Introduction to Copulas in Finance

Seminar in Financial Econometrics - LMU

Simone Benzi

January 2024

Abstract

This research article is geared toward student who has an essential background in Probability and Statistics, assuming the reader is not familiar with the main topic. The primary objective is to deliver a thorough introduction to Copulas, allowing the audience to delve deeper into this subject independently. The theory is allied with intuitions and examples in order to show how Copulas are applied to real world data. Throughout this dissertation I explained what Copula is, its main features and the estimation methods. Finally, I illustrated how to employ Copulas in Finance through R software. The main purpose of this example is to estimate the joint distribution of stock returns through a Survival Gumbel and T student Copulas. Additionally, I have simulated synthetic returns to assess the ability of the models to capture the behavior of historical data. I harvested Apple and Exxon returns, from 1 January 2019 to 18 December 2023. As anticipated, all the presented results stem from my research using R as statistical software. The main libraries I relied upon are VineCopula and Copula. Noteworthy is the intentional omission of detailed proofs, preventing the article from becoming excessively mathematical. Nevertheless, for interested readers, comprehensive proofs can be found in the provided references.

Contents

1 Introduction

1.1 Why Copulas are useful?

In financial markets, is well known that returns are neither normally distributed nor symmetric, and is not seldom to have dependent assets that exhibit tail dependence and outliers. In the context of credit risk, it is essential to model the tail dependence between defaults of different financial instruments, such as bonds or credit derivatives. Having said that, we can surmise that the modeling of **multivariate distributions** and their **dependence structure** is one of the most critical concerns in probability theory and financial applications. In this paper we are going to deepen as Copula functions unravel this two issues, namely how to gauge dependence and how to deal with multivariate distribution.
The main appeal of copulas is that by using them you can model the correlation structure and the marginals separately. This separation allows for more flexible modeling, as changes in marginal distributions or dependence structure can be addressed independently. In addition, different Copula families can capture various types of dependence, including positive or negative correlation, tail dependence, and asymmetry.

Moreover, from a probability theory point of view, this can be an advantage because for many combinations of marginals there no exist a closed form to generate the desired multivariate distribution, albeit there exist several univariate distributions that are parametric.
For instance, it is feasible to generate random samples from a joint normal distribution. In this case, each marginal is normal, and the multivariate distribution is fully characterized by its mean vector and covariance matrix. Therefore, one may carry out Cholesky decomposition to generate random samples from a multivariate. Notwithstanding, it's not straightforward to do the same when marginals are different, for instance Beta, Gamma, Weibull and Gumbel. In this case the joint distribution has to take into account manifold parameters and this usually lead to a complex dependence structures, which for non elliptical distribution is difficult to be described by a simple covariance matrix.

2 Useful facts in Probability and Statistics

2.1 Probability Overview

Definition 2.1.1: Cumulative Distribution Function represents the probability that a random variable is less or equal of a certain value.
It always takes values between zero and one.
It can be defined as : $(F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt)$

In the bivariate case, given (X_1, X_2), the 2-dimensional CDF is defined as:

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leq x_1, X_2 \leq x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{\mathbf{X}}(t_1, t_2)\,dt_2\,dt_1$$

In addition, we reflect on a further significant way of defining the CDF.
Regarding the bivariate case, since the latter is the probability that, at the same time, $P(X_1 \leq x_1, X_2 \leq x_2)$, we surmise that the CDF depends on the marginals of X_1, X_2 and their dependence structure.

Remark 1 : One may decompose a multivariate distribution as:

- Multivariate distribution function = Marginals + Dependence Structure

Definition 2.1.3: Probability integral transform. Let X being a random variable with cumulative distribution function $F_X(x)$, then the random variable $Y = F_X(X)$ follows a uniform distribution on the interval $(0, 1)$.
In other words: $Y = F_X(X) \sim U(0, 1)$.
Roughly speaking, if we apply the CDF as transformation to the original random variable, we obtain a new random variable that behaves uniformly in (0,1). The link with Copulas is to transform all random variables X by their $F_X(X)$ obtaining all uniform variables that contain the same information as the starting random variable. This method allows us to get rid of marginal distributions, obtaining only uniform random variables that preserve the dependence structure.

Definition 2.1.4: Inverse Probability integral transform. Let X being a random variable, then $X = F_X^{-1}(U)$ where U is a random variable uniformly distributed on the interval $(0, 1)$. In other words, the IPIT maps values from a uniform distribution to the corresponding values of the original random variable X through the inverse CDF.

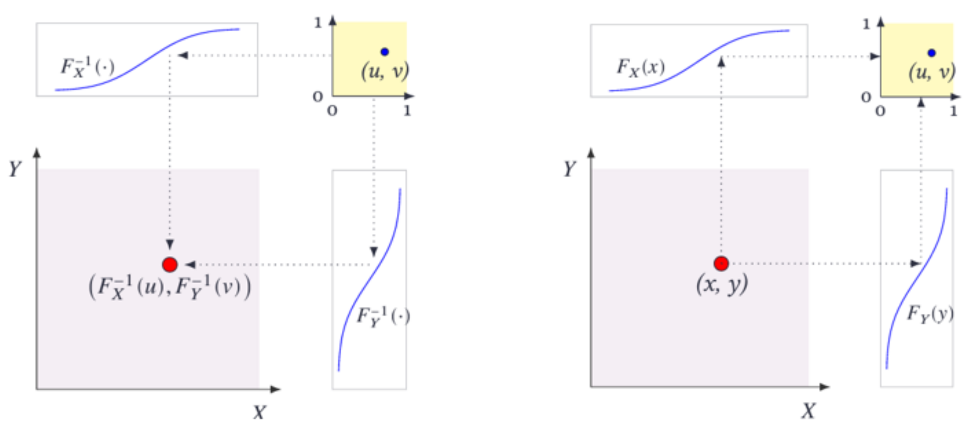

Figure 1: From the left: We can use the marginal inverse CDFs to map from (u, v) to $(F_X^{-1}(u), F_Y^{-1}(v))$. Conversely, the second picture show how can use the marginal CDFs to map from (x, y) to a point (u, v) on the unit square.

2.2 Measures of dependence

In quantitative finance, correlation is employed to construct well-diversified portfolios, manage risk, optimize asset allocation, hedging and algorithmic trading. Moreover, in financial econometrics and stochastic calculus the dependence is often introduced by instantaneous Brownian shocks, which are Gaussian. Two different types of dependency measures are covered in this section: linear correlation and rank correlation.
For a couple of random variables, each of these dependent measurements produces a scalar measurement, even though each case has unique characteristics.

2.2.1 Pearson Correlation

The Pearson correlation coefficient (ρ) between two random variables X and Y is given by the formula:

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

where $\mathrm{cov}(X,Y)$ is the covariance between X and Y, and σ_X and σ_Y are the standard deviations of X and Y, respectively.

It's a measure of linear dependence and it assumes that:

- Data is normally distributed;
- The relation among the variables is linear;
- There are no outliers.

