UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea Magistrale in Matematica

Modelling Dependence of Stock Returns through Copulas

Supervisor:
Prof. Patrik Albin

Cosupervisor:
Prof. Giacomo Aletti

Elaborato finale di
Rossana D'Avino
Matricola 805584

Anno Accademico 2012/2013

Contents

Motivation

Imagine to own a portfolio P with only two constituents, say stock S_1 and stock S_2 [1], then the risk associated to P is directly related to the dependence between S_1 and S_2.
For that reason, selecting an appropriate model for dependence is of key importance for portfolio managment and portfolio selection.
Traditionally, the Pearson correlation coeffcient is used to describe dependence between two random variables X and Y with nonzero finite variances. The coefficient is defined as[2]

$$\rho_{XY} := \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$$

and it measures the strength and the direction of a *linear* relationship between X and Y.
Although profusely used because of the ease of which it can be calculated, the Pearson coefficient is neither a satisfactory nor a complete measure of the dependence among different securities an it is often a misunderstood measure of dependence for several reasons:

1. When the variance of returns tends to be infinite, that is, when extreme events are frequently observed, the linear correlation between these securities is undefined.

2. The correlation is a measure for linear dependence only.

3. The linear correlation is not invariant under nonlinear strictly increasing transformations, implying that returns might be uncorrelated whereas prices are correlated or vice versa.

4. Linear correlation only measures the degree of dependence but does not clearly discover the structure of dependence.

[1] S_1 and S_2 are random variables representing the two stock returns.

[2] $cov(V, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ is the covariance of (X, Y) whereas $Var(X)$ and $Var(Y)$ are the variances of X and Y.

5. It has been widely observed that market crashes or financial crises often occur in different markets and countries at about the same time period even when the correlation amog those markets is fairly low.

6. It is a natural scalar measure fully describing the dependence of (X, Y) when they are elliptical jointly distributed (such as bivariate normal), but it is totally wrong outside that case. Unfortunately, securities (S_1, S_2) often fail in being jointly elliptical distributed, thus using linear correlation as the only measure of dependence in such situations might yield misleading conclusions.

The latter point is related to the fact that correlation coefficient does not tell anything about dependence in the tails (that is when the two random variables reaches their extreme values), likely the Gaussian distribution, but in the real market instead securities often have strong tail dependence.
A more prelevant approach that overcomes the disadvantages of linear correlation is to model dependence by using copulas. With copula method the nature of dependence that can be modelled is more general as copulas offer much more flexibility then the correlation approach, in particular the dependence of extreme events can be considered.

Chapter 1

Bivariate copulas

The standard "operational" definition of a bivariate copula (briefely copula) is a bivariate distribution function defined on the unit square $[0,1]\times[0,1]$ with margins distributed as uniform on $[0,1]$. This definition however masks some of the problems one faces when costructing copulas using other techniques. For that reason, we start with a slightly more abstract definition and then prove that the "operational" one holds true.
Chapter 1 is therefore an introductory chapter to copulas and it is mainly organized in three parts: the first one defines the concept of bivariate copulas and points out first remarks; in the second part a bridge between copulas and distribution functions is built, so that the "operational" definition of copulas makes sense; finally, the third and last part of the chapter links bivariate copula with random variates by defining the copula associated with two given random variates.

1.1 Definitions

In order to define copulas, notions of *grounded* and *2-increasing* functions need to be introduced.
Let A_1 and A_2 be two nonempty subsets of $\overline{R}$ and $f : A_1 \times A_2 \longrightarrow \mathbb{R}$ a real function.

Definition 1.1. *If A_1 has a least element a_1 and A_2 has a least element a_2, the function f is said to be* grounded *if and only if*

$$f(x, a_2) = f(a_1, y) = 0 \quad \forall (x, y) \in A_1 \times A_2.$$

In other words, a two-place function f is grounded if it vanishes on the left and on the bottom boundary of its domain, i.e. on the set $\{a_1\} \times A_2$ and

$A_1 \times \{a_2\}$.

Definition 1.2. *f is said to be* 2-increasing *if and only if*
$\forall\ x_1, x_2 \in S_1$ *and* $\forall\ y_1, y_2 \in S_1 \times S_2$ *with* $x_1 \leq x_2, \quad y_1 \leq y_2$

$$f(x_1, y_1) + f(x_2, y_2) - f(x_1, y_2) - f(x_2, y_1) \geq 0.$$

Unlike groundedness, it is not so easy to recognize a 2-increasing function f just looking at its graph, but it can be proved that any 2-increasing and grounded function f is both nondecreasing in each argument and uniformely continous on its domain.
We are now ready to define copulas.

Definition 1.3 (bivariate copulas). *A bivariate copula (or briefly copula) C is any real function C defined on the unit square* $I^2 := [0, 1] \times [0, 1]$

$$C : I^2 \longrightarrow \mathbb{R}$$

fulfilling the following properties

i. *C is grounded;*

ii. $\forall u, v \in I \quad C(u, 1) = u$ *and* $C(1, v) = v$;

iii *C is 2-increasing.*

According to Definition 1.3 and the previous remark about grounded and 2-increasing functions, the graph of any copula is a continous surface within the unit cube I^3 i.e. $0 \leq C(u, v) \leq 1 \quad \forall (u, v) \in I$ whose boundary is the skew quadrilater with vertices $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, $(1, 1, 1)$ and it is a nondecreasing function in each place.
Theorem 1.1 guarantees that each copula C has is own graph liying between two main surfaces $z = max(u + v - 1, 0)$ and $z = min(u, v)$ called *Frchet-Hoeffding bounds.*

Theorem 1.1 (Frecht- Hoeffding bounds inequality)**.** *Let C be a copula. Then for every* $(u, v) \in I^2$

$$max(u + v - 1, 0) \leq C(u, v) \leq min(u, v).$$

Proof. From the nondecreasing property of C in both argument it follows that $\forall (u,v) \in I^2$ either $C(u,v) \leq C(u,1) = u$ and $C(u,v) \leq C(1,v) = v$, and so the second inequality yields $C(u,v) \leq min(u,v)$.
Now, from the 2-increasing property yields $\forall (u,v) \in I^2$ $0 \leq C(u,v) + C(1,1) - C(u,1) - C(1,v) = C(u,v) + 1 - u - v$ and so $C(u,v) \geq u + v - 1$ which, combined with $C(u,v) \geq 0$ becomes $C(u,v) \geq max(u+v-1,0)$. $\square$

It is tedious but not hard at all to prove that the two boundary functions

$$M(u,v) := min(u,v)$$

$$W(u,v) := max(u+v-1,0)$$

fulfill properties from *i.* to *iii.* in Definition 1.3 hence they are copulas. In particular, M is well known as *Frecht-Hoeffing upper bound* or shortly *the maximum copula*, whereas W is well known as *Frecht-Hoeffing lower bound* or *the minimum copula*.
Figures 1.1 and 1.2 report the Frecht-Hoeffing bounds graphs toghether with their level sets.

