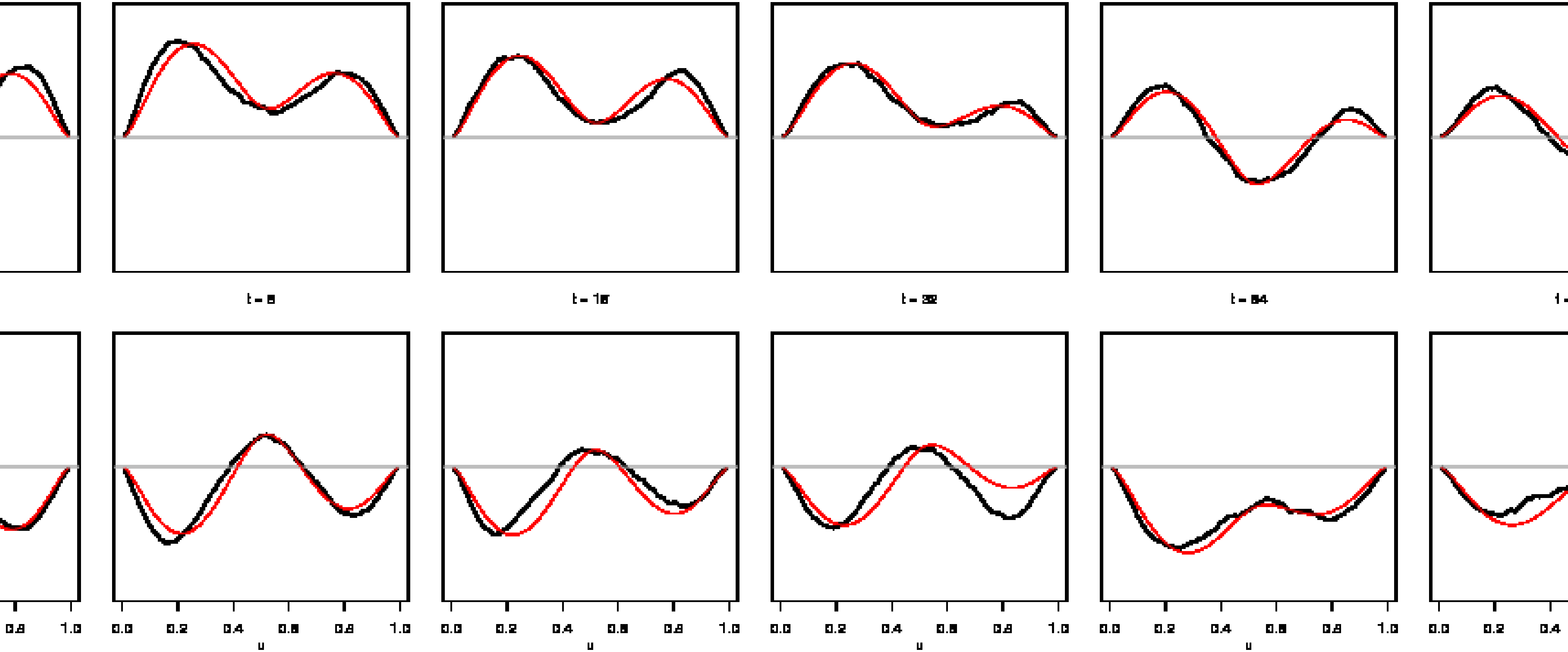
$$C_t(u,v) - uv \approx \rho R(u,v) + \alpha A(u,v) - \beta B(u,v)$$

Three corrections to independence

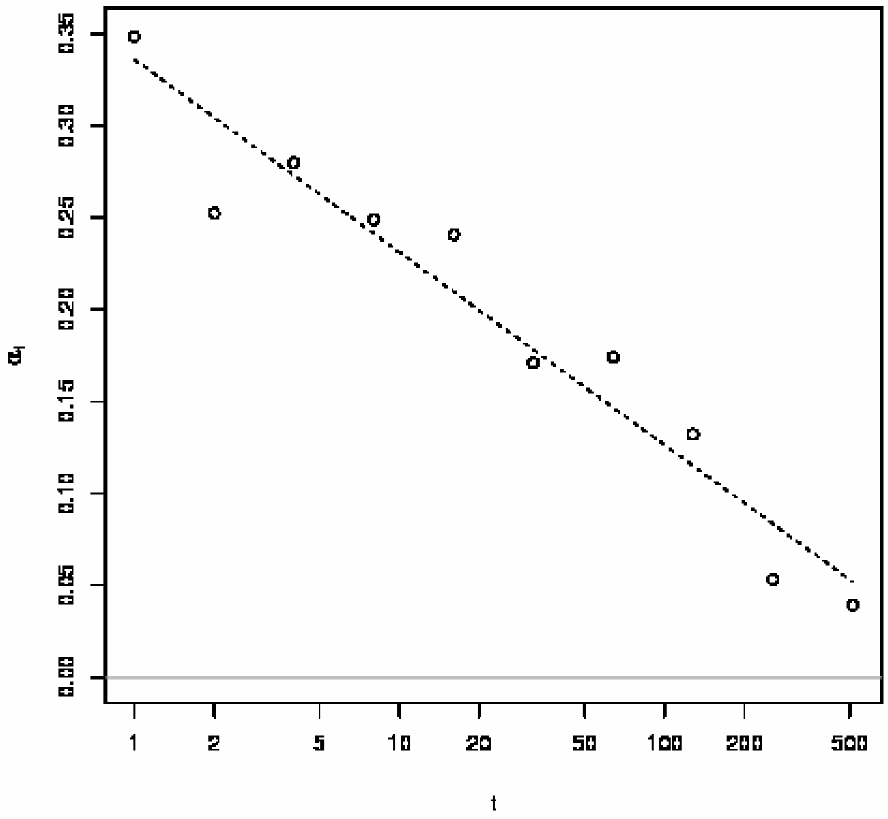


$$\rho, \alpha, \beta$$

Empirical self copulas



Long range (multifractal) memory



Self-copulas

- A direct application: GoF tests (Kolmogorov-Smirnov/Cramer-von Mises) for dependent variables

- The relevant quantity is $\sum_t (C_t(u,v) + C_{-t}(u,v) - 2uv)$

- The test is dependent on the self-copula

- $\Rightarrow$ Significant decrease of the effective number of independent variables

Conclusion – Open problems

- GoF tests for two-dimensional copulas: max of "Brownian sheets" (some progress with Rémy)

- Structural model: requires analytical progress (possible thanks to the tree structure) and numerical simulations

- Extension to account for U/L asymmetry

- Extension to describe defaults and time to defaults – move away from silly models and introduce some underlying *structure*