From these hypothesis we can deduct that it gives a comprehensive characterization of dependence only in the multivariate Normal case where zero correlation also implies independence. It's also suitable in some non-Gaussian situations, such as elliptical distribution.
However, the stylized facts in financial markets are against these assumptions, hence for heavy-tailed distributions and for nonlinear dependence, it may produce misleading results. Furthermore, it is not invariant under monotone transformations of original variables, making it inadequate in many cases. In recent years a number of alternatives have been proposed such as rank correlation measures.

2.2.2 Rank Correlation: Spearman and Kendall tau

Alternative dependence measures that do not suffer from the preceding assumptions are Kendall tau and Spearman rho. Essentially, they are pairwise measures of concordance based on ranks, which make no assumptions about the marginal distributions. This is the crucial connection with Copulas.

Definition: Spearman's rank correlation coefficient (ρ) between two variables X and Y with ranked data is given by:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where d_i is the difference between the ranks of corresponding observations, and n is the number of observations. It gauges the strength and direction of the monotonic relationship between two variables. It ranges from -1 to 1. A monotonic relationship between two variables entails that as one variable increases (or decreases), the other variable tends to consistently increase (or decrease) as well but not at a constant rate.

Definition: The Kendall's tau for two variables X and Y is given by:

$$\tau = \frac{\text{Number of concordant pairs} - \text{Number of discordant pairs}}{\frac{1}{2}n(n-1)}$$

where n is the number of observations. Thus, it measures the similarity in the ordering of data points between two variables. It ranges from -1 to 1. If, for instance, we obtain $\tau =$ - 0.2, It means that, on average, as X increases, there tends to be a slight decrease in Y.

Let's just recapitulate the essential point. Rank correlation gauges the strength of the 'direction' of two random variables. It's more rough than the linear correlation. Yet, Spearman's rho and Kendall tau measures are more robust and suitable to model non linear dependence. In addiction they don't hinge upon the shape of marginal distributions of the random variables. The latter is a paramount connection with Copulas.

3 Copulas function

In order to have a quick grasp of Copula, imagine you're building a house. The margins are akin the foundational pillars of the house, whereas the Copula is the blueprint that dictates how these pillars are connected. Regardless of the specific choice of materials (marginal distributions) you use for the pillars, the blueprint (copula) remains consistent in connecting them. Moreover, once you pin down a blueprint (Copula) you can change the materials (marginals) without modifying the Copula structure **(see Thr 3)**.

Now, we can kick off reasoning in financial terms. Let's think about the returns of two stocks, like Apple and Exxon. Then, imagine to plot them on a scatter plot. The result can be seen as a simultaneously interactions of the **dependence structure** among the two stocks and their own **marginal distributions.** This is consistent with reality owing to a stock price basically depends on its core business (idiosyncratic risk) and the market condition (market risk). If we change the correlation, data will change, as well as if we modify the marginals, the data set will have different values.
Therefore, the main appeal of Copulas is that by using them, you can model the correlation structure of the two stocks and their marginals separately. Moreover, let's mull over the previous decomposition of multivariate distribution **Remark 1.** Linking these two ways of reasoning, we can infer that Copulas can be seen as a multivariate distribution function, where Copulas depicts the correlation structure. From the above idea, we can surmise that this separation allows for more flexible modeling, as changes in marginal distributions or dependence structure can be addressed independently.

3.1 Copulas and its properties

Definition 3.1: Copula is a multivariate cumulative distribution function for which the marginal probability distribution of each variable is uniform on the interval [0, 1]. In probabilistic terms, a Copula, denoted as C, is a map from $[0,1]^d \to [0,1]$.
The common notation for a copula is:

$$C(u_1, u_2, u_3, \ldots, u_m) = P(U_1 \leq u_1, \ldots, U_d \leq u_d)$$

such that $u_i \sim \text{Uniform}(0,1)$.
Basically, it receives a vector of uniform marginals and returns a value between zero and one. Hence, any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula, where the latter describes the dependence structure between the two variables.

Theorem 1: Sklar Theorem: Let F be a multivariate distribution function with margins $F_1, \ldots, F_d$. Then, there exists a copula C such that:

$$F_{XY}(x_1, x_2, \ldots, x_d) = C\{F_1(x_1), F_2(x_2), \ldots, F_d(x_d)\}, \quad x_1, \ldots, x_d \in R.$$

Moreover, if F_i are continuous, then C is unique. Therefore each joint distribution $F_{XY\ldots}$ can be written as a copula function $C(F_1(x_1, F_2(x_2), \ldots)$ taking the marginal distributions as arguments, and vice versa, every copula function taking univariate distributions as arguments yields a joint distribution.
Proof Thr 1. (only for $F_1, \ldots, F_d$ continuous):

$$\begin{aligned} F(x_1, \ldots, x_d) &= P(X_1 \leq x_1, \ldots, X_d \leq x_d) \\ &= P(F_1(X_1) \leq F_1(x_1), \ldots, F_d(X_d) \leq F_d(x_d)) \\ &= C(F_1(x_1), \ldots, F_d(x_d)) \end{aligned}$$

Notwithstanding, the interpretation of **Sklar Theorem** may be not straightforward in the multivariate case, owing to an objective difficult to imagine how a n-dimensional joint distribution works. Let's break down it in the two-dimensional case. Assume $X \sim$ **Gumbel and** $Y \sim$ **Gamma**.

Problem: **We don't know a priori their joint distribution** $F_{XY}(x, y)$, since it is challenging to gauge because the latter has to take into account manifold parameters and this usually lead to a complex dependence structure.

Solution: Yet, we know their individual cumulative distribution functions (CDF). Thus, we need a tool able to tie together the univariate CDF.
This is exactly what Copula does. It models the correlation structure and simultaneously gathers the univariate CDFs $\to C\{F_1(x_1), F_2(x_2)\}$ in order to obtain a representation of a multivariate distribution that we didn't know a priori.

Remark 2: For an arbitrary continuous multivariate distribution, we can determine its Copula by applying Definition 2.1.4:

$$C(u_1, \ldots, u_d) = F\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\}, \quad u_1, \ldots, u_d \in [0, 1]$$

For the next definition, we recall that the density (pdf) is indeed the derivative of the cumulative distribution function (CDF). Mathematically, the relationship is given by:

$$f(x) = \frac{d}{dx}F(x)$$

Definition 3.2: The Copula density and the **density of the multivariate distribution with respect to copula** are:

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}, \quad u_1, \ldots, u_d \in [0, 1],$$

$$f(x_1, \ldots, x_d) = c\{F_1(x_1), \ldots, F_d(x_d)\} \prod_{i=1}^{d} f(x_i).$$

The copula c models the joint dependence between the variables, while the marginal densities f_i describe the individual behavior of each variable.

Theorem 2: Fréchet-Hoeffding bounds: Let C be a copula, then we can bound it as:

$$\max\left(1 - d + \sum_{i=1}^{d} u_i, 0\right) \leq C(u_1, \ldots, u_d) \leq \min\{u_1, \ldots, u_d\}, \quad \forall u \in [0, 1]^d$$

The upper bound represents a copula with stronger positive dependence, whereas the lower bound represents a copula with stronger negative dependence. The value of d depicts the dimension of the Copula.

Theorem 3: Invariant Principle: Given a random vector $X = (X_1, \ldots, X_d)$ one can transform it, by Definition 2.1.3, to $U = (F_1(X_1), \ldots, F_d(X_d))$ without changing the Copula.
Therefore, we study the dependence between the components of X by studying the dependence between the components of U, independently of the marginals $F_1, \ldots, F_d$.

Let's attempt to combine **Thr 1. and Thr 3.**
We have shown that Copula remains constant even when individual distributions change, enabling us to adapt models without affecting the dependency structure. Furthermore, let X and Y be two correlated stocks.
Let assume that $X \sim$ Log-Normal and $Y \sim$ T-student, and that a Clayton Copula is suitable to model their dependency.
According to **Sklar's Theorem**, we can construct their joint distribution function $F_{XY}(x, y)$ using the Clayton copula and the marginal distributions.
However, if we afterwards discover that X actually follows a Pareto distribution, only the marginal for X needs to be updated. The Clayton copula capturing the dependency between X and Y remains valid.

3.2 Copulas Families

Different Copula families can capture various types of dependence, including positive or negative correlation, tail dependence, and asymmetry. In this section we are going to present the two main group of Copulas, their features and properties.

3.2.1 Elliptical Copulas

Intuitively, an elliptical distribution can be seen as any probability distribution that generalize the multivariate normal distribution. Elliptical copulas are defined for the aforementioned family of distribution. The key advantage of them is that the measure of dependence is fully determined by the correlation matrix. On the other hand, a disadvantage is that they typically don't have a simple closed-form expressions. The most commonly used elliptical distributions are the Gaussian and Student-t distributions.

Definition 3.2.1 Gaussian Copulas is expressed as:

$$C_{\text{Gauss } \Sigma}\,(u_1, u_2, \ldots, u_d;) = \Phi_\Sigma\{\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_d)\}$$

Φ_Σ is the multivariate Gaussian distribution function with positve semidefinite correlation matrix Σ, and Φ^{-1} is the inverse standard normal CDF.

Definition 3.2.2 T-Student Copula is defined as:

$$C_{\text{T-Student } \Sigma}(u_1, u_2, \ldots, u_d; \nu) = T_\nu\{t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_d)\}$$

where T_ν is the multivariate t-Student distribution function with ν degrees of freedom, t_ν^{-1} is the inverse of the t-Student cumulative distribution function, and Σ is the correlation matrix. It is used to model the joint distribution of random variables with symmetrically heavy tails and is meaningful when dealing with data that deviates from normality, exhibiting fat tails or outliers. The parameter ν controls the tail thickness, with higher values indicating lighter tails.

Breaking Down Gaussian Copulas: An Example

Let suppose to have Apple and Exxon stocks, each with a t-distribution for returns:
Exxon: - Mean: 0.04 - Scale parameter (σ^*): 0.15 - (df): 8
Apple: - Mean: 0.07 - Scale parameter (σ^*): 0.17 - (df): 5

In addiction, we take heed that scale parameter for the t distribution is not the standard deviation, as these are based also on degrees of freedom. The exact relationship is that $s = \sigma \left(\frac{df-2}{df}\right)^{0.5}$, where σ is the standard deviation, s is the scale, and df is the degrees of freedom.

Problem: Assume we want to employ a normal copula, with the stock's correlation matrix $\Sigma = 0.7$. **What is the probability that both assets produce a loss (namely the returns are less than zero)?**

Solution:

- **Step 1:** Compute the probability that Exxon and Apple's returns are less than zero according to the Student's t-distribution.
$\mathbf{P}_{\mathbf{Exxon}}(0) : \left(0 - \frac{0.4}{0.15}\right) = -0.2667 \quad \rightarrow \quad F_{\mathbf{t\text{-}stud}_8}(-0.2667) = 0.3982$
$\mathbf{P}_{\mathbf{Apple}}(0) : \left(0 - \frac{0.7}{0.17}\right) = -0.4118 \quad \rightarrow \quad F_{\mathbf{t\text{-}stud}_5}(-0.4118) = 0.3488$**.**

- **Step 2: Convert these probabilities to standard normal variables, so** $\Phi^{-1}[P_{\mathbf{Exxon}}(0)] = -0.2579$ **and** $\Phi^{-1}[P_{\mathbf{Apple}}(0)] = -0.3886$**.**

- **Step 3: Use these as inputs to a standard bivariate normal distribution, so** $\Phi_\rho\{\Phi^{-1}[P_{\mathbf{Exxon}}(0)], \Phi^{-1}[P_{\mathbf{Apple}}(0)]\} = \Phi_{0.7}\{-0.2579, -0.3886\} = 0.2522$**.**

This result entails that, according to all our hypothesis and data we posited, the joint probability that both asset produce a loss is around 25%.

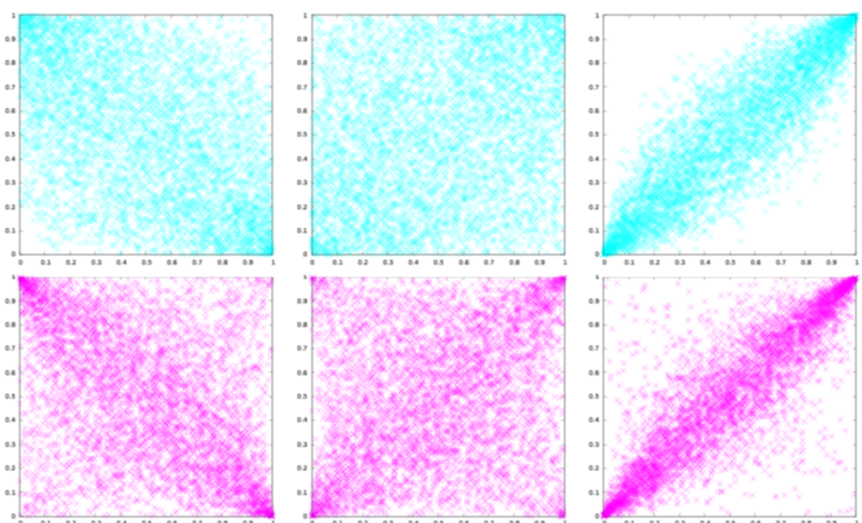

Figure 2: Simulation of Gaussian (up) and Student,= 2 (bottom) Copula, with correlation $-0.5(l)$, 0.3(m) and 0.9(r). As we can see the T Copula has fatter (symmetrical) tail than the Gaussian.

3.2.2 Archimedean Copulas

Archimedean copulas are a paramount class of copulas with a great quality: they can be expressed in closed form and are defined by using a generator function. Gumbel, Frank and Clayton are Copulas that belong to this class.
The general form of an Archimedean copula C is given by:

$$C(u_1, u_2, \ldots, u_n) = \psi^{-1}(\psi(u_1) + \psi(u_2) + \ldots + \psi(u_n))$$

where $u_1, u_2, \ldots, u_n$ are the marginal uniform distributions, and ψ is the Archimedean generator function.

Gumbel Copula:

$$C_{\text{Gumbel}}(u_1, u_2) = \exp\left(-\left((-\log u_1)^\theta + (-\log u_2)^\theta\right)^{1/\theta}\right)$$

Clayton Copula:

$$C_{\text{Clayton}}(u_1, u_2) = \max\left(u_1^{-\theta} + u_2^{-\theta} - 1, 0\right)^{-1/\theta}$$

Frank Copula:

$$C_{\text{Frank}}(u_1, u_2) = \frac{-1}{\theta}\log\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right)$$

In these formulas, θ is a parameter that controls the tail dependence of the Copula. Different values of θ yield copulas with different tail behaviors.
In most cases, Gumbel Copulas can be used to model the upper tail dependence or the max behaviour of a phenomena. Conversely, the Clayton one can be employed to manage left tail dependency.

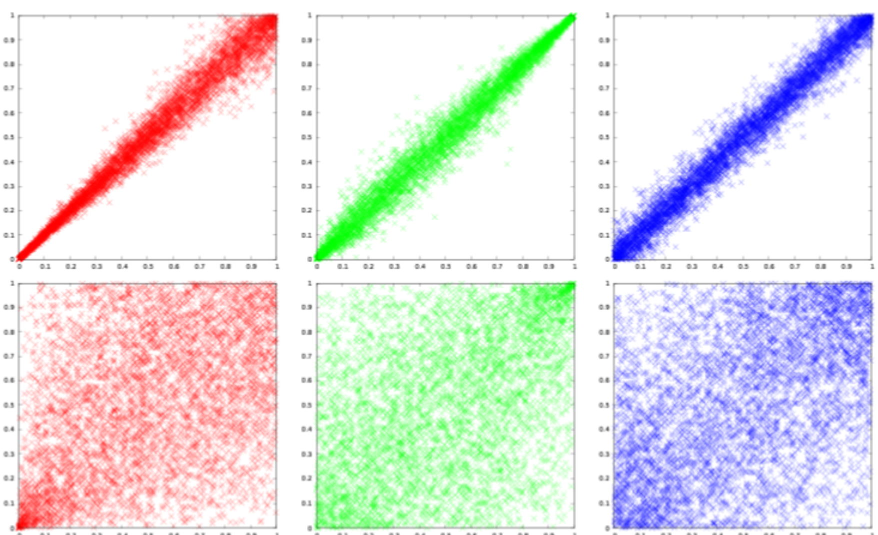

Figure 3: From the left, a simulation of Clayton,Gumbel and Frank with dependence measure of 0.9(up) and 0.3(bottom).

3.3 Estimation of Copula parameters

The literature on the Copula estimation methods is rather vast. Given that there exists several academic papers that deal with this topic, in this section we are presenting an introduction of two main methods. In general, the estimation involves both the estimation of the Copula parameters θ and the estimation of the marginals.

3.3.1 Kendall Tau method of moments

In the bivariate case, a standard method of estimating the univariate parameter is based on Kendall's statistic (see Genest and Rivest, 1993). For most copula functions with a single parameter there is a one-to-one relationship between and the Kendall's tau. This entails that from the latter, we can infer an estimate for the Copula parameter.

Theorem 4: Let (X_1, X_2) be a pair of random variable vector with continuous marginal cdfs F_1, F_2, and copula C, then:

$$\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2)\, dC(u_1, u_2) - 1$$

This theorem provides a direct link between Kendall tau and Copula. Both are measures of dependency and don't rely on marginal distributions since the integrand is expressed only in terms of the Copula, and the integral is defined in the unit square.

Greiner's Theorem 5: Consider a meta-Gaussian Copula (Gaussian Copula and arbitrary marginals) we can estimate its correlation matrix Σ of elements ρ_{ij}, then:

$$\tau(X_1, X_2) = \frac{2}{\pi} \arcsin(\rho_{ij})$$

This means that Kendall's can be used as estimator for the matrix Σ. Furthermore, the elements of the correlation matrix ρ_{ij} are defined as:

$$p_{ij} = \sin\left(\frac{1}{2}\pi\tau_{ij}\right)$$

Yet, there is no guarantee that this componentwise transformation of the matrix of Kendall's rank correlation coefficients will remain positive definite. In order to overcome this issue, there is a standard procedure that uses the eigenvalue decomposition to transform the correlation matrix into one that is positive definite. If Σ is not positive semidefinite, use Algorithm 5.55 from McNeil, Frey pag 231.

3.3.2 Maximum Likelihood Estimation

In Statistics, Maximum likelihood estimation (MLE) is a methodology used to estimate the parameters of a model. This approach seeks to find parameter values that maximise the probability that the data are actually observed.

Following Appendix B, let apply MLE in the context of Copulas.
Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_d)$ depicts the vector of parameters of marginal distributions and $\boldsymbol{\theta}$ parameters of the Copula. We want to estimate all the parameters **simultaneously**. Solving for the parameters involves numerical optimization, where we find the values that make the gradient (score) equal to zero.

The ML estimator $\hat{\boldsymbol{\eta}}$=($\boldsymbol{\alpha}, \boldsymbol{\theta}$) solves the system:

$$\frac{\partial \mathcal{L}(\boldsymbol{\eta}, \mathbf{X})}{\partial \boldsymbol{\eta}^{\top}} = \mathbf{0}$$

where:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\eta}, \mathbf{X}) &= \sum_{i=1}^{n} \log \left[c \left\{ F_1\left(x_{1i}, \boldsymbol{\alpha}_1\right), \ldots, F_d\left(x_{di}, \boldsymbol{\alpha}_d\right), \boldsymbol{\theta} \right\} \prod_{j=1}^{d} f_j\left(x_{ji}, \boldsymbol{\alpha}_j\right) \right] \\
&= \sum_{i=1}^{n} \left[\log c \left\{ F_1\left(x_{1i}, \boldsymbol{\alpha}_1\right), \ldots, F_d\left(x_{di}, \boldsymbol{\alpha}_d\right), \boldsymbol{\theta} \right\} \right. \\
&\left. + \sum_{j=1}^{d} \log f_j\left(x_{ji}, \boldsymbol{\alpha}_j\right) \right]
\end{aligned}$$

In line to the classical properties of ML estimation, the estimator is efficient and asymptotically normal. Moreover, plugging back the estimates into the L($\boldsymbol{\eta}$, X) we obtain a value of the log likelihood function. This can be used to compute AIC, BIC and the Likelihood ratio test. Notwithstanding, a drawback of this method is that it is often computationally demanding to solve, especially when we deal with several distributions. The main issue is that both Copula and marginals parameters are gauged simultaneously. For comprehensive presentation of MLE for each class of Copulas see (Cherubini, Luciano, Vecchiato, 2004, Chapter7)

Alternatively, in order to cope with this flaw, we can employ **Inference of Margins** estimation. It is basically a two stages optimization.
First, we estimate separately the parameters of the margins, and then use them in the estimation of the Copula parameters as known quantities. The above optimization problem is then replaced by:

$$\left(\frac{\partial \mathcal{L}_1}{\partial \boldsymbol{\alpha}_1}, \ldots, \frac{\partial \mathcal{L}_d}{\partial \boldsymbol{\alpha}_d}, \frac{\partial \mathcal{L}_{d+1}}{\partial \boldsymbol{\theta}} \right) = \mathbf{0}$$

where:

$$\begin{aligned}
&\mathcal{L}_j = \sum_{i=1}^{n} l_j\left(\mathbf{X}_i\right), \text{ for } j = 1, \ldots, d+1, \\
&l_j\left(\mathbf{X}_i\right) = \log f_j\left(x_{ji}, \boldsymbol{\alpha}_j\right), \text{ for } j = 1, \ldots, d, i = 1, \ldots, n, \\
&l_{d+1}\left(\mathbf{X}_i\right) = \log\left[c\left\{F_1\left(x_{1i}, \boldsymbol{\alpha}_1\right), \ldots, F_d\left(x_{di}, \boldsymbol{\alpha}_d\right)\right\}\right], \text{ for } i = 1, \ldots, n.
\end{aligned}$$

The first d components correspond to the usual ML estimation of the parameters of the marginal distributions. The last component reflects the estimation of the Copula parameters. Detailed discussion on this method could be found in Joe (1997).

4 Sampling from Gaussian and T Student Copulas

Throughout this article, we outlined that a Copula can be seen as a joint cumulative distribution which ties altogether different univariate marginals. Therefore, to figure out how a given Copula (with arbitrary marginals) behaves in practise, it is extremely useful to sample from it, namely to simulate its values. This procedure is beneficial for scenario analysis and model validation. Sampling from a Copula allows to compare synthetic (i.e. simulated) data generated using the Copula with observed data. If the simulated data closely matches the historical data, it suggests that the Copula captures the dependence structure effectively.

4.1 Algorithm for Sampling

In this section, we present algorithms for sampling from Gaussian and T-student Copula. For Archimedean Copulas, where ϕ is the generator, the simulation method is carried out by Laplace-Stieltjes transformation: $\int_0^\infty e^{-tx}\, dF(x)$. For a comprehensive explanation, refer to Marshall and Olkin (1988) and Nolan (2010).

4.2 Practical Example

Sampling from Normal Copula: the input of the simulation is the correlation matrix Σ. The normal Copula can be simulated by the following steps:

1. Generate a multivariate normal vector $\boldsymbol{Z} \sim N(0, \Sigma)$, where Σ is an m-dimensional correlation matrix. This step can be carried out by the Cholesky decomposition of the correlation matrix $\Sigma = LL^T$, where L is a lower triangular matrix with positive elements on the diagonal. If $\tilde{Z} \sim N(0, I)$, then $L\tilde{Z} \sim N(0, \Sigma)$. See Appendix B.

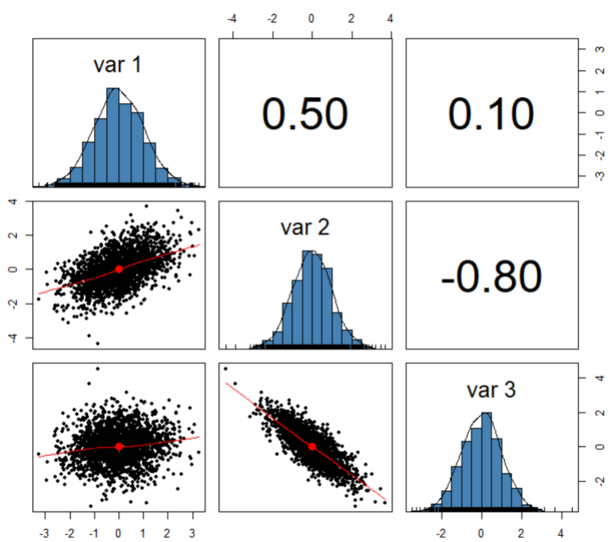

Figure 4: Simulation of a three-dimensional multivariate normal, $(0.5, 0.10, -0.6)$ are pairwise correlation values we set. In doing that one must specify a positive definite correlation matrix.

2. Transform the vector $\boldsymbol{Z}$ into $\boldsymbol{U} = (\Phi(Z_1), \ldots, \Phi(Z_m))$, where Φ is the distribution function of the univariate standard normal. By means of the Gaussian CDF, we get rid of marginals and have a pure representation of the dependence structure, as shown by the scatterplot. This is essentially a Gaussian Copula representation.

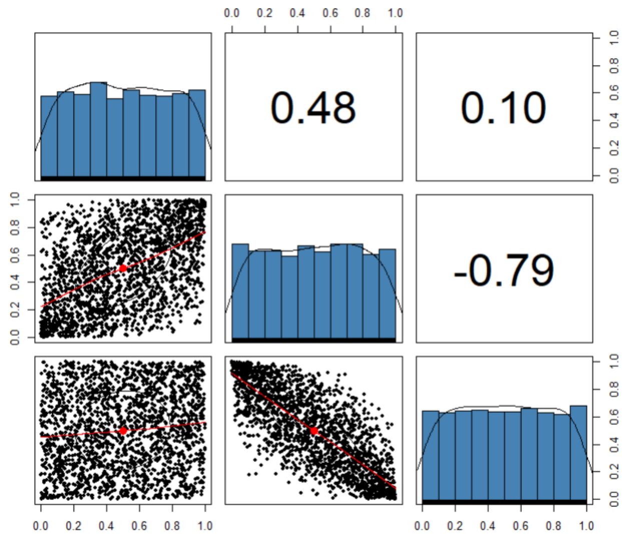

Figure 5: Each distribution is uniformly distributed in [0, 1]. It's paramount to recognize that the correlation remains the same. The applied transformation did not alter the correlation structure among the random variables. Essentially, what remains is the pure dependence structure (scatter plots).

3. Transform the uniform vector in the marginal you want applying the inverse CDF: $(\mathcal{F}_1^{-1}(u_1), \ldots, \mathcal{F}_n^{-1}(u_n))$.

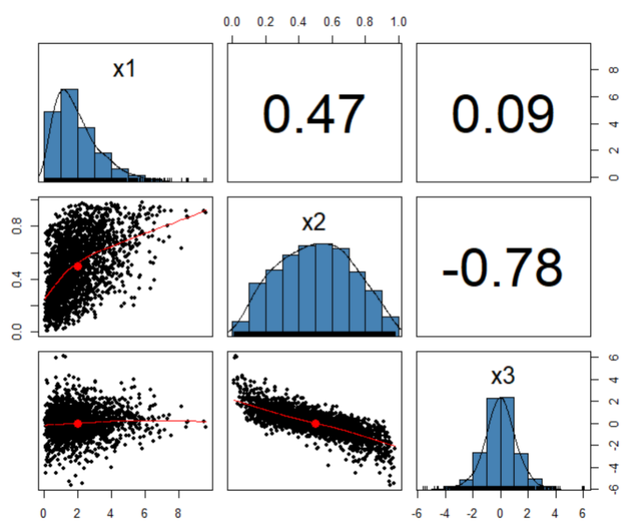

Figure 6: Inverse CDF (definition 2.1.4) applied to the uniform (0,1) data.
For this example we posit arbitrary distributions: Gamma (2,1), Beta(2,2) and T-stud (5).

To sum up, we define a correlation structure of our choice. Then we exploit the fact that normal random variables are feasible to simulate with Cholesky decomposition. We applied integral probability transform to factor out the dependence on gaussian marginals, but we preserve the **same** correlation structure. Finally, by applying the univariate inverse CDF, the result is a simulation of the distribution we selected linked by the initial correlation structure. To utterly appreciate it from a graphical point of view, one can compare scatter plot of figure 4 and 6.

Sampling from T-student Copula: the input parameters for the simulation are (v, Σ). The t copula can be simulated as follow:

1. Generate a multivariate vector $X \sim t_m(v, 0, \Sigma)$ following the centered t distribution with v degrees of freedom and correlation matrix Σ.
2. Transform the vector $\boldsymbol{X}$ into $\boldsymbol{U} = \left(t_v\left(X_1\right), \ldots, t_v\left(X_m\right)\right)^T$, where t_v is the distribution function of univariate t distribution with v degrees of freedom.
3. Transform the uniform vector in the marginal you want applying the inverse CDF : $(\mathcal{F}_1^{-1}(u_1), \ldots, \mathcal{F}_n^{-1}(u_n))$.

To simulate centered multivariate t random variables, you can use the property that $X \sim t_m(v, 0, \Sigma)$ if $\boldsymbol{X} = \sqrt{v/s}\boldsymbol{Z}$, where $\boldsymbol{Z} \sim N(0, \Sigma)$ and the univariate random variable $s \sim \chi_v^2$.

5 Sampling Apple and Exxon returns from Survival Gumbel and T student Copula

The main purpose of this example is to estimate the joint distribution of stock returns through a Survival Gumbel and T student Copulas employing Maximum Likelihood Estimation (see 3.3.2). In addition, we simulated synthetic returns to test whether the Copulas are able to capture the behaviour of historical data. Apple and Exxon returns, from 1 January 2019 to 18 December 2023 are considered. As financial markets plunge in periods of crisis, I have intentionally included the Covid pandemic in our data to assess whether our models allow for these fat tail events, sometimes called 'Black Swan'.

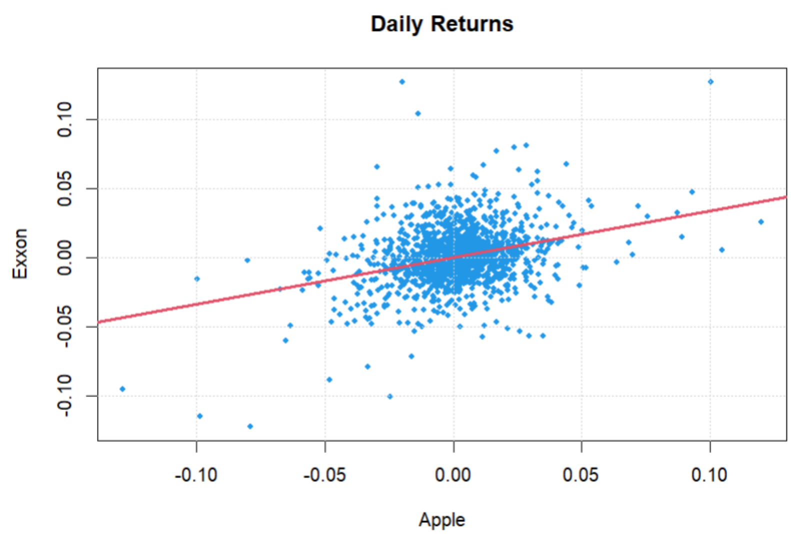

Figure 7: Apple and Exxon returns from January 1, 2019 to December 18, 2023.

In the previous illustration (4.2), the selection of a normal Copula model was executed without exhaustive deliberation. Nonetheless, practical applications of these models require an algorithm that pins down the most appropriate Copula. **VineCopula** package in R provides a valuable tool for Copula selection. The **BiCopSelect** function facilitates an informed decision-making process by systematically evaluating different Copulas. At first, for any kind of Copulas all parameters are estimated via MLE. Then the function elects the most appropriate model according to AIC and BIC. This approach ensures an efficient selection of the best model from a wide range of Copula families.

Given our data, the most deliberate Copula selected by the aforementioned algorithm is the **Survival Gumbel** with parameter $\hat{\theta} = 1.24$. In addition, estimation with the kendall's tau is feasible since this model hinges upon one single parameter, and it leads to an almost identical estimate for θ. For the sake of completeness, we estimate an additional model by the means of **BiCopSelect**, to single out the second most suitable Copula. It turns out to be a T-student Copula with parameters $\hat{\rho} = 0.25$ and df $= 5$.

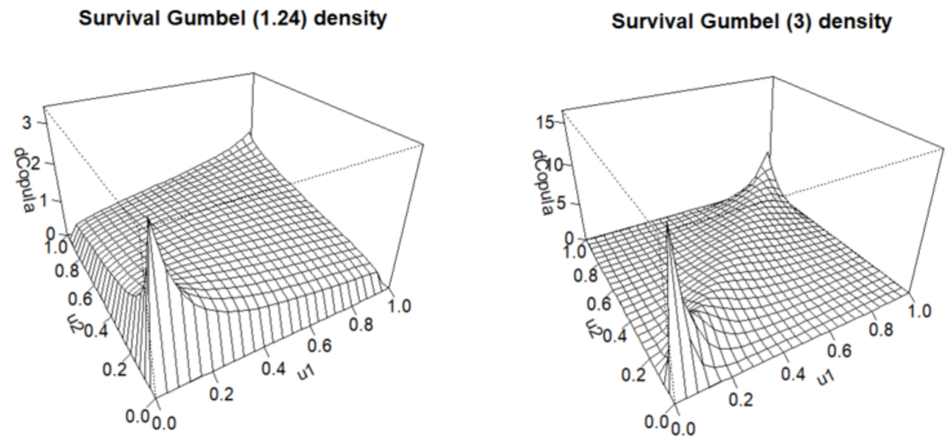

Figure 8: Survival Gumbel Copula density with parameter 1.24 and 3

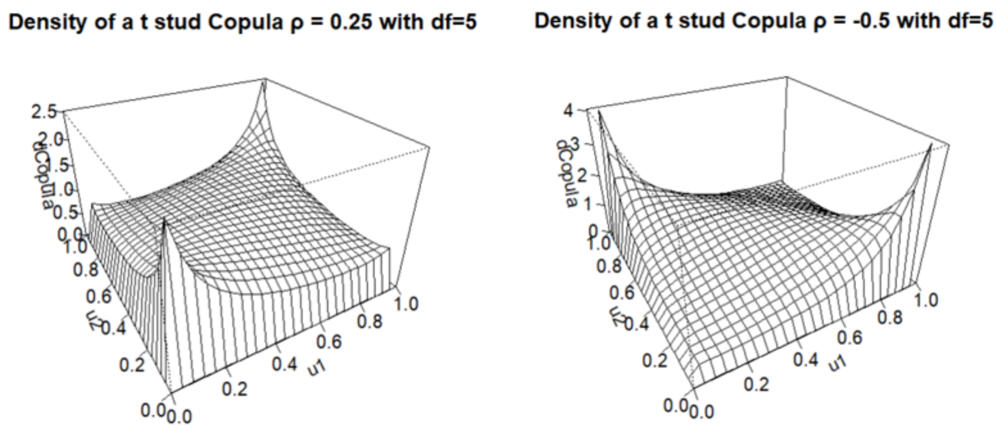

Figure 9: T student Copula density with df= 5 and $\hat{\rho}$ 0.25 and -0.5

The density sheds the light on the characteristics of the Copula . The Survival Gumbel Copula has harsh left tail dependency, whereas the T student Copula has a symmetrical tail dependency. Nonetheless, both copulas have fat tails. Furthermore, as $\hat{\theta}$ increases, the probability to experience extreme events is higher in the Survival Gumbel (see Figure 8). For the T Copula, as $\hat{\rho}$ increases in absolute value, the probability of experiencing extreme events also rises. When the parameter is negative, we basically invert the direction of dependency.(see Figure 9)

Moreover, I assumed that the t-student distribution would be a suitable marginal to model the returns of Exxon and Apple. MLE has been used to estimate the parameters based on our empirical dataset. Let's simulate the returns with the fitted distribution and compare them to empirical stock returns.

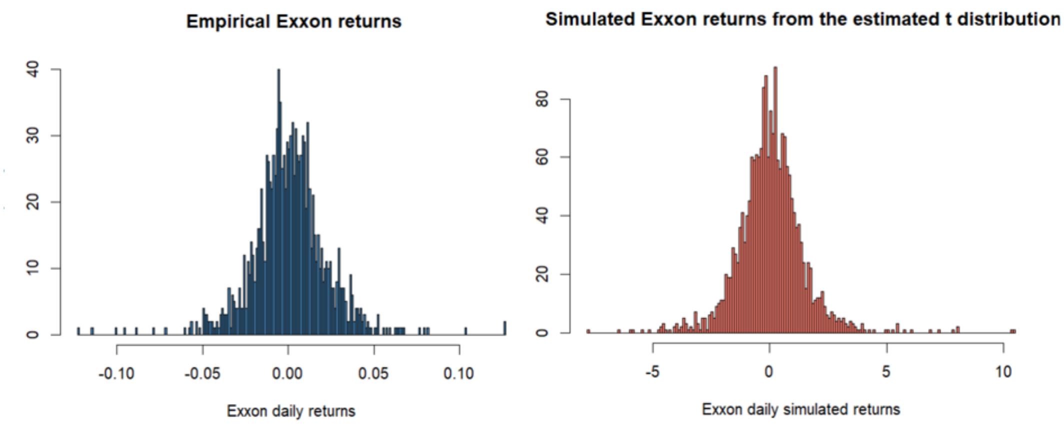

Figure 10: Empirical and simulated returns from a t student distribution

At this point we have both Copula and t marginals. Let's combine them altogether by Sklar's Theorem to have a representation of the multivariate distribution. We could see in practise that by the latter theorem each joint distribution $F_{XY\dots}$ can be written as a copula function $C(F_1(x_1, F_2(x_2), \dots)$ taking the marginal distributions as arguments.

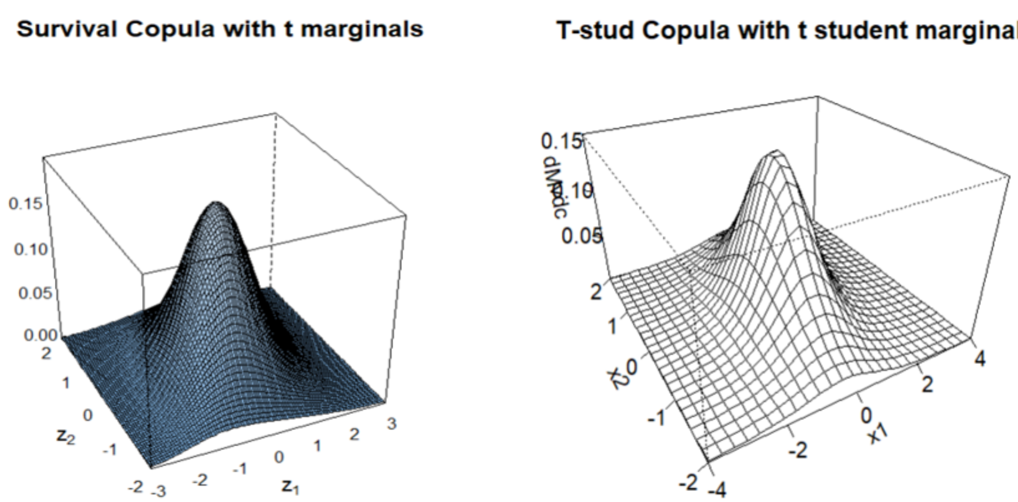

Figure 11: Multivariate distribution representation of Survival Gumbel ($\hat{\theta} = 1.24$) and T-student Copula (df$= 5$ and $\hat{\rho} = 0.25$) with the estimated t students marginals.

Last step regards the simulation from our Copulas to verify whether the simulated returns are in line with empirical stock data. Copula and VineCopula libraries in R don't allow to simulate from a Survival Gumbel Copula with t student marginals since the family of survivals has been recently implemented and this feature is not available yet. Despite it was seemingly the most suitable model, we simulated synthetic returns from the other T Copula with t marginals we presented throughout this article.

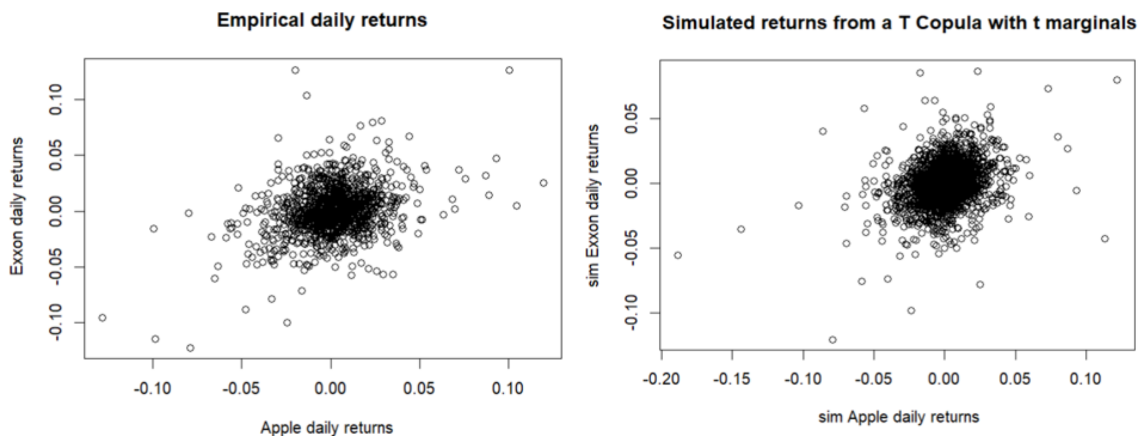

Figure 12: Apple and Exxon empirical and simulated returns from January 1, 2019 to December 18, 2023.

Let overlap the two graphs to have a more comprehensive interpretation.

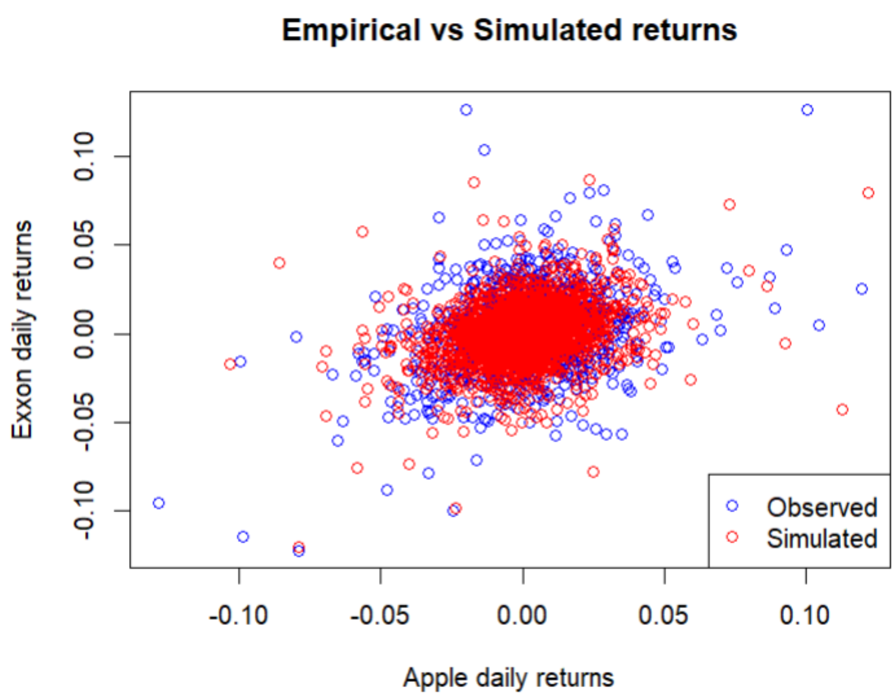

Figure 13: Apple and Exxon empirical and simulated returns from January 1, 2019 to December 18, 2023.

5.1 Conclusion

In conclusion, the estimated T Copula leads to results rather close to the actual observations. Nonetheless, there are two extreme negative observations where stocks sank simultaneously by -10% that were not captured by our model. Notwithstanding, our model seems to work adequately when Apple and Exxon plummet or soar by -5%. These features of the model stems from the fact the T Copulas tends has symmetrical tail dependence (see Figure 9). Therefore, my conclusion is that our model seems to behave appropriately when returns range from -5% to 5%. Furthermore, it also accounts for extremely negative returns, but no so frequently as seems to happen in financial markets. Nevertheless, I recall the main aim of this article is to strive to apply Copulas to financial data.

At this point, one could appreciate how flexible and powerful these functions are. We have modeled the joint behavior of stocks, which are rather rough random variables. A practitioner may enhance and employ this procedure for model validation, stress testing, risk management and pricing multi-asset derivatives. However, Copulas are not a panacea; they are a paramount tool in the statistical toolkit, but they are not a universal solution. Depending on the characteristics of the data and the specific research question, other methods might be more appropriate.

6 APPENDIX

6.1 Appendix A - Cholesky Decomposition

A square matrix A is said to have a Cholesky decomposition if it can be written as the product of a lower triangular matrix and its transpose. The lower triangular matrix is required to have strictly positive real entries on its main diagonal. For a given symmetric positive definite matrix , the Cholesky decomposition is expressed as:

$$A = LL^T$$

where:

- A is a symmetric positive definite matrix,
- L is a lower triangular matrix,
- L^T is the transpose of L.

The decomposition is not unique, as there are multiple ways to factorize A into LL^T where L is a lower triangular matrix.
Consider the symmetric positive definite matrix:

$$A = \begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix}$$

The Cholesky decomposition results in the lower triangular matrix L:

$$L = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix}$$

To verify the decomposition, let's check that $LL^T = A$:

$$L \cdot L^T = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 5 & 3 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix}$$

The product LL^T indeed equals the original matrix A.

6.2 Appendix B - Maximum Likelihood Estimation

Let θ be a vector of unknown parameters. When we have a set of jointly continuous random variables $X_1, X_2, X_3, \ldots, X_n$ we utilize the joint probability density function (PDF) to define the likelihood:

$$L(x_1, x_2, \ldots, x_n; \theta) = f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n; \theta)$$

For independent and identically distributed random variables, the likelihood simplifies to:

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$$

To enhance mathematical flexibility, we commonly apply the logarithmic transformation. This step preserves the structure of the likelihood, as the logarithm is a monotonically increasing function. Consequently, we work with the log-likelihood:

$$\ell(x_1, x_2, \ldots, x_n; \theta) = \log L(x_1, x_2, \ldots, x_n; \theta)$$

Then we calculate the partial derivatives with respect to the parameter and set them equal to zero. Solving for the parameters involves numerical optimization, where we find the values that make the gradient (score) equal to zero. These solutions provide the parameter estimates.

References

[1] McNeil, Rudiger Frey, Paul Embrechts, 2005, Quantitative risk management : concepts, techniques, and tools, by Princeton University Press

[2] Ruey S. Tsay, 2010, Analysis of Financial Time Series, Third Edition by John Wiley & Sons

[3] Härdle_Okhrin (2009), Modeling Dependencies with Copula

[4] Umberto Cherubini, Elisa Luciano, Walter Vecchiato, 2004, Copula Method in Finance by John Wiley & Sons

[5] R Copula Package, 2023-12-07, 'Multivariate Dependence with Copulas', <https://cran.r-project.org/web/packages/copula/copula.pdf>

[6] Roger B. Nelsen, 2010, An Introduction to Copulas by Springer

[7] R VineCopula Package, 2023-10-1, Statistical Inference of Vine Copulas, <https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf>

[8] Genest, C., & Rivest, L.-P. (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. Journal of the American Statistical Association

[9] Joe, H. (1997) Multivariate Models and Dependence Concepts. Monographs in Statistics and Probability. by Chapman and Hall, <http://dx.doi.org/10.1201/b13150>

[10] Marshall, A.W. and Olkin, I. (1988) Families of Multivariate Distributions. Journal of the American Statistical Association