Note that, $\forall t \in I$ the points $(t,1)$ and $(1,t)$ are each members of the level set corresponding to the constant t. Hence we do not need to label the level sets in the diagram, as the boundary conditions $C(1,t) = t = C(t,1)$ readily provide the constant for each level set.
As a consequence of Theorem 1.1, the graph of the level set

$$\left\{(u,v) \in I^2 \ \ s.t. \ C(u,v) = t\right\}$$

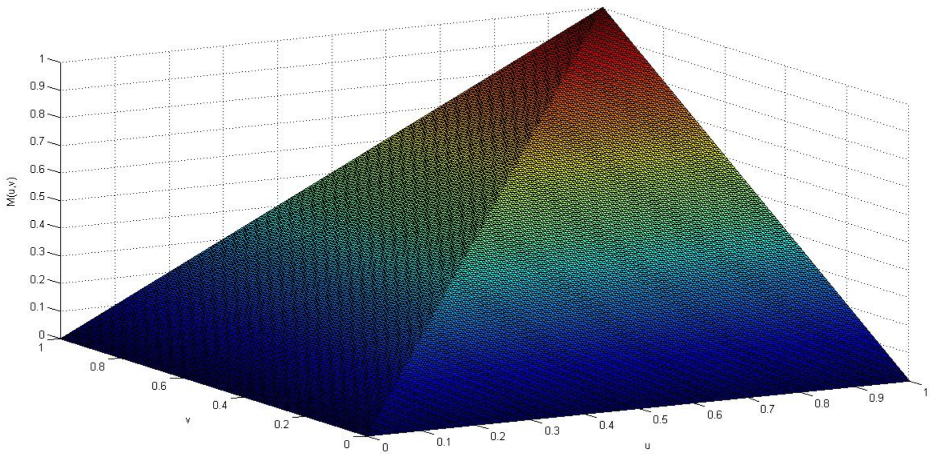
of any copula C at level $t \in I$ must lie in the shaded triangle drawn in Figure 1.3. Its hypotenuse is the level set $W(u,v) = t$, and the other two sides are the level set $M(u,v) = t$.

A third important copula frequently encountered is the *product copula*

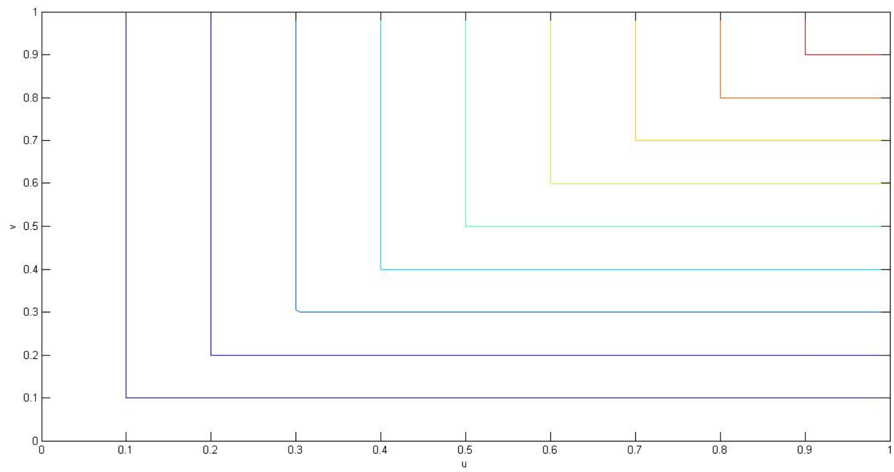
$$\Pi(u,v) := uv \quad \forall (u,v) \in I. \tag{1.1}$$

whose plot is drawn in Figure 1.4.

Chapeter 2 will clarify the reason why M, W and Π play an important role when studying the dependence of two random variates.

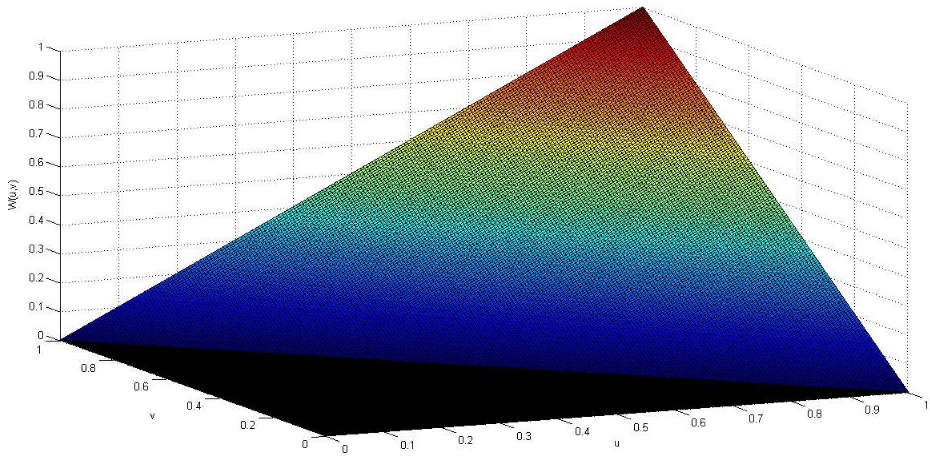


(a)

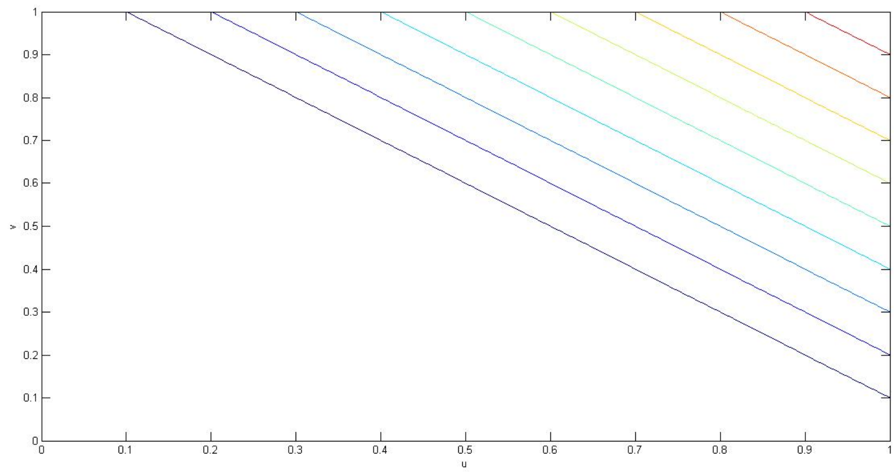


(b)

Figure 1.1: M copula and its level sets.



(a)



(b)

Figure 1.2: W copula and its level sets.

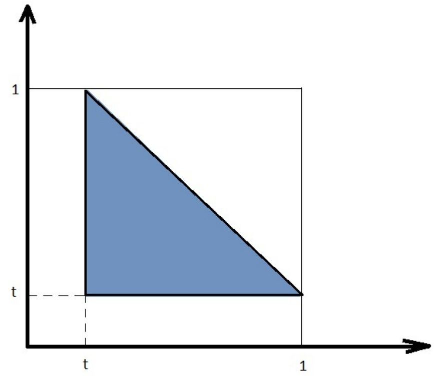


Figure 1.3: The region containing the level set $\{(u,v) \in I^2 \ s.t. \ C(u,v) = t\}$.

1.2 Copulas and distribution functions

At the very beginning of the chapter copulas have been alternatively defined as distribution functions with margins distributed as uniform on $[0,1]$. The followig section aims at proving this equivalence.
The *Easy implication* part shows why copulas are particular distribution functions with uniform margins on $[0,1]$ and the *easy* denotation stems from the fact that it trivially holds true.
The *Hard implication* part instead proves the converse, that each distribution functions with margins distributed as uniform on [0,1] can be viewed as copulas. The *hard* denotation underlines the non-triviality of the proof which needs instead the *Sklar's theorem* statement.

Easy implication

Let's recall first the definition of bivariate distribution function.

Definition 1.4. *A bivariate distribution function is any function two-place and real function*

$$H : \overline{\mathbb{R}}^2 \longrightarrow \mathbb{R}$$

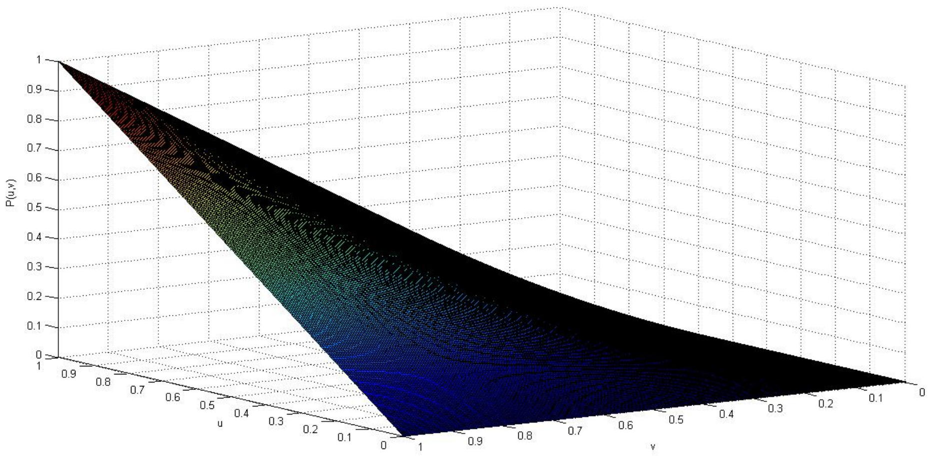
such that[1]

i. $\forall (x,y) \in \mathbb{R}^2 \quad H(-\infty, y) = H(x, -\infty) = 0$ *i.e. H is grounded;*

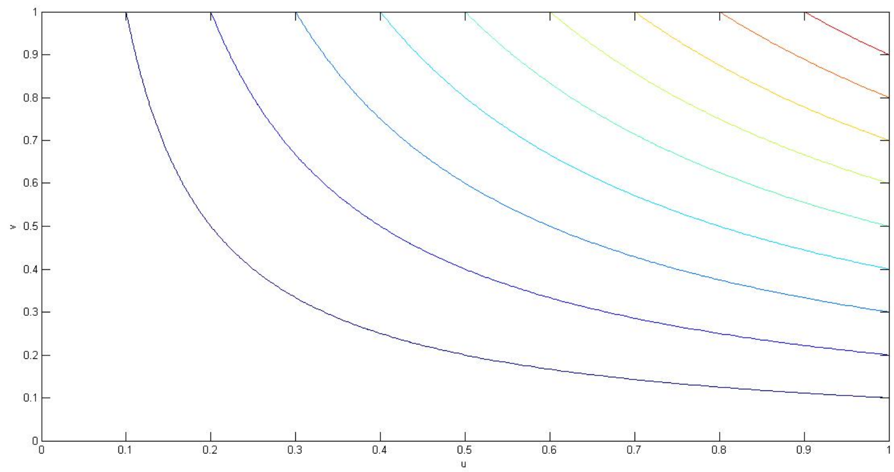
ii. $H(\infty, \infty) = 1$;

iii. *H is 2-increasing.*

[1] $H(\infty,\infty) := \lim_{x,y\to\infty} H(x,y)$.



(a)



(b)

Figure 1.4: Π copula and its level sets.

The margins of H are defined as

$$F(x) := H(x, \infty) \quad \forall x \in \mathbb{R}$$

$$G(y) := H(\infty, y) \quad \forall y \in \mathbb{R}$$

and it can be easily checked that they are univariate distribution functions.
If C is a copula and we extend its domain to $\mathbb{R}^2$ as follows
$\forall (x, y) \in \mathbb{R}^2$

$$H_C(x, y) := \begin{cases} 0 & \text{if } x \leq 0 \text{ or } y \leq 0 \\ C(x, y) & \text{if } (x, y) \in [0, 1] \\ x & \text{if } y \geq 1 \text{ and } x \in [0, 1] \\ y & \text{if } x \geq 1 \text{ and } y \in [0, 1] \\ 1 & \text{if } x \geq 1 \text{ and } y \geq 1, \end{cases}$$

then H_C satisfies conditions from *i.* to *iii.* in Definition 1.4, hence H_C is a distribution function and by definition its margins

$$F(x) = H_C(x, \infty) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x \in [0, 1] \\ 1 & \text{if } x \geq 1 \end{cases}$$

$$G(y) = H_C(\infty, y) = \begin{cases} 0 & \text{if } y \leq 0 \\ y & \text{if } y \in [0, 1] \\ 1 & \text{if } y \geq 1 \end{cases}$$

are distributed as uniform on $[0, 1]$. Thus, copulas are particular distribution functions with uniform margins in the sense that they are restrictions on I^2 of those ones.

Hard implication

Theorem 1.2 (Sklar's theorem). *Suppose F and G are two univariate and continous cdfs.*
Let $\mathbb{C}$ denote the entire class of copulas and $\mathbb{H}$ that one of joint cdfs with margins F and G.
Then the corrispondence $f : \mathbb{C} \longrightarrow \mathbb{H} : \; C \mapsto f(C) =: H$ defined as

$$H(x, y) = C(F(x), G(y)) \quad \forall (x, y) \in \mathbb{R}^2 \tag{1.2}$$

is a bijective function, i.e.

1. *f is well posed, that is H in (1.2) is a distribution function with margins F and G;*

2. *f is bijective, equivalently* $\forall H \in \mathbb{H} \;\; \exists! \; C \in \mathbb{C}$ *s.t. (1.2) holds.*

Hence, given any couple of continous margins, the above theorem states that there exists a one to one correspondence between copulas and cummulative distribution functions. In other words, given two continous margins F and G, each copula C univocally identifies a cdf H_C and viceversa each cdf H can be associated to only one copula C_H.
In particular, when F and G are $\sim U([0,1])$ Slar's theorem states exactly the result we were aiming at.
Notice that, being F and G continous, their inverse F^{-1} and G^{-1} make sense and equation (1.2) can be equivalently written as

$$C(u,v) = C(F^{-1}(u), G^{-1}(v)) \qquad \forall (u,v) \in [0,1]^2 \tag{1.3}$$

by setting $u := F(x)$ and $v := G(y)$.

Since copulas are distributions, they admit the notion of density. Precisely, if $C(u,v)$ is a copula, the density $c(u,v)$ associated with C is

$$c(u,v) := \frac{\partial^2 C(u,v)}{\partial u \partial v} \qquad \forall (u,v) \in I^2. \tag{1.4}$$

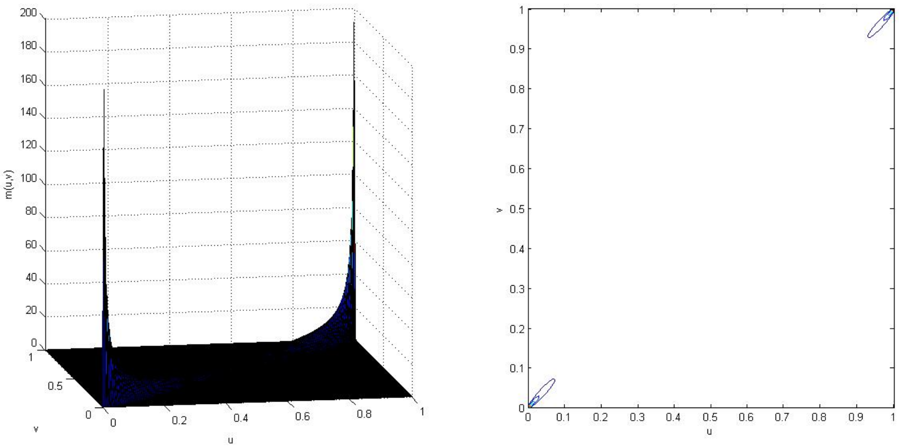
Hence, c is a non negative function on the unit square with zero value elsewhere on $\mathbb{R}^2$. As an example, Figure 1.5 illustrates the densities of M and W together with their level sets.

The name *copula* was choosen to emphasize the manner in which a copula *couples* a distribution function H to its univariate margins F and G. Indeed, while writing $H(x,y) = C(F(x), G(y))$ one splits the H into the margins and a copula, so that the latter only represents the "association" between F and G. Copulas separate margins behaviour from their association: at the opposite, the two cannot be disentagled in the usual representation of joint probabilities via distribution functions.
For that reasion, copulas are also called *dependence functions*.

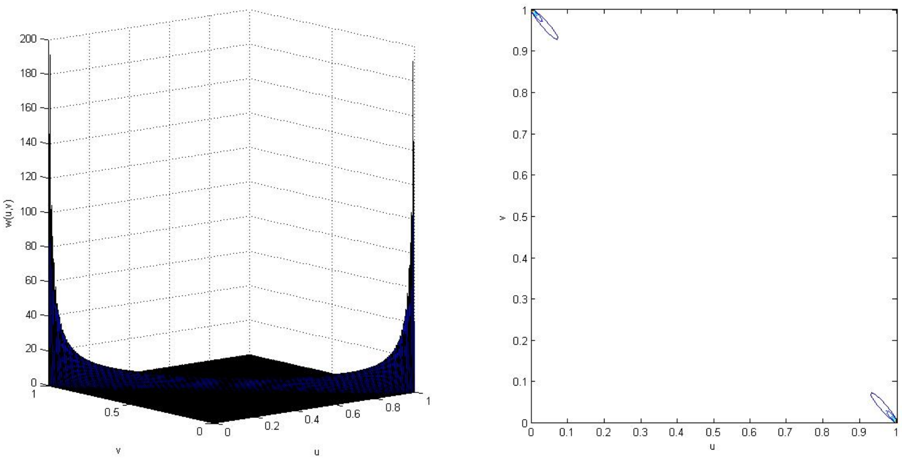
A function closely related to a given copula is its *survival copula*.

Definition 1.5. *The survival copula $\overline{C}$ associate with the copula C is*

$$\overline{C}(u,v) := u + v - 1 + C(1-u, 1-v) \qquad \forall (u,v) \in I^2. \tag{1.5}$$



(a) M copula density



(b) W copula density

Figure 1.5: Copula densities.

It is easy to prove that $\overline{C}$ verifies conditions from $i.$ to $iii.$ in Definition 1.3, hence it is a copula, and it therefore admits density

$$\begin{aligned}\overline{c}(u,v) &:= \frac{\partial^2 \overline{C}(u,v)}{\partial u \partial v} = \frac{\partial^2 [u+v-1+C(1-u,1-v)]}{\partial u \partial v} \\ &= c(1-u,1-v),\end{aligned}$$

thus $\overline{c}(u,v) = c(1-u,1-v)$.
Last equivalence clarifies - better than the expression of $\overline{C}$ in (1.5) did - the connection between a copula and its survival version: they are one the rotation of the other, or better said one the simmetrical of the other one.
Next chapter will give few examples of copulas and survival copulas so to better notice the high symmetry from their plots.

Again, since copulas are special distribution functions it is true that convex combination of two orcopulas are still copulas. Formally speaking, if $C_1, C_2 \dots C_n$ are n copulas and $\alpha_1, \alpha_2 \dots \alpha_n$ are real numebers in $[0,1]$ such that $\sum \alpha_i = 1$, then the convex combination

$$C(u,v) := \sum_{i=1}^{n} \alpha_i C_i(u,v)$$

is still a copula.

1.3 Copulas and random variables

Thanks to Skorohod representation there exsist a bijection between the cdfs H with margins F and G and the joint cdfs of $X \sim F$ and $Y \sim G$ that is

$$H(x,y) = \mathbb{P}(X \leq x, Y \leq y).$$

In the light of what has just been said, a new version of Sklar's theorem can be restated with random variables.

Theorem 1.3 (**Sklar's theorem**). *Suppose X and Y are two random variates with continous distributions F and G respectively. Let $\mathbb{C}$ denote the class of all copulas and $\mathbb{H}$ that one of all possible joint distribution functions of X and Y.*
Then the corrispondence $f : \mathbb{C} \longrightarrow \mathbb{H} : C \mapsto f(C) = H$ defined as follows,

$$H(x,y) = C(F(x), G(y)) \qquad \forall (x,y) \in \mathbb{R}^2 \tag{1.6}$$

is a bijective function.

The copula C just claimed is usually called *copula of X and Y* and denoted by C_{XY}.
Since (1.2) can be equivalently written as (1.3), C_{XY} is exactly the cdf of $(F(X), G(Y))$ where $F(X)$ and $G(Y)$ are uniform because of the continouity of X and Y. In conclusion, the bijective function f assignes to each H of (X, Y), the distribution function of the uniform coupled variates $(F(X), G(Y))$.

Much of the usefulness of copulas comes from the fact that for strictly monotone transformations of the random variables, copulas either are invariant or change in predictable ways, and that is what basically next theorem states[2].

Theorem 1.4. *Let X and Y be continous random variables with copula C_{XY}. If α and β are strictly increasing function on $Ran(X)$ and $Ran(Y)$ respectively, then*

$$C_{\alpha(X)\beta(Y)} = C_{XY} \tag{1.7}$$

When at least one of α and β is strictly decreasing we obtain results in which the copula of the random variables $\alpha(X)$ and $\beta(Y)$ is a simple transformation of C_{XY}. Specifically, we have

Theorem 1.5. *Let X and Y be continous random variables with copula C_{XY}. Let α and β be strictly monotone on $Ran(X)$ and $Ran(Y)$, respectively.*

1. *If α is strictly increasing and β is strictly decreasing, then*

$$C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v)$$

2. *If α is strictly decreasing and β is strictly increasing, then*

$$C_{\alpha(X)\beta(Y)}(u, v) = v - C_{XY}(1 - u, v)$$

3. *If α and β are both strictly decreasing, then*

$$C_{\alpha(X)\beta(Y)} = u + v - 1 + C_{XY}(1 - u, 1 - v).$$

[2] Recall that if X is a continous random variable with distribution function F and if α is a strictly monotone function whose domain contains $Ran(X)$, then $\alpha(X)$ is still a continous random variable.

It should be now more clear the reason why "... it is precisely the copula which captures those properties of the joint distribution function which are invariant under almost surely strictly increasing transformations" (Scheizer and Wolff, 1981).

Chapter 2

Copulas and dependence

Copulas play an important role when studying the relationship between two continous random variables X and Y and the following chapter helps in understanding how a copula can capture a particular dependence structure. Precisely, the study starts with analyzing the independence case, considered the easiest one, and then it moves to the dependence case, much more complex because of the huge variety of dependence structures existing. In his last part, the chapter introduces some of the most important copula families. From now on X and Y will be assumed continous throughout the work.

2.1 Independence

Let's recall the usual definition of independence.

Definition 2.1. *Suppose X and Y have joint distribution function H. The variables X and Y are said to be independent if and only if*

$$H(x, y) = F(x)G(y) \quad \forall(x, y) \in \mathbb{R}^2$$

The Sklar's theorem version (1.1) and the definition of *product copula* Π trivially yield

Corollary 2.1. *Suppose X and Y have joint distribution function H. The variables X and Y are independent if and only if*

$$C_{XY}(u, v) = \Pi(u, v) \quad \forall(u, v) \in I^2.$$

The Corollary states that the independent random variables are all and only those having as their copula the product copula $C_{XY} = \Pi$, hence the Π copula completely characterizes the independence structure.

2.2 Dependence

When X and Y are not independent, they are said to be *associated* or *dependent.*
Two random variates can be associated in several ways, each of them capturing a different meaning of dependence. Three examples of three different kinds of dependence structure are:

linear dependence which answers to the question “is Y a linear function of X?”;

monotone dependence -or *concordance*- which answers to “is Y a monotone function of X?”;

tail dependence which answers to “is there any relatioship between X and Y when they assume their extreme values?”.

Not every dependence structure are “scale-invariant”, in the sense that they remain unchanged under strictly increasing transformations of X and Y, for istance the linear dependence is one of those. All the other ones which are instead scale-invariant, thanks to Theorem (1.4), can be expressed in terms of C_{XY}.
The present work limits the study only to the scale-invariant dependence structures, in particular the concordance and the tail dependence, and it defines the indices common used to measure them.

2.2.1 Concordance

Informally, X and Y are said to be *concordant* if “large values” of one tend to occur with “large values” of the other, and viceversa “small values” of one with “small values” of the other. On the other hand, X and Y are said to be *discordant* if “large values” of one tend to occur with “small values” of the other, and viceversa “small values” of one with “large values” of the other.
In other words, concordance aims at capturing the probability of having large (or small) values of both X and Y is high, while the probability of having large values of X together with small values of Y is low -or viceversa-. Shortly, concordance is otherwise known as monotonicity.
To measure the degree of concordance between X and Y many indices can be defined but the two most known are the *Kendall's tau* and the *Spearman's rho.*

Kendall's tau

Let (X_1, Y_1) and (X_2, Y_2) be two iid copies of (X, Y).
The *Kendall's tau* coefficient of X and Y is defined as

$$\tau^{XY} = \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0) \in [-1, 1] \quad (2.1)$$

By defining two couples (a, b) and (c, d) concordand if and only if $a < c$ and $b < d$ - or $a > c$ and $b > d$ - and discordant if and only if $a < c$ and $b > d$ - or $a > c$ and $b < d$ -, the expression (2.1) is exactly the difference between the probability of concordance and the probability of discordance for (X_1, Y_1) and (X_2, Y_2).
Theorem 2.2 states how Kendall's tau coefficient of X and Y can be computed by C_{XY}

Theorem 2.2. *If X and Y have copula C_{XY}, then*

$$\tau^C = 4 \int\int_{I^2} C(u, v) dC(u, v) - 1.$$

By applying the previous theorem to $M(u, v) = min(u, v)$, $W(u, v) = max(u + v - 1, 0)$ and $\Pi(u, v) = uv$ copulas we find that

$$\tau^M = 1 \qquad \tau^W = -1 \qquad \tau^\Pi = 0.$$

To be more precise

$$\tau^{XY} = 1 \quad \text{if and only if } C_{XY} = M$$

$$\tau^{XY} = -1 \quad \text{if and only if } C_{XY} = W$$

Definition 2.2. *X and Y are said to be* comonotone *(*countermonotone*) if and only if $C_{XY} = M$ $(C_{XY} = W)$*

What does the above concepts mean in terms of dependence structure? What does it mean that two variets have M or W copula?
$C_{XY} = M$ if and only if Y is a.s. an increasing function of X, and $C_{XY} = W$ if and only if Y is a.s. a decreasing function of X. Thus, while we have previousely seen how Π copula characterizes the independence structure, here we have seen how M and W copulas characterize the perfect dependence structures, comonotonicity and countermonotonicity.
The two cases above are the only ones such that τ reaches its bounds (1 and -1), in all the other ones $\tau \in (-1, 1)$. We have now two equivalent ways to characterize perfect dependence, either by checking that the copula is one

of Frchet-Hoeffding bounds, or by checking that Kendall's tau reaches the bounds.
Moreover it is not true that $\tau = 0$ only in case in independence, indeed it might vanishes even though X and Y are dependent.

As for the survival copula, Theorem 2.3 proves that the Kendall's tau coefficient remain unchanged when considering the survival version of a copula

Theorem 2.3. *If C is a copula and τ^C represents its Kedall's tau coefficient, then*

$$\tau^C = \tau^{\overline{C}}.$$

Spearman's rho

Let (X_1, Y_1) (X_2, Y_2) and (X_3, Y_3) be three iid copies of (X, Y).
The *Spearman's rho* coefficient related to X and Y is defined as

$$\rho_S^{XY} = 3(\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0]). \quad (2.2)$$

From (2.2), Spearman's coefficient measures (up to a multiplicative constant) the difference between the probability of concordance and the probability of discordance for the random couples (X_1, Y_1) and (X_2, Y_3).
Also for Spearman's coefficient $\rho_S \in [-1, 1]$ and it reached the bounds if and only if X and Y are respectively counetrmonotonic and comonotonic random variates

$$\rho_S^{XY} = 1 \text{ if and only if } C_{XY} = M$$
$$\rho_S^{XY} = -1 \text{ if and only if } C_{XY} = W$$

The following theorem states how Spearman's rho coefficient of X and Y can be computed by C_{XY}

Theorem 2.4. *If X and Y have copula C_{XY}, then*

$$\rho_S^{XY} = 12 \int\int_{I^2} C(u, v) du dv - 3.$$

As for the survival copula, a risult similar to the Kendall's tau holds true also in this case: the Spearman parameter associated to a copula remain unchanged when considering the survival copula.

Theorem 2.5. *If C is a copula and ρ_S^C represents its Spearman's rho coefficient, then*

$$\rho_S^C = \rho_S^{\overline{C}}.$$

PQD/NQD

Given C_{XY}, is there any graphical tip suggesting the sign of Kendall's tau or Spearman's rho for the couple (X, Y)?
In this section we partially answer to that question introducing the concept of positive and negative quadrant dependence.
The random variables X and Y are said to be *positive quadrant dependent* - shortly *PQD* - and *negative quadrant dependent* - shortly *NQD* - if and only if respectively $\forall (u, v) \in I^2$

$$C_{XY}(u, v) \geq uv$$

$$C_{XY}(u, v) \leq uv$$

that is if the graph of C_{XY} lies completely above or below the Π one.
PQD implies the non-negativity of Kendall's tau and Spearman's rho coefficients, and NQD implies their non-positivity. In other words, if the graph of C_{XY} lies above the Π one, then X is likely an increasing function of Y; if the graph of C_{XY} lies instead below the Π one, then X is likely an increasing function of Y.

2.2.2 Tail dependence

Tail dependence looks at concordance in tail, or extreme values, of X and Y. To be more precise, while concordance describes how large (or small) values of one random variable appear with large (or small) of the other, tail dependence describes instead how extreme (either large or small) values of one tends to occur with extreme values (again either large or small) of the other.
In order to measure the degree of tail dependence of two random variates X and Y, two indices will be introduced, the *Upper tail* dependence coefficient λ_U and the *Lower tail* dependence coefficient λ_L.

Upper and Lower tail coefficients

The *upper tail dependence function* associated to X and Y is defined as the conditional probability that the distribution function of X exceeds the threshold v given that the distribution function of Y does, in symbols,

$$\Lambda_U^{XY}(v) := \mathbb{P}(Y \geq G^{-1}(v)|X \geq F^{-1}(v)) \in [0,1] \qquad \forall v \in [0,1].$$

From the definition follows that Λ_U^{XY} is a decreasing function on $[0,1]$ reaching its maximum value on the boundary point $v = 0$ where $\Lambda_U^{XY}(v) = 1$.
The *upper tail dependence parameter* is defined to be the limit (if it exists) of $\Lambda_U^{XY}(v)$ when v tends to one,

$$\lambda_U^{XY} := \lim_{v \to 1^-} \Lambda_U^{XY}. \tag{2.3}$$

X and Y are said to have no upper tail dependence if and only if $\lambda_U^{XY} = 0$, in all other cases they are said to have and the larger λ_U^{XY}, the stronger the dependence.
Analogousely, the *lower tail dependence function* associated to X and Y is defined as the conditional probability that the distribution function of X falls below the threshold v given that the distribution function of Y does, in symbols, $\forall v \in [0,1]$

$$\Lambda_L^{XY}(v) := \mathbb{P}(Y \leq G^{-1}(v)|X \leq F^{-1}(v)) \in [0,1].$$

Here Λ_U^{XY} is an increasing function on $[0,1]$ reaching its maximum value on the boundary point $v = 1$ where $\Lambda_L^{XY}(v) = 1$. The *lower tail dependence parameter* instead, is the limit (if it exists) of $\Lambda_L^{XY}(v)$ when v tends to zero,

$$\lambda_L^{XY} := \lim_{v \to 0^+} \Lambda_L^{XY} \in [0,1]. \tag{2.4}$$

X and Y are said to have no lower tail dependence if and only if $\lambda_L^{XY} = 0,$, in all other cases they are said to have and the larger λ_L^{XY}, the stronger the dependence.
Intuitively, the more X and Y are comonotone ($\tau^{XY},\ \rho_S^{XY} \ \to 1$) and so Y tends to be an increasing function of X, the more they are upper and lower tail dependent ($\lambda_U^{XY},\ \lambda_L^{XY} \to 1$); similarly, the more they are countermonotone ($\tau^{XY},\ \rho_S^{XY} \ \to -1$) and so Y tends to be a decreasing function of X, the less they are upper and lower tail dependent ($\lambda_U^{XY},\ \lambda_L^{XY} \to 0$).
However, the converse of the previuos remarks could not hold true because, indeed X and Y might have tail dependence and at the same time not be

comonotone or they might not have tail dependence and at the same time be not countermonotone.
Let's now formally check that for the limit cases of perfect comonotonicity and countermonotonicity upper and lower tail dependence behave as intuition suggests. To that end λ_U^{XY} and λ_L^{XY} needed to be expressed as function of C_{XY}.

Theorem 2.6. *If X and Y have copula C_{XY} and the two limits in (2.3) (2.4) exist, then*

$$\lambda_U^{XY} = 2 - \lim_{v \to 1^-} \frac{1 - C_{XY}(v,v)}{1-v} =: \lambda_U^{C_{XY}}$$

$$\lambda_L = \lim_{v \to 0} \frac{C_{XY}(v,v)}{v} =: \lambda_L^{C_{XY}}.$$

Now, if X and Y are comonotone, equivalently $C_{XY}(u,v) = M(u,v)$ then, by applying the previous theorem to that case, we find $\lambda_U^M = \lambda_L^M = 1$ that is M copula has perfect upper and lower tail depencence; if X and Y are countermonotone instead, equivaleently $C_{XY}(u,v) = W(u,v)$, by applying the same procedure we find $\lambda_U^W = \lambda_L^W = 0$. that is the W copula has no tail dependence at all.
Furthermore, if X and Y are independent, $C_{XY}(u,v) = \Pi(u,v)$, then $\lambda_U^W = \lambda_L^W = 0$ that is, if the two tail coefficient vanish it does not mean that the two random variates are necessarely countermonotone, they might be also independent.

Being a special copula, we can introduce the coefficients of tail dependency also for the survival copula $\overline{C}$:

$$\lambda_U^{\overline{C}} := \lim_{v \to 1^-} \frac{1 - 2v + \overline{C}(v,v)}{1-v}$$

$$\lambda_L^{\overline{C}} := \lim_{v \to 0^+} \frac{\overline{C}(v,v)}{v}$$

if these limits exist and are finite. The following property holds trivially:

Theorem 2.7. *If $\overline{C}$ is the survival copula associated with C, then:*

$$\lambda_U^{\overline{C}} = \lambda_L^C$$

$$\lambda_L^{\overline{C}} = \lambda_U^C$$

Proof. From the definition of λ_U^C

$$\begin{aligned}
\lambda_U^{\overline{C}} &:= \lim_{v\to 1^-} \frac{1-2v+\overline{C}(v,v)}{1-v} \\
&= \lim_{v\to 1^-} \frac{C(1-v,1-v)}{1-v} \\
&= \lim_{v\to 0^+} \frac{C(v,v)}{v} := \lambda_L^C.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\lambda_L^{\overline{C}} &:= \lim_{v\to 0^+} \frac{\overline{C}(v,v)}{v} = \lim_{v\to 0^+} \frac{2v-1+C(1-v,1-v)}{v} \\
&= \lim_{v\to 1^-} \frac{1-2v+C(v,v)}{1-v} =: \lambda_L^C.
\end{aligned}$$

□

This result ensures that if a copula C has upper tail, then its survival version has lower tail and viceversa.

Moreover, if C is a convex combination of two copulas one having upper tail and the other having lower tail, then C has upper and lower tail proportionally to the combination coefficients. Formally, simple computations prove that the following theorem holds true:

Theorem 2.8. *If $C = \sum_{i=1}^n \alpha_i C_i$ is a convex combination of n copulas each of whom having upper and lower tail parameters λ_U^i and λ_L^i, then*

$$\lambda_U^C = \sum_{i=1}^n \alpha_i \lambda_U^i$$

$$\lambda_L^C = \sum_{i=1}^n \alpha_i \lambda_L^i.$$

A very interesting consequence of the last two results is that if a copula C has upper tail, its survival version $\overline{C}$ has lower tail, hence any convex combination of the two give rise to a copula having both upper and lower tail.

2.3 Copula families

Previousely we have seen that Π copula and Frchet-Hoeffding copulas can completely characterize some specific dependence structures, in particular X and Y have copula $C_{XY} = \Pi$ iff they are independent and $C_{XY} = M$ (W) iff they are perfectly positive (negative) dependent.
But what if C_{XY} is not among the previous ones? If X and Y are neither independent nor perfectly dependent, how does C_{XY} look like and which dependence structure does it identify?
The answer of that questions is not trivial. In this section the most commonly used copulas will be described and their properties will be presented. The presentation is far from complete, but covers the copulas that are considered in most application on the litterature.For exhaustive lists of copula functions and various methods for constructing copulas the books by Joe (1997) and Nelson (1999) may be consulted.

Definition 2.3. *A k-parameters copula family* $\{C_{\alpha_1 \dots \alpha_k}(u,v)\}_{\alpha_1 \dots \alpha_k}$ *is said to be* comprehensive *when it encompasses the minimum, product and maximum copulas.*

An example of comprehensive families is built as convex combination of M, W and Π and it is known as the two-parameters *Frchet family*

$$F := \{pW(u,v) + (1-p-q)\Pi(u,v) + qM(u,v) \mid p, q \in [0,1]\}$$

The definition obviousely yield

$$C_{01} = M \qquad C_{10} = W \qquad C_{00} = \Pi.$$

2.3.1 Elliptical copula class

Elliptical copulas are simply the copulas defined starting from elliptical distributions. The two most important families included in the class are the *Gaussian family* and the *Student's t family.*

Gaussian family

Let F and G be two univariate standard normal distribution functions, i.e. $\forall x \in \mathbb{R}$

$$F(x) = G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp^{-\frac{y^2}{2}} dy$$

and suppose their joint distribution function H_ρ is a bivariate normal with correlation coefficient $\rho \in [-1,1]$, i.e. $\forall (x,y) \in \mathbb{R}^2$

$$H_\rho(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{x} \int_{-\infty}^{y} \exp^{\left[-\frac{1}{2(1-\rho^2)}(s^2+t^2-2\rho st)\right]} dsdt.$$

By Sklar's theorem 1.2, *Gaussian copula* with parameter ρ is simply defined as $\forall (u,v) \in [0,1]^2$

$$\begin{aligned} C_\rho^{GA}(u,v) &= H(F^{-1}(u), G^{-1}(v)) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{F^{-1}(u)} \int_{-\infty}^{G^{-1}(v)} \exp^{\left[-\frac{1}{2(1-\rho^2)}(s^2+t^2-2\rho st)\right]} dsdt \end{aligned}$$

It is a one-parameter (depending only on ρ) and comprehensive family, indeed

$$C_{-1}^{GA} = W \quad C_0^{GA} = \Pi \quad C_1^{GA} = M.$$

As for Kendall tau and Spearman's rho coefficients, from Theorems 2.2 and 2.4 applied to a Gaussian copula C_ρ^{GA} we find that $\forall \rho \in [-1,1]$

$$\tau^\rho = \frac{2}{\pi} arcsin(\rho)$$

$$\rho_S^\rho = \frac{6}{\pi} arcsin(\frac{\rho}{2})$$

meaning that the larger the correlation ρ the more concordant F and G are. Moreover F and G are PQD if and only if $\rho \geq 0$.

As for tail dependence, from Theorem 2.6 applied to the same Gaussian copula C_ρ^{GA} we find that, unless $\rho = 1$ in which case

$$\lambda_U^1 = \lambda_L^1 = 1$$

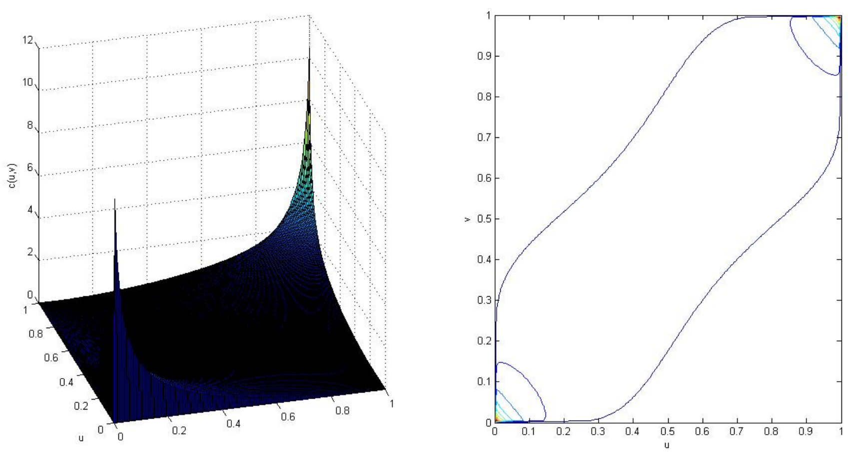
in all other cases $\rho \neq 1$

$$\lambda_U^\rho = \lambda_L^\rho = 0.$$

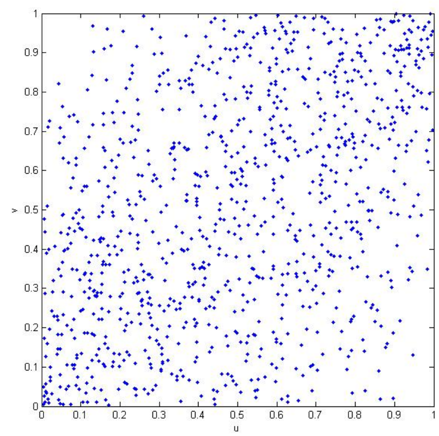
Thus, exeptionally for the trivial case of comonotonicity, in all other cases Gaussian copula has neither upper noe lower tail dependence.

If two random variables X and Y have Gaussian copula with a “non trivial” correlation coefficient ρ they do not have dependence in extreme values but, according to how large is ρ they can be more or less comonotonic. In this case we the correlation coefficient is enough to describe their dependence.

Figure 2.1 illustrates an example of Gaussian copula obtained by setting $\rho = 0.5$.



(a) Gaussian copula density $\rho = 0.5$



(b) Scatter plot

Figure 2.1: Gaussian family: an example.

t Student family

Let F and G be two univariate Student's t distribution function both with n degrees of freedom, i.e. $\forall x \in \mathbb{R}$

$$F_n(x) = G_n(x) = \int_{-\infty}^{x} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})} \left(1 + \frac{s^2}{n}\right)^{-\frac{n+1}{2}} ds$$

where Γ is the Euler function. Furthermore, suppose that their joint distribution function $H_{n,\rho}$ is a bivariate Student's t with parameter $\rho \in [-1,1]$ and $n \in \mathbb{N}$, i.e. $\forall (x,y) \in \mathbb{R}^2$.

$$H_{n,\rho}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x}^{-\infty} \int_{-\infty}^{y} \left(1 + \frac{s^2+t^2-2\rho st}{n(1-\rho^2)}\right)^{-\frac{n+2}{2}} dsdt$$

By Sklar's theorem 1.2, *Student's t copula* with parameters (ρ, n) is simply defined as $\forall (u,v) \in [0,1]^2$

$$\begin{aligned} C_{\rho,n}(u,v) &= H_{n,\rho}(F_n^{-1}(u), G_n^{-1}(v)) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{F_n^{-1}(u)} \int_{-\infty}^{G_n^{-1}(v)} \left(1 + \frac{s^2+t^2-2\rho st}{n(1-\rho^2)}\right)^{-\frac{n+2}{2}} dsdt. \end{aligned}$$

When the number of degrees of freedom diverges $n \longrightarrow \infty$, the Student's t copula converges to the Gaussian one $C_{\rho,n} \longrightarrow C_\rho^{GA}$. Moreover, $n > 2$ each margin admits a (finite) variance, $n/(n-2)$ and ρ can be interpreted as a linear correlation coefficient.

Unlikely the Gaussian family, the Student's t is a two-parameters family and contains the Frchet-Hoeffding bound copulas

$$C_{-1,n}(u,v) = W(u,v) \quad C_{1,n}(u,v) = M(u,v)$$

whatever fixed n, but the it does not contain the product copula ($C_{0,n} \neq \Pi$ for finite n), thus it is not a comprehensive family.

As for Kendall tau and Spearman's rho coefficients, from Theorems (2.2) and (2.4) applied to a Student's t copula $C_{\rho,n}$ we find the same results as for the Gaussian copula: $\forall \rho \in [-1,1]$ and $\forall n \in \mathbb{N}$

$$\tau^\rho = \frac{2}{\pi} arcsin(\rho)$$

$$\rho_S^\rho = \frac{6}{\pi} arcsin(\frac{\rho}{2})$$

meaning that independently from the degrees of freedom, the larger the correlation ρ the more concordant F and G are.
Similarly outcomes hold true for the PQD in the sense that F and G are PQD if and only if $\rho \geq 0$. As for the tail dependence, instead, from Theorem 2.6 we find that unless $\rho = 1, -1$ where

$$\lambda_U^{(1,n)} = \lambda_L^{(1,n)} = 1$$

$$\lambda_U^{(-1,n)} = \lambda_L^{(-1,n)} = 0$$

as expected, whatever n, in all other cases $\rho \neq 1, -1$, whatever n

$$\lambda_U^{(\rho,n)} = \lambda_L^{(\rho,n)} > 0.$$

Thus, exeptionally for the trivial cases of comonotonicity and countermonotonicity, in all other cases Student's t copula has both upper and lower tail dependence and the strength of that dependence decreases as the d.o.f. increases. The equality of the two parameters depends on the symmetry of the copula.
Figure 2.2 shows an example of Student's t copula obtained by setting the parameters $\rho = 0.5$ and $\nu = 3$. It is rather clear the similarity with the Gaussian case in the previous subsection in Figure 2.1 but the Student's t express more dependence in the tails than in the central part of the distribution.

2.3.2 Archimedean copula class

In this section we discuss an important class of copulas known as Archimedean copulas. What makes those copula interesting in applications is:

1. the ease with which they can be constructed;

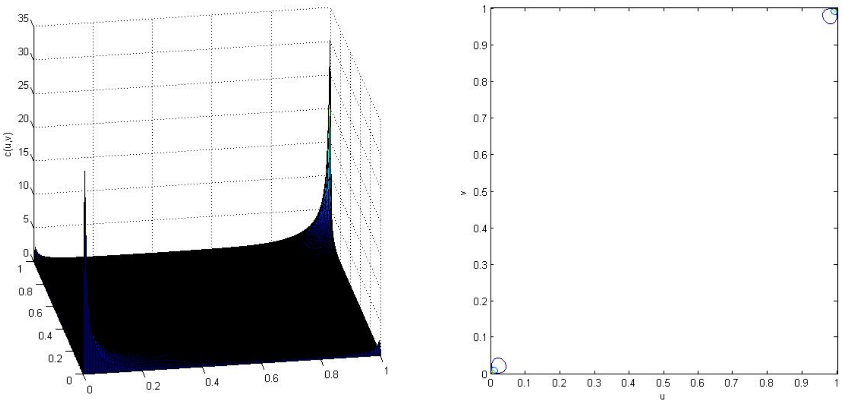
2. the great variety of copulas belonging to that class;

3. the many nice properties owned by members of this class.

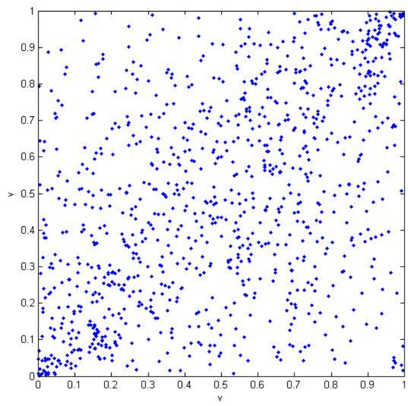
In order to define an Archimedean copula we need to introduce the concept of *generator* and *pseudo-inverse* of a generator. Recall that $I = [0, 1]$.

Definition 2.4 (Generator)**.** *A function $\varphi : I \longrightarrow \mathbb{R}^+$ is said to be a* generator *if and only if it is continous, strictly decreasing and such that $\varphi(1) = 0$.*

A *strict generatore* is a generator φ such that $\varphi(0) = +\infty$.



(a) Student's t copula density $(\nu, \rho) = (3, 0.5)$



(b) Scatter plot

Figure 2.2: Student's t family: an example.

Definition 2.5 (Pseudo-inverse)**.** *If φ is a generator, its* pseudo-inverse *is defined as follows:* $\forall v \in \mathbb{R}^+$

$$\varphi^{-1}(v) := \begin{cases} \varphi^{-1}(v) & \textit{if } 0 \leq v \leq \varphi(0) \\ 0 & \textit{if } \varphi(0) \leq v \leq \infty. \end{cases}$$

Notice that by definition whenever φ is a generator, its pseudo-inverse satisfies $\varphi^{-1}(\varphi(v)) = v \;\; \forall v \in I$ and, when φ is a strict generator, it coincides with the usual inverse.

Definition 2.6 (Archimedean copulas)**.** *Let φ be a convex generator and φ^{-1} its pseudo-inