DOCTORAL PROGRAM
SWISS FINANCE INSTITUTE

FINANCIAL ECONOMETRICS

Eric Jondeau

FINANCIAL ECONOMETRICS

Lecture 4: Multivariate Models

swiss:finance:institute

Eric Jondeau

Objectives of the lecture

Main characteristic of the multivariate distribution: the **dependency parameter**. It measures the strength of the link between two series.

For the normal distribution, dependency is simply measured by the **Pearson's (or linear) correlation**. For this reason, we often refer to the **correlation** instead of **dependency**.

How does correlation between market returns vary when markets become agitated? This is a crucial issue from an asset-management perspective.

Many asset allocation approaches are based on the use of a sample correlation matrix. However, if correlation increases during turbulent periods, the benefits of diversification disappear when they are the most needed, i.e., during crashes.

Important concept: Tail dependence. It could be generated by

- dynamic correlations
- a distribution with different levels of dependence

Objectives of the lecture

How to test that the dependency parameter is constant over time?

- Test the equality of linear correlation coefficients computed before and after a crash. This approach may be misleading, however, because conditioning the estimation of the correlation coefficient on the sample period induces an estimator bias, if the variance changes over the 2 subperiods.

- Test in a conditional model:

  (1) Estimate the joint dynamics of stock-market returns and

  (2) Describe how conditional correlation vary over time.

We need to model the joint dynamics of several series:

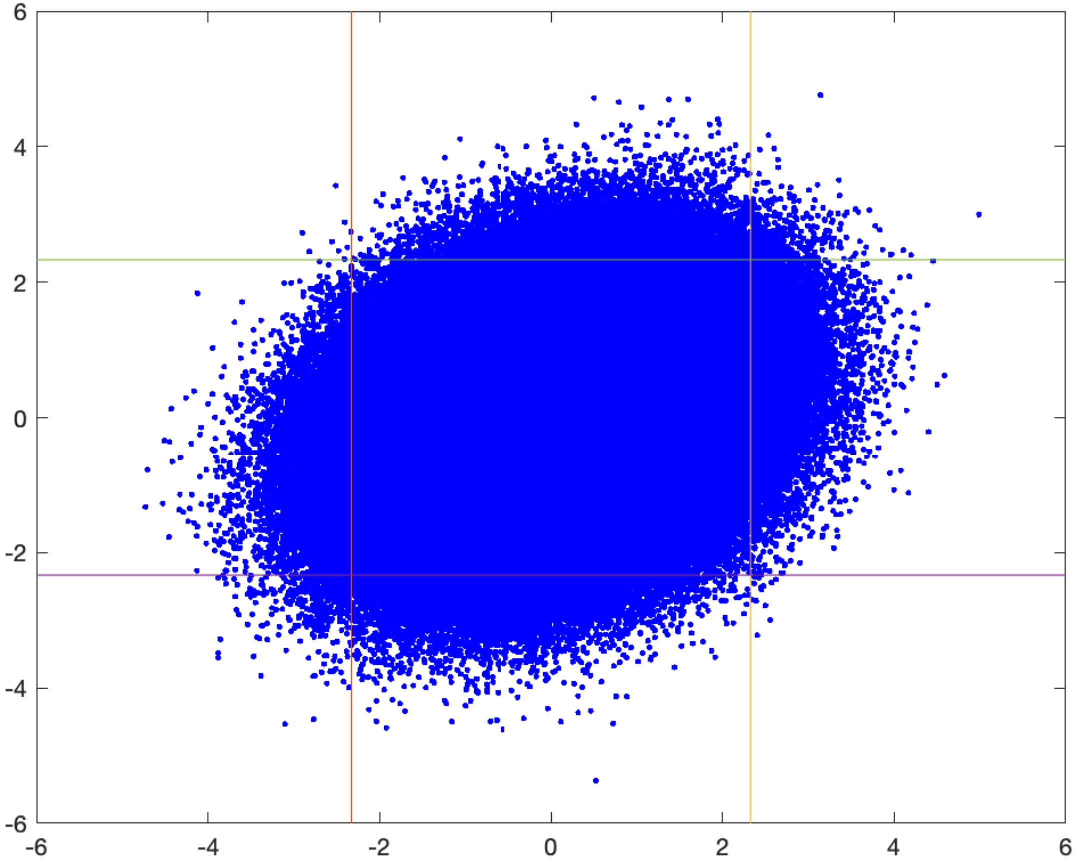
- Multivariate GARCH models (for the covariance matrix)
- Multivariate distributions or copulas models (for the distribution)

Objectives of the lecture

Simulation of a $N(\mu, \Sigma)$ with $\mu = (0; 0)$ and $\Sigma = (1\ \rho\ , \rho\ 1)$ and $\rho = 0.25$

Correlation below 1% quantile = 6.2%   Correlation above 99% quantile = 4.4%

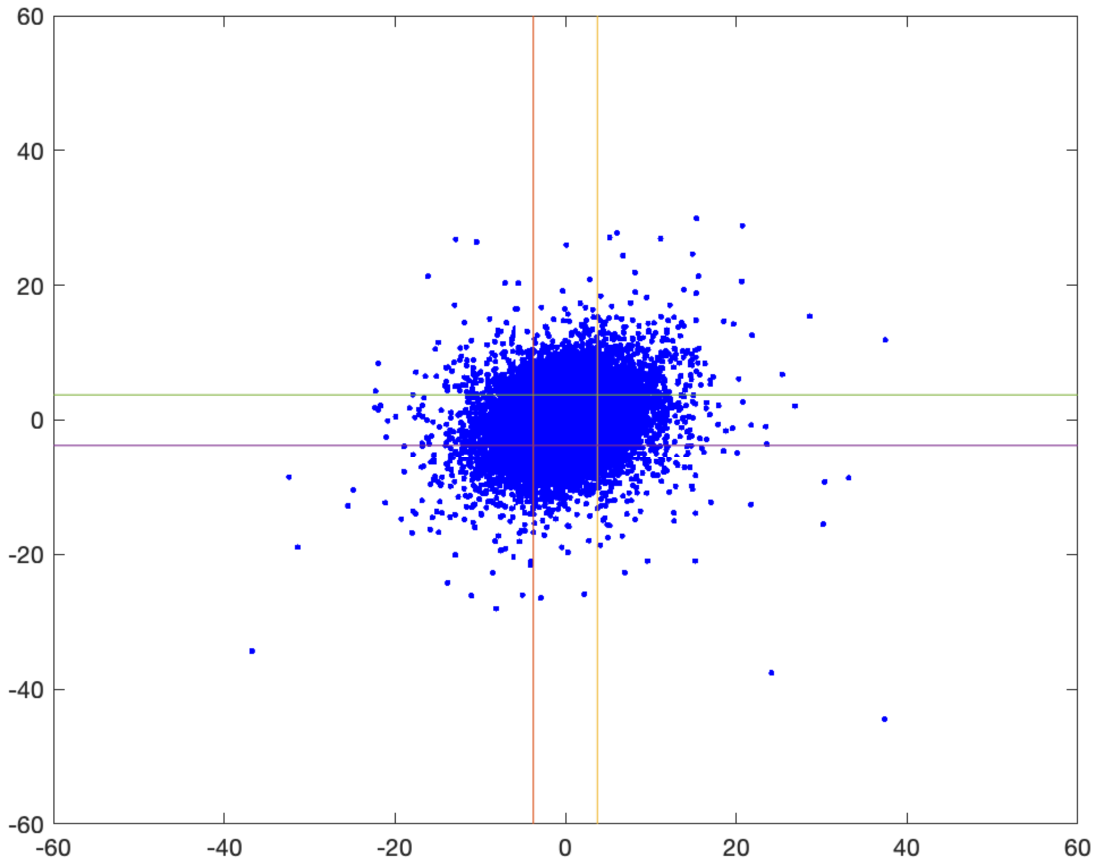
Correlation between 1% and 99% quantiles = 22%



Objectives of the lecture

Simulation of a $t(\mu, \Sigma)$ with $\mu = (0; 0)$, $\Sigma = (1\ \rho\ , \rho\ 1)$, $\rho = 0.25$, $\nu = 4$

Correlation below 1% quantile = 38.2% Correlation above 99% quantile = 38.1%
Correlation between 1% and 99% quantiles = 21.9%



Objectives of the lecture

➔ **Multivariate GARCH models**

- **Estimation issues**

- **Copulas**

Multivariate GARCH models

We consider a random vector $r_t = (r_{1,t},...,r_{n,t})'$ with joint dynamics

$$r_t = \mu_t(\theta) + \varepsilon_t \qquad \text{with} \qquad \mu_t(\theta) = E[r_t \mid I_{t-1}] = \mu(\theta, I_{t-1})$$

$$\varepsilon_t = \Sigma_t^{1/2}(\theta) z_t \qquad \text{with} \qquad \Sigma_t(\theta) = E[(r_t - \mu_t(\theta))(r_t - \mu_t(\theta))' \mid I_{t-1}] = \Sigma(\theta, I_{t-1})$$

where $\mu_t(\theta)$ is the $(n \times 1)$ vector of conditional means

$\Sigma_t(\theta)$ is the $(n \times n)$ conditional covariance matrix of the error term ε_t

θ is the vector of unknown parameters

Standardized innovation vector $z_t = \Sigma_t^{-1/2}(\theta)\varepsilon_t$ is i.i.d. with $E[z_t] = 0$ and $V[z_t] = I_n$

We assume that z_t is multivariate normal $N(0, I_n)$

Multivariate GARCH models

Multivariate normal distribution

The random vector $Z = (Z_1, \cdots, Z_n)$ is multivariate normal with zero mean and identity covariance matrix, denoted $Z \sim N_n(0, I_n)$, if its distribution is

$$\varphi(z) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} z'z\right)$$

With normal distribution, **dependency** is measured by the **covariance matrix**.

The random vector $X = (X_1, \cdots, X_n)$ is multivariate normal with $(n \times 1)$ mean vector μ and $(n \times n)$ covariance matrix Σ, denoted $X \sim N_n(\mu, \Sigma)$, if

$$X = \mu + A\,Z \qquad \text{with} \qquad \Sigma = AA'$$

so that

$$f(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

Multivariate GARCH models

Square root of a covariance matrix

What is the meaning of $\Sigma_t^{1/2}$?

This is the matrix A_t such that $A_t A_t' = \Sigma_t$.

Two possibilities:

- **Cholesky decomposition**, where A_t is lower triangular with positive diagonal entries.

- **Spectral (or Schur) decomposition**, where $\Sigma_t = V_t \Lambda_t V_t'$ and $A_t = V_t \Lambda_t^{1/2} V_t'$, with V_t the matrix of eigenvectors and Λ_t the diagonal matrix of eigenvalues.

Recommendation: The Cholesky decomposition is more appropriate when there is a "natural" ranking of the assets (most important ones first). Otherwise, the spectral decomposition is probably the safest approach.

Multivariate GARCH models

Several parameterizations have been proposed for $\Sigma_t(\theta)$.

Main issue: **dimension of the parameter vector** as the number of variables n increases

Trade-off between

- capturing the main statistical features of the distribution
- estimating a large number of parameters
- adding constraints for the covariance matrix to be positive at each date t.

Other issues:

- Should the conditional correlation be modeled instead of conditional covariance?
- Is the conditional correlation time-varying?

Multivariate GARCH models − Notations

Define D_t the $(n \times n)$ diagonal matrix with the conditional variances

$$D_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{n,t}^2 \end{pmatrix}$$

Also define Γ_t the $(n \times n)$ matrix of conditional correlations of ε_t as

$$\Gamma_t = D_t^{-1/2} \Sigma_t D_t^{-1/2} = \{\rho_t\}_{ij}.$$

We deduce the $(n \times 1)$ vector of normalized innovations $u_t = D_t^{-1/2} \varepsilon_t$.

ⓘ u_t differs from standardized innovation $z_t = \Sigma_t^{-1/2} \varepsilon_t$, as it is not orthogonalized

Multivariate GARCH models − Vech GARCH

Kraft and Engle (1982, *WP*), Bollerslev, Chou, and Kroner (1988, *J. Econometrics*).

Each element of the covariance matrix is a linear function of the most recent past cross-products of errors and conditional variances and covariances.

The **Vech GARCH(1,1) model** is defined as

$$vech(\Sigma_t) = vech(\Omega) + A\ vech(\varepsilon_{t-1}\varepsilon_{t-1}') + B\ vech(\Sigma_{t-1})$$

where Ω is an $(n \times n)$ positive definite and symmetric matrix

A and B are $(n(n+1)/2 \times n(n+1)/2)$ matrices

vech (for *vector half*) is the operator which stacks the lower triangular elements of an $(n \times n)$ matrix as an $(n(n+1)/2 \times 1)$ vector

Multivariate GARCH models − Vech GARCH

In the case $n = 2$, it can be written as

$$\begin{pmatrix} \sigma_{1,t}^2 \\ \sigma_{12,t} \\ \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} \omega_{11} \\ \omega_{12} \\ \omega_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{12,t-1} \\ \sigma_{2,t-1}^2 \end{pmatrix}$$

Number of unknown parameters: $[n(n+1)/2]\times[1+2n(n+1)/2]$ (21 for $n = 2$).

Very flexible specification, but the number of parameters is proportional to n^4.

In addition, conditions which ensure a positive definite conditional covariance matrix are difficult to verify and impose.

Multivariate GARCH models − BEKK model

Engle and Kroner (1995, *Econometric Theory*).

$$\Sigma_t = \tilde{\Omega} + \tilde{A}'(\varepsilon_{t-1}\varepsilon_{t-1}')\tilde{A} + \tilde{B}'\Sigma_{t-1}\tilde{B}$$

where $\tilde{\Omega}$, $\tilde{A}$ and $\tilde{B}$ are $(n \times n)$ matrices.

Main advantage: Σ_t is positive definite as long as $\tilde{\Omega}$ is positive definite.

In the case $n = 2$, it can be written as

$$\begin{pmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$
$$+ \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

Number of unknown parameters: $[n(n+1)/2] + 2n^2$ (11 for $n = 2$).

Multivariate GARCH models − BEKK model

Special cases

Diagonal BEKK model

$$\begin{pmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}$$
$$+ \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}$$

(7 unkonwn parameters).

Scalar BEKK model

$$\begin{pmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} + a^2 \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} + b^2 \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix}$$

(5 unkonwn parameters).

Multivariate GARCH models − The constant term

Engle and Mezrich (1996, *Risk*).

Vech and BEKK models can be estimated with the additional constraint that the expected covariance matrix is equal to the sample covariance matrix.

This constraint reduces the number of parameters dramatically and often improves the finite-sample performance.

We have the following parameterizations:

$$vech(\Omega) = \left(I_{n(n+1)/2} - A - B\right) vech(S) \qquad \textbf{(Vech model)}$$

$$\tilde{\Omega} = S - \tilde{A} S \tilde{A}' - \tilde{B} S \tilde{B}' \qquad \textbf{(BEKK model)}$$

where $S = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$ is the sample covariance matrix of residuals, and $e = (1,\ldots,1)'$.

Remark: In the univariate case, we have $\omega = \sigma^2(1 - \alpha - \beta)$.

Multivariate GARCH models − CCC model

Bollerslev (1990): **Constant conditional correlation model**.

Time-varying conditional covariances are parameterized to be proportional to the product of the corresponding conditional standard deviations.

1- Assume that each conditional variance is a univariate GARCH(1,1) model as

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2$$

where ω_i, α_i, and β_i are non-negative and $\alpha_i + \beta_i < 1$ for all $i = 1,\ldots,n$.

2- Then assume that the conditional correlation matrix $\Gamma_t = \Gamma = \{\rho\}_{ij}$ is a time-invariant $(n \times n)$ positive definite parameter matrix with unit diagonal elements

$$\Gamma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{n-1,n} \\ \rho_{1n} & \cdots & \rho_{n-1,n} & 1 \end{pmatrix}$$

Multivariate GARCH models − CCC model

3- Finally, the temporal variation in Σ_t is determined solely by the conditional variances

$$\Sigma_t = D_t^{1/2} \Gamma D_t^{1/2}$$

Σ_t is almost surely positive definite for all t. It is covariance stationary iif $\alpha_i + \beta_i < 1$ for all $i = 1,...,n$. Finally, since ε_t has finite, time-independent second moments $\sigma_{ij} = E[\sigma_{ij,t}]$ it is also a covariance stationary process.

This approach decreases the number of parameters dramatically: We only estimate the dynamics of the n conditional variances and the constant correlation matrix, so the number of parameters reduces to $3n+n(n-1)/2$ (7 for $n = 2$).

In addition, the correlation matrix Γ is estimated in a preliminary step using the sample correlation matrix of residuals!!!

Remark: the CCC model is a useful starting point for multivariate modeling, but the constancy of conditional correlation is an unrealistic feature.

Multivariate GARCH models − DCC model

Engle and Sheppard (2001): **Dynamic conditional correlation model**.

Basic idea: The conditional correlation matrix Γ_t is time varying, so the conditional covariance matrix is

$$\Sigma_t = D_t^{1/2} \Gamma_t D_t^{1/2}$$

The conditional correlation matrix is defined by:

$$\Gamma_t = diag(Q_t)^{-1/2} \times Q_t \times diag(Q_t)^{-1/2}$$

$$Q_t = (1 - \delta_1 - \delta_2)\bar{Q} + \delta_1 (u_{t-1} u'_{t-1}) + \delta_2 Q_{t-1}$$

$$u_t = D_t^{-1/2} \varepsilon_t = \left\{ \varepsilon_{i,t} / \sigma_{i,t} \right\}_{i=1,\ldots,n}$$

where $\bar{Q}$ is the sample covariance matrix of u_t and δ_1 and δ_2 satisfy $0 \leq \delta_1, \delta_2 \leq 1$ and $\delta_1 + \delta_2 \leq 1$. $diag(Q_t)$ of an $(n \times n)$ matrix is the $(n \times 1)$ vector containing its diagonal.

Once the restrictions are imposed, the conditional correlation matrix (Γ_t) is guaranteed to be positive definite.

Objectives of the lecture

- Multivariate GARCH models

➔ **Estimation issues**

- Copulas

Multivariate GARCH models − Estimation issues

Multivariate GARCH models are estimated in a very similar way as univariate models.

Suppose an $(n\times 1)$ vector of time series $\{r_t\}_{t=1}^{T}$. Unknown parameters are θ. With a multivariate normal distribution, the **log-likelihood function** for r_t is

$$\log(L_T(\theta)) = \sum_{t=1}^{T}\log(\ell_t(\theta))$$

where $$\log(\ell_t(\theta)) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[\log\left|\Sigma_t(\theta)\right| + (r_t-\mu_t(\theta))'\Sigma_t^{-1}(\theta)(r_t-\mu_t(\theta))\right]$$

The ML estimator $\hat{\theta}_{ML}$ satisfies $L_T(\hat{\theta}_{ML}) \geq L_T(\theta), \forall\theta\in\Theta$, where Θ is the domain of definition of θ.

$\hat{\theta}_{ML}$ is asymptotically normal with distribution: $\sqrt{T}(\hat{\theta}_{ML}-\theta_0) \overset{d}{\sim} N(0, A_0^{-1})$

where A_0 is the information matrix at θ_0: $A_0 = -\frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial^2\log(\ell_t(\theta_0))}{\partial\theta\partial\theta'}\right]$

Multivariate GARCH models − Estimation of DCC model

Two-step estimation: based on the idea that parameters of the conditional variances (θ_V) and the conditional correlations (θ_C) can be estimated separately.

Reason: The log-likelihood, with $\theta = (\theta_V', \theta_C')'$

$$\log(L_T(\theta)) = \sum_{t=1}^{T} \log(\ell_t(\theta)) = -\frac{1}{2}\sum_{t=1}^{T}\left[n\log(2\pi) + \log\left|\Sigma_t\right| + (r_t - \mu_t)'\Sigma_t^{-1}(r_t - \mu_t)\right]$$

can be written as the sum of a volatility part and a correlation part.

As $\Sigma_t = D_t^{1/2}\Gamma_t D_t^{1/2}$ and D_t is diagonal, we have

(1) $$\log\left|\Sigma_t\right| = \log\left|D_t\right| + \log\left|\Gamma_t\right|$$

(2) $$\begin{aligned}(r_t - \mu_t)'\Sigma_t^{-1}(r_t - \mu_t) &= (r_t - \mu_t)'(D_t^{1/2}\Gamma_t D_t^{1/2})^{-1}(r_t - \mu_t) \\ &= u_t'\Gamma_t^{-1}u_t + (r_t - \mu_t)'D_t^{-1}(r_t - \mu_t) - u_t'u_t\end{aligned}$$

where $u_t = D_t^{-1/2}\varepsilon_t$ is vector of normalized innovations.

Multivariate GARCH models − Estimation of DCC model

The last two terms of the second equality are clearly equal, but this expression allows the log-likelihood to be broken down in the two following terms:

$$\log(L_T(\theta_V,\theta_C)) = \log(L_T^V(\theta_V)) + \log(L_T^C(\theta_V,\theta_C))$$

with

$$\log(L_T^V(\theta_V)) = -\frac{1}{2}\sum_{t=1}^{T}\left[n\log(2\pi) + \log\left|D_t\right| + (r_t-\mu_t)' D_t^{-1}(r_t-\mu_t)\right]$$

$$= -\sum_{i=1}^{n}\left[\frac{T}{2}\log(2\pi) + \frac{1}{2}\sum_{t=1}^{T}\left[\log(\sigma_{i,t}^2) + \left(\frac{r_{i,t}-\mu_{i,t}}{\sigma_{i,t}}\right)^2\right]\right]$$

$$\log(L_T^C(\theta_V,\theta_C)) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log\left|\Gamma_t\right| + u_t'\Gamma_t^{-1}u_t - u_t'u_t\right]$$

$\log(L_T^V(\theta_V))$ is the sum of log-likelihoods of the individual GARCH equations.

Multivariate GARCH models − Estimation of DCC model

The two-step estimation then relies on maximizing the log-likelihood as follows.

(1) Estimate the volatility parameters

$$\hat{\theta}_V \in \arg\max_{\{\theta_V\}} \log(L_T^V(\theta_V))$$

As squared residuals do not depend on correlation parameters, these parameters can be ignored in the estimation of the conditional volatility dynamics.

(2) Estimate the correlation parameters

$$\hat{\theta}_C \in \arg\max_{\{\theta_C\}} \log(L_T^C(\hat{\theta}_V, \theta_C)) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log\left|\Gamma_t\right| + \hat{u}_t{}'\Gamma_t^{-1}\hat{u}_t - \hat{u}_t{}'\hat{u}_t\right]$$

with $\hat{u}_{i,t} = (r_{i,t} - \hat{\mu}_{i,t}) / \hat{\sigma}_{i,t}$

Multivariate GARCH models − Estimation of DCC model

The two-step estimator is consistent and asymptotically normal, with distribution

$$\sqrt{T}(\hat{\theta}_{TS}-\theta_0) \to N(0, A_0^{-1} B_0 A_0^{-1}{}')$$

where
$$A_0 = -\frac{1}{T} E \sum_{t=1}^{T} \begin{bmatrix} \dfrac{\partial^2 \log(\ell_t^V(\theta_{V0}))}{\partial\theta_V \partial\theta_V{}'} & 0 \\ \dfrac{\partial^2 \log(\ell_t^C(\theta_0))}{\partial\theta_V \partial\theta_C{}'} & \dfrac{\partial^2 \log(\ell_t^C(\theta_0))}{\partial\theta_C \partial\theta_C{}'} \end{bmatrix}$$

$$B_0 = \frac{1}{T} E \sum_{t=1}^{T} \begin{bmatrix} \dfrac{\partial \log(\ell_t^V(\theta_{V0}))}{\partial\theta_V} \dfrac{\partial \log(\ell_t^V(\theta_{V0}))'}{\partial\theta_V} & \dfrac{\partial \log(\ell_t^C(\theta_0))}{\partial\theta_C} \dfrac{\partial \log(\ell_t^V(\theta_{V0}))'}{\partial\theta_V} \\ \dfrac{\partial \log(\ell_t^V(\theta_{V0}))}{\partial\theta_V} \dfrac{\partial \log(\ell_t^C(\theta_0))'}{\partial\theta_C} & \dfrac{\partial \log(\ell_t^C(\theta_0))}{\partial\theta_C} \dfrac{\partial \log(\ell_t^C(\theta_0))'}{\partial\theta_C} \end{bmatrix}$$

Due to the structure of A_0, the asymptotic variances of the GARCH parameters θ_V for each series are the standard robust covariance matrix estimators. For the second stage parameters, however, the asymptotic variance involves all parameters.

Multivariate GARCH models − Illustration

Daily index returns from January 1980 to July 2021

	SP500-DAX		SP500-FT-SE	
	estimate	std err.	estimate	std err.
Diagonal BEKK model				
ω_{11}	0.1151	0.0111	0.1047	0.0139
ω_{12}	0.0574	0.0089	0.0590	0.0123
ω_{22}	0.1261	0.0130	0.0803	0.0176
a_{11}	0.2752	0.0149	0.2520	0.0175
a_{22}	0.2514	0.0137	0.1567	0.0129
b_{11}	0.9561	0.0046	0.9630	0.0054
b_{22}	0.9626	0.0035	0.9832	0.0036
log-lik.	-29780	-	-28323	-
Scalar BEKK model				
ω_{11}	0.1104	0.0094	0.0825	0.0125
ω_{12}	0.0577	0.0102	0.0509	0.0091
ω_{22}	0.1307	0.0132	0.0980	0.0145
a	0.2627	0.0115	0.1896	0.0135
b	0.9596	0.0034	0.9778	0.0037
log-lik.	-29783	-	-28416	-

Multivariate GARCH models − Illustration

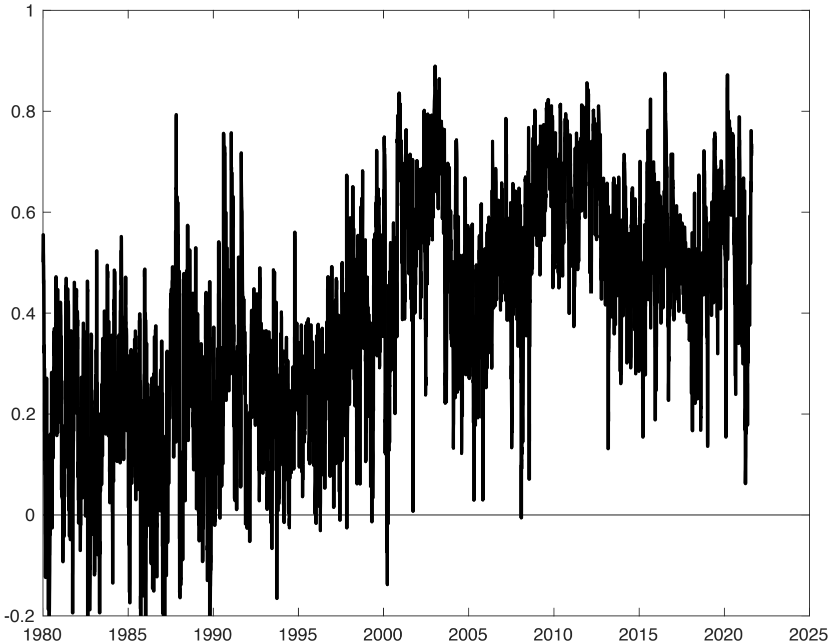
Daily index returns from January 1980 to July 2021

	SP500-DAX		SP500-FT-SE	
	estimate	std err.	estimate	std err.
Univariate GARCH models				
ω_1	0.0153	0.0035	0.0153	0.0035
a_1	0.0801	0.0086	0.0801	0.0086
b_1	0.9058	0.0103	0.9058	0.0103
ω_2	0.0287	0.0075	0.0092	0.0047
a_2	0.0965	0.0140	0.0249	0.0050
b_2	0.8879	0.0142	0.9669	0.0089
CCC model				
ρ	0.3871	0.0131	0.3750	0.0269
log-lik.	-29976	-	-28540	-
DCC model				
δ_1	0.0088	0.0025	0.0055	0.0040
δ_2	0.9904	0.0029	0.9942	0.0045
log-lik.	-29597	-	-28245	-

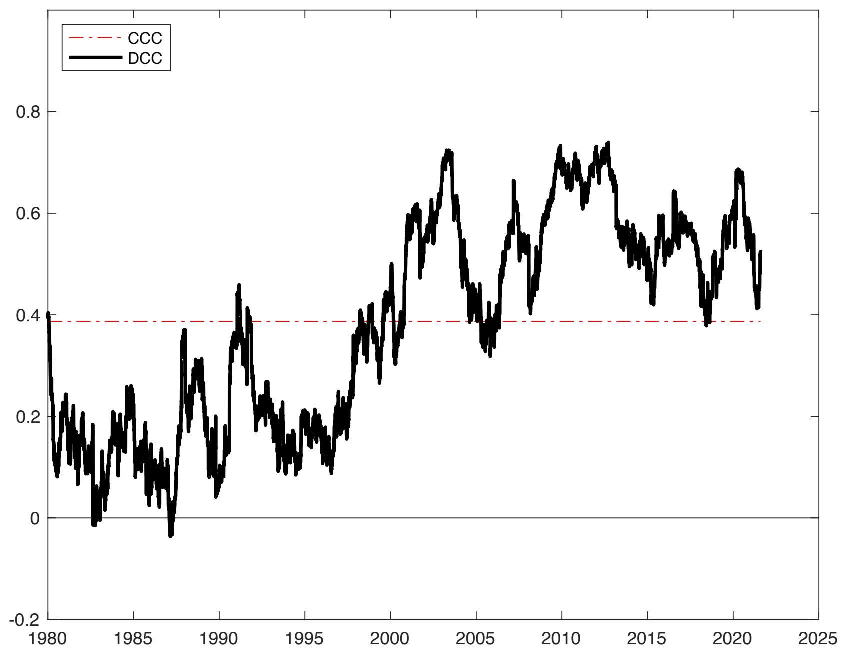
Multivariate GARCH models − Illustration

SP500 and DAX daily returns from January 1980 to July 2021

BEKK model



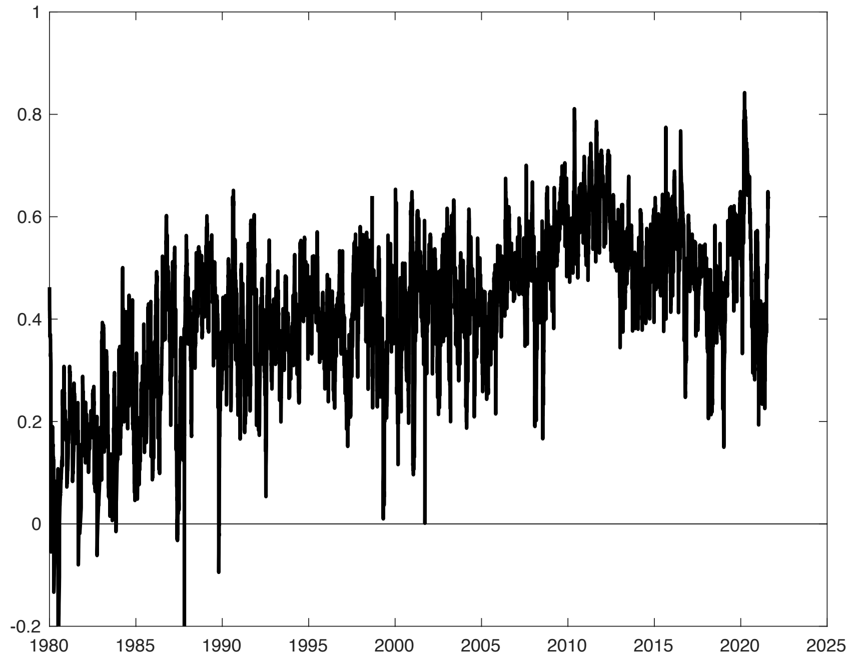
CCC and DCC GARCH models



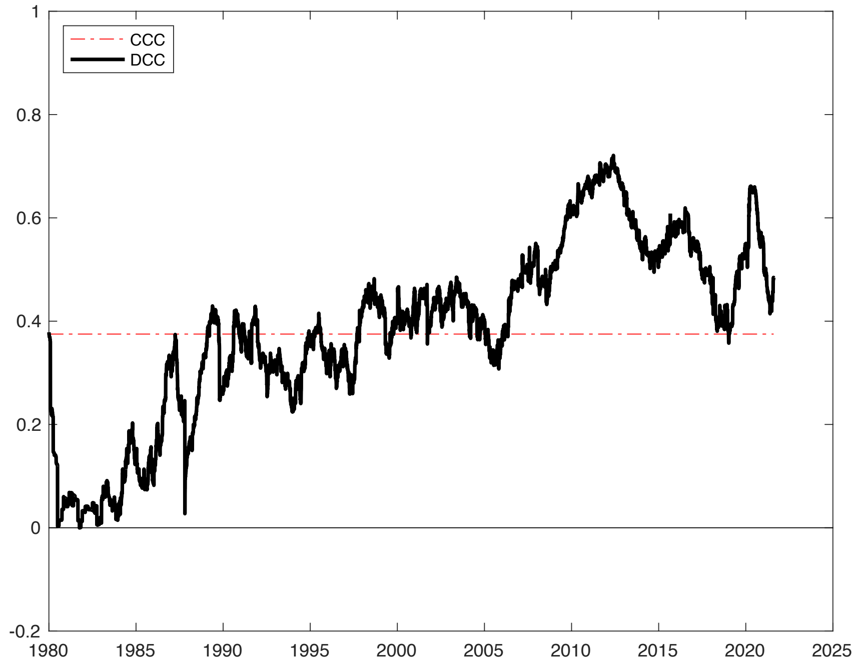
Multivariate GARCH models − Illustration

SP500 and FTSE daily returns from January 1980 to July 2021

BEKK model



CCC and DCC GARCH models



Objectives of the lecture

- Multivariate GARCH models

- Estimation issues

➔ **Copulas**

Copulas

In many situations where marginal distributions are not Gaussian, it is simply impossible to define a joint distribution.

This is the case when the two variables have **different marginal distributions** (for instance, a Student *t* variate and a Pareto variate). This is also the case for several marginal distributions, for which a multivariate extension does not exist.

In such contexts, a solution is to use **copula models**. These functions relate two marginal distributions instead two series directly. Therefore, once margins have been estimated, no reference is made to their true functional form. **Copula models can relate any kind of margins.**

With copula models, it is possible to generate **non-linear dependence**. In general, the Pearson's correlation is not the appropriate measure of dependence across variables.

Definitions and properties

Consider two random variables X and Y with marginal distributions, or margins, $F(x) = \Pr\left[X \leq x\right]$ and $G(y) = \Pr\left[Y \leq y\right]$. The cdf are continuous.

X and Y also have a joint distribution function, $H(x, y) = \Pr\left[X \leq x, Y \leq y\right]$.

All the cdf, $F(.)$, $G(.)$, and $H(.,.)$ have as range the interval [0,1].

In some cases, a multivariate distribution exists, so that the function $H(.,.)$ has an explicit expression (for instance, the multivariate normal distribution).

In many cases, however, an analytical expression of the margins $F(.)$ and $G(.)$ is relatively easy to obtain, while the joint distribution $H(.,.)$ does not exist.

This is where copulas are useful because they link margins into a multivariate distribution function.

Definitions and properties

Definition: A **bivariate copula** is a function $C:[0,1]\times[0,1]\rightarrow[0,1]$ with the three following properties:

1. $C(u,v)$ is increasing in u and v.

2. $C(u,0)=0, \quad C(u,1)=u, \quad C(0,v)=0, \quad C(1,v)=v$.

3. $\Pr\left[u_1 \leq U \leq u_2, v_1 \leq V \leq v_2\right]=C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1)\geq 0$, $\forall\, u_1,v_1,u_2,v_2$ in $[0,1]$ such that $0\leq u_1\leq u_2\leq 1$ and $0\leq v_1\leq v_2\leq 1$.

Property 1: When one marginal distribution is constant, the joint probability will increase provided that the other marginal distribution increases.

Property 2: If one margin has zero probability to occur, then it must be the same for the joint occurrence. Also, if on the contrary, one margin is certain to occur, then the probability of a joint occurrence is determined by the remaining margin probability.

Property 3: If both u and v increase, then the joint probability also increases. This property is, therefore, a multivariate extension of the condition that a cdf is increasing.

Definitions and properties

If $u = F(x)$ and $v = G(y)$, then $C\left(F(x), G(y)\right)$ describes the joint distribution of (X,Y).

Theorem (Sklar, 1959): Let H be a joint distribution function of X and Y with marginal distributions F and G, respectively. Then,

- there exists a copula $C:[0,1]\times[0,1]\rightarrow[0,1]$ such that, for all real numbers (x,y),

$$H(x,y) = C\left(F(x), G(y)\right)$$

Furthermore, if F and G are continuous, then C is unique.

- conversely, if C is a copula and F and G are univariate distribution functions, then $H(x,y) = C\left(F(x), G(y)\right)$ is a joint distribution function with marginal distributions F and G.

The **density c of the copula** is:

$$c(u,v) = \frac{\partial^2 C\left(u,v\right)}{\partial u\, \partial v}$$

The **density h of the distribution** H is:

$$h(x,y) = c(F(x), G(y)) \times f(x) \times g(y)$$

Example 1: Gaussian copula

Cdf of the Gaussian copula

$$C_\rho(u,v) = \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) ds dt$$

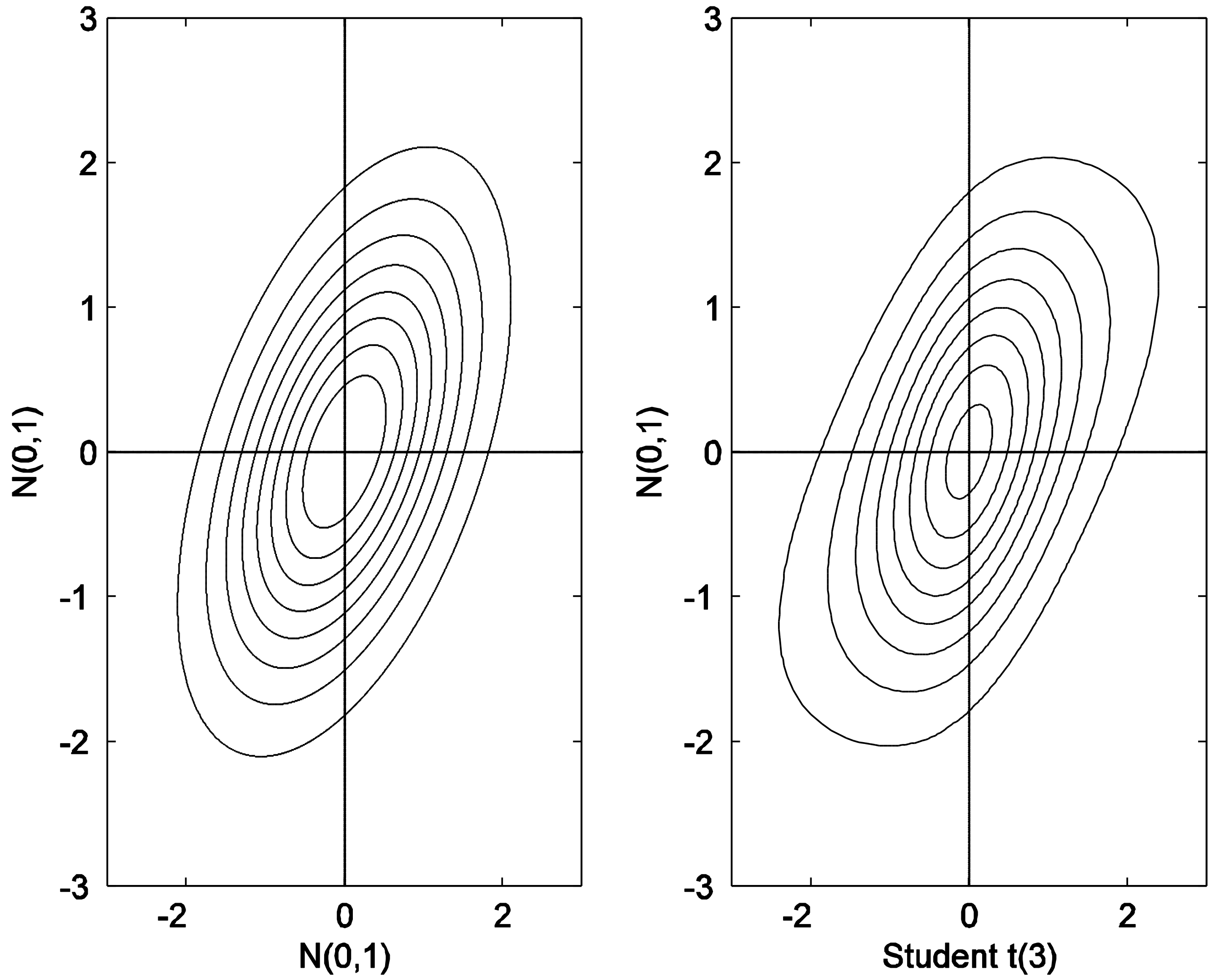
where Φ_ρ is the bivariate standardized Gaussian cdf with correlation $\rho \in [-1,1]$ and Φ^{-1} denotes the inverse of the cdf of the univariate standard normal distribution.

The **density** of the Gaussian copula is given by

$$c_\rho(u,v) = \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2}\psi'\left(R^{-1} - I_2\right)\psi\right) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)} + \frac{s^2 + t^2}{2}\right)$$

where $\psi = \left(\Phi^{-1}(u), \Phi^{-1}(v)\right)'$, and R is the (2,2) correlation matrix between u and v with ρ as correlation parameter.

Example 1: Gaussian copula



Example 2: Student *t* copula

Cdf of the Student *t* copula

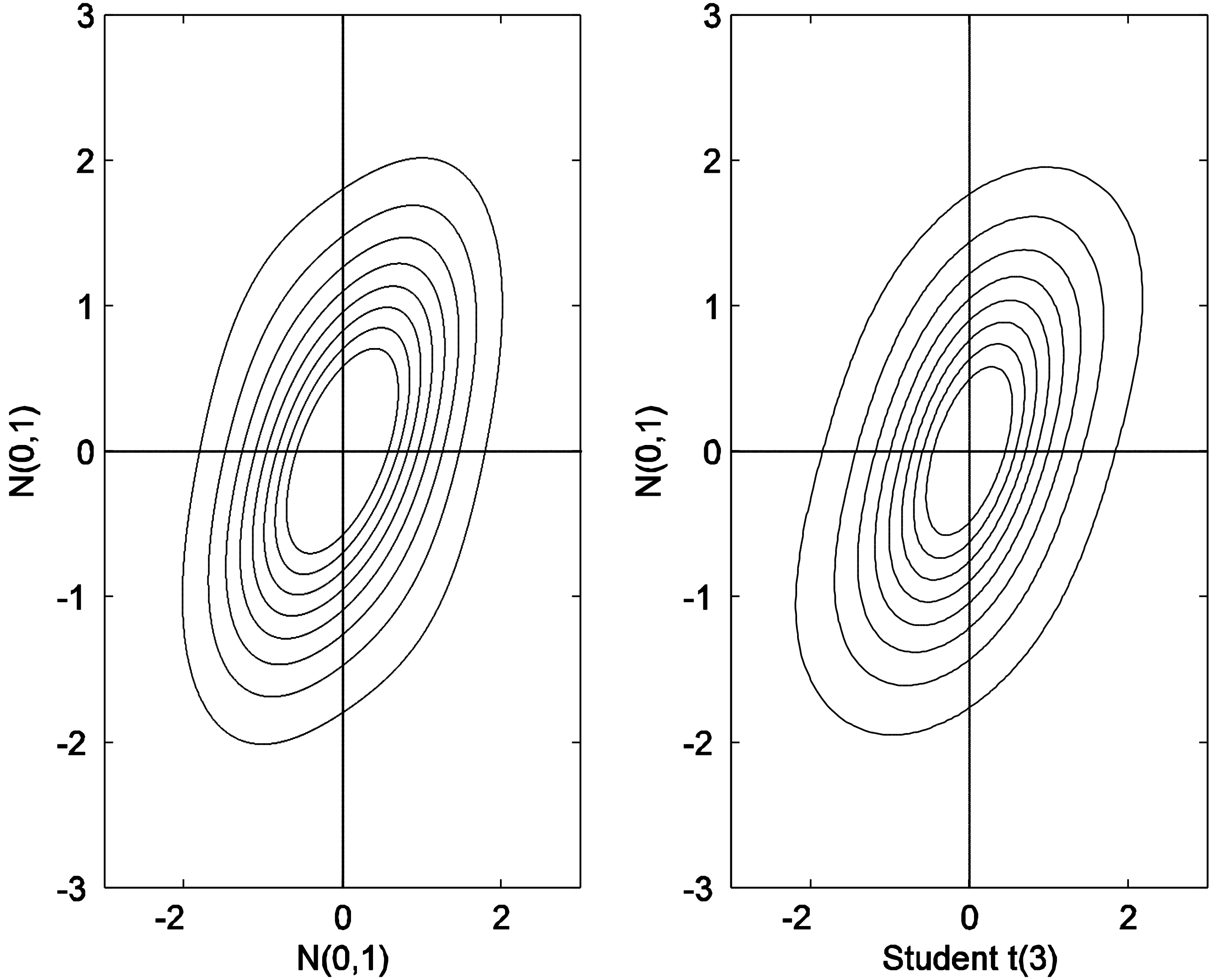
$$C_{\rho,n}(u,v) = t_{\rho,n}\left(t_n^{-1}(u), t_n^{-1}(v)\right) = \int_{-\infty}^{t_n^{-1}(u)} \int_{-\infty}^{t_n^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{n(1-\rho^2)}\right)^{-\frac{n+2}{2}} dsdt$$

where $\psi = \left(t_n^{-1}(u), t_n^{-1}(v)\right)'$, $t_{\rho,n}$ is the bivariate Student *t* cdf with n degrees of freedom and correlation ρ, and t_n^{-1} denotes the inverse of the cdf of the univariate Student *t* distribution with n degrees of freedom.

The **density** of the Student *t* copula is given by

$$c_{\rho,n}(u,v) = \frac{1}{|R|^{1/2}} \frac{\Gamma\left(\frac{n+2}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\left(\Gamma\left(\frac{n+1}{2}\right)\right)^2} \frac{\left(1 + \frac{1}{n}\psi' R^{-1}\psi\right)^{-\frac{n+2}{2}}}{\prod_{i=1}^{2}\left(1 + \frac{1}{n}\psi_i^2\right)^{-\frac{n+1}{2}}}$$

Example 2: Student *t* copula

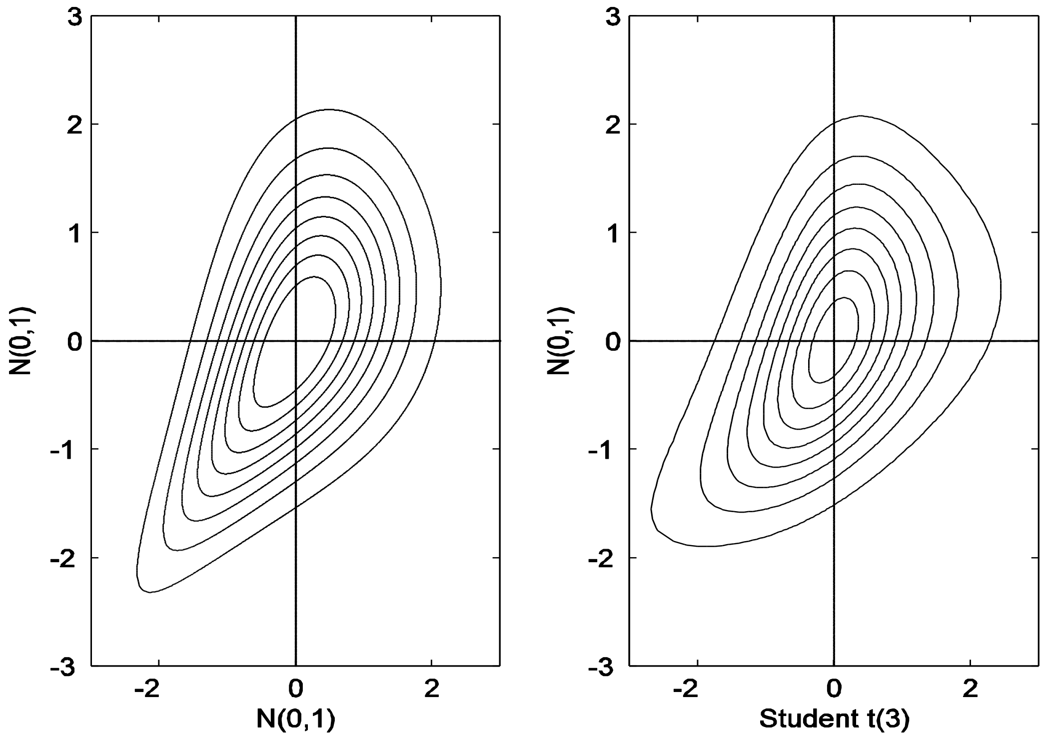


Example 3: Clayton copula

Cdf of the Clayton copula: $$C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}$$

Density of the copula: $$c_\theta(u,v) = (1+\theta)(uv)^{-\theta-1}\left(u^{-\theta} + v^{-\theta} - 1\right)^{-2-1/\theta}$$

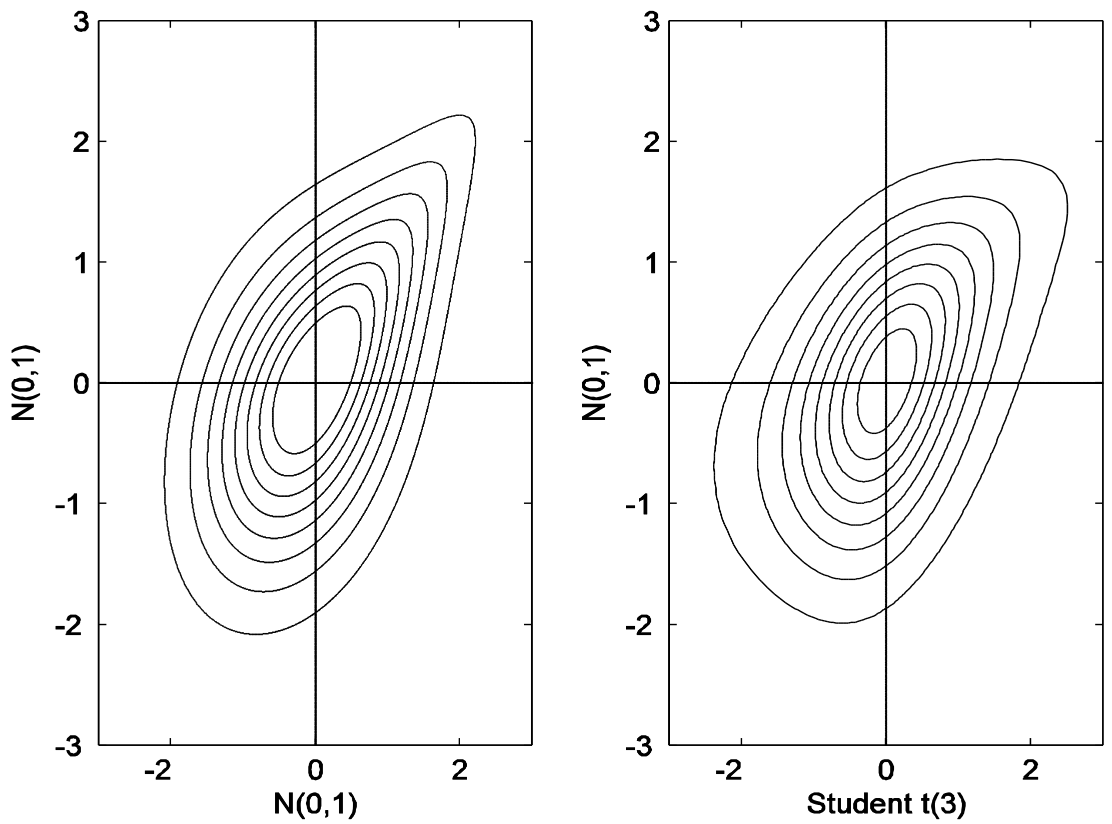
The Clayton copula generates dependence in the lower tail but not in the upper tail.



Example 4: Gumbel copula

Cdf of the Gumbel copula: $$C_\theta(u,v) = \exp\left\{-\left[(-\log u)^\theta + (-\log v)^\theta\right]^{1/\theta}\right\}$$

The Gumbel copula generates dependence in the upper tail but not in the lower tail.



Estimation

Several approaches to estimate the parameters of a copula model, depending on the type of information available:

- Standard ML estimation,

- Two-step estimation procedure (Shih and Louis, 1995, Joe and Xu, 1996),

- Semi-parametrically estimation (Genest, Ghoudi, and Rivest, 1995),

ML Estimation

Parameters of the marginal densities f and g are denoted θ_x and θ_y, respectively, and parameters of the copula are denoted θ_γ. We define the vector $\theta = (\theta_x, \theta_y, \theta_\gamma)'$.

The ML estimate of a model is obtained by maximizing the conditional log-likelihood function:

$$\log(L_T(\theta)) = \sum_{t=1}^{T} \log\left(\ell_t(\theta)\right) = \sum_{t=1}^{T} \log\left(c_{\theta_\gamma}\left(F(x_t,\theta_x), G(y_t,\theta_y)\right) \times f(x_t,\theta_x) \times g(y_t,\theta_y)\right)$$

$$= \sum_{t=1}^{T} \log\left(c_{\theta_\gamma}\left(F(x_t,\theta_x), G(y_t,\theta_y)\right)\right) + \sum_{t=1}^{T}\left[\log\left(f(x_t,\theta_x)\right) + \log\left(g(y_t,\theta_y)\right)\right]$$

The ML estimator $\hat{\theta}_{ML}$ is asymptotically normal with asymptotic distribution

$$\sqrt{T}(\hat{\theta}_{ML} - \theta) \to N(0, A_0^{-1})$$

where A_0 is the information matrix of Fisher.

Two-step Estimation (or Inference Functions for Margins)

In practical applications, MLE may be difficult to implement. First, the dimension of the optimization problem can be very large. Second, in general, there is no analytical expression of the gradient of the likelihood.

In some cases, the vector of parameters can be split into different parts: those associated with the margins and those associated with the copula. This gives rise to the following **two-step estimator** (Shih and Louis, 1995, Joe and Xu, 1996):

1- The first step is the estimation of the margins:

$$\tilde{\theta}_x \in \arg \max_{\{\theta_x \in \Theta_x\}} \sum_{t=1}^{T} \log\left(f(x_t, \theta_x)\right)$$

$$\tilde{\theta}_y \in \arg \max_{\{\theta_y \in \Theta_y\}} \sum_{t=1}^{T} \log\left(g(y_t, \theta_y)\right)$$

Two-step Estimation (or Inference Functions for Margins)

2- In a second step, parameters θ_γ of the copula model can be estimated conditionally on the margin parameters

$$\tilde{\theta}_\gamma \in \arg \max_{\{\theta_\gamma \in \Theta_\gamma\}} \sum_{t=1}^{T} \log\left(c_{\theta_\gamma}\left(F(x_t,\tilde{\theta}_x),G(y_t,\tilde{\theta}_y)\right)\right)$$

If the model is correctly specified, then under rather mild assumptions, $\tilde{\theta}_{IFM} = (\tilde{\theta}_x,\tilde{\theta}_y,\tilde{\theta}_\gamma)'$ is consistent and asymptotically normal (Joe, 1997):

$$\sqrt{T}(\tilde{\theta}_{IFM} - \theta) \to N(0,\Omega_0^{-1})$$

where $\Omega_0 = A_0^{-1}B_0(A_0^{-1})'$ is the estimator of the covariance matrix, with

$$A_0 = E\left[\frac{\partial g(\theta)}{\partial \theta}\right] \qquad \text{and} \qquad B_0 = E\left[g(\theta)'g(\theta)\right]$$

and $g(\theta) = \left(\frac{\partial \log(f_x(\theta))}{\partial \theta_x}, \frac{\partial \log(f_y(\theta))}{\partial \theta_y}, \frac{\partial \log(c(\theta))}{\partial \theta_\gamma}\right)$ is the gradient of the model.

Semi-parametric ML (or Canonical ML)

If a parametric marginal distribution is not needed, we can use the marginal empirical cdf (Genest, Ghoudi, and Rivest, 1995):

$$\hat{F}_T(x_{(\tau)}) = \frac{1}{T}\sum_{t=1}^{T} 1_{\{x_t \le x_{(\tau)}\}}$$

where $x_{(1)} \le \ldots \le x_{(T)}$ is the ordered sample of observations. In other words, $\hat{F}_t(x_{(\tau)})$ represents the frequency of observations below or equal to $x_{(\tau)}$ in the sample $\{x_t\}_{t=1}^{T}$.

Working with the empirical margins avoids the estimation of the parameters of the margins. Then, θ_γ is estimated by maximizing the pseudo log-likelihood

$$\tilde{\theta}_\gamma \in \arg \max_{\{\theta_\gamma \in \Theta_\gamma\}} \sum_{t=1}^{T} \log\left(c_{\theta_\gamma}\left(\hat{F}(x_t), \hat{G}(y_t)\right)\right)$$

The estimator $\tilde{\theta}_\gamma$ is asymptotically normal, with a larger asymptotic variance than the ML estimator (obtained assuming that the margins are known).

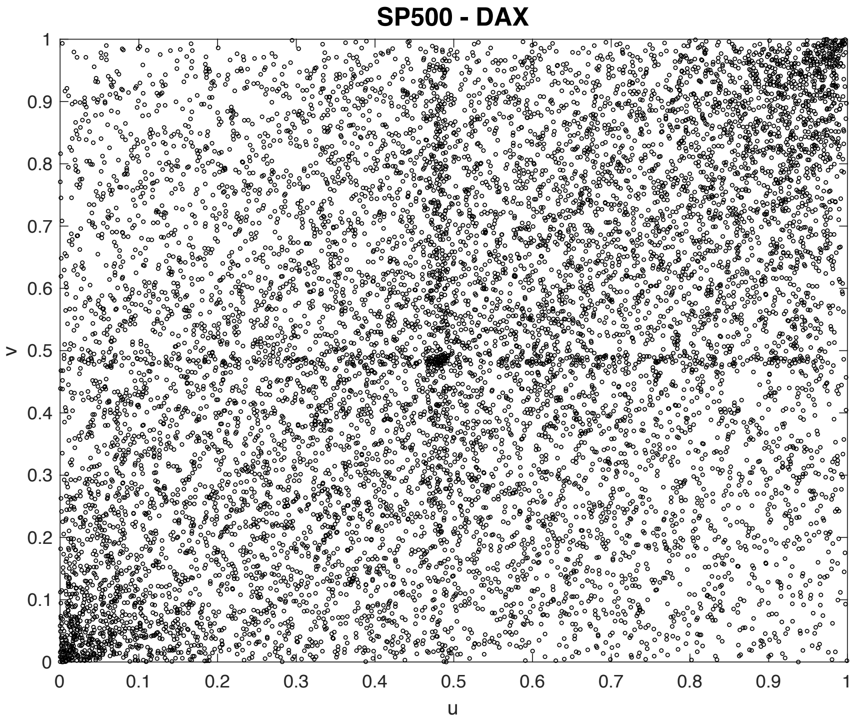
Illustration

Copy models for two pairs of daily returns from January 1980 to July 2021

- SP500 and DAX
- SP500 and FTSE

Constant conditional mean with simple GARCH(1,1) model for the margins.

Scatterplot of margins (with Student t distribution)



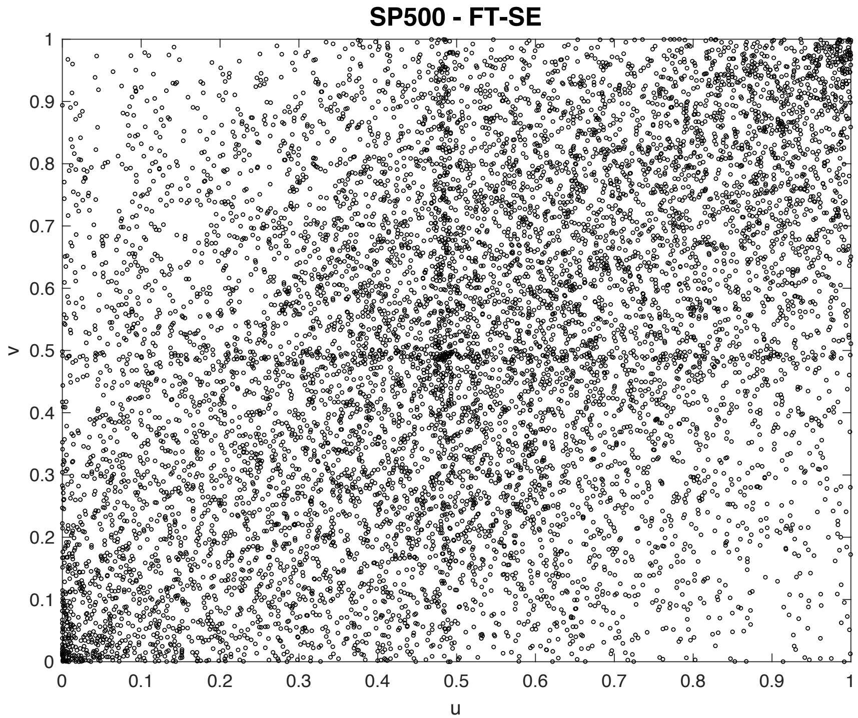


Illustration – Comparison of copula models

Daily returns from January 1980 to July 2021

	S&P - DAX		S&P - FTSE	
	Log-lik.	**Param. estimate**	**Log-lik.**	**Param. estimate**
Gaussian margins				
Gaussian	880.9055	0.3872	1037.9723	0.4457
Student *t*	**989.5955**	0.4046	**1116.8460**	0.4522
Clayton	550.8052	0.3435	704.9097	0.4367
Gumbel	877.5087	1.3443	970.5052	1.4011
Student *t* margins				
Gaussian	866.9396	0.3848	1030.5599	0.4354
Student *t*	**979.5820**	0.3850	**1111.1889**	0.4363
Clayton	723.7297	0.4639	870.7841	0.5534
Gumbel	877.0488	1.3277	987.7457	1.3830
Empirical		0.3838		0.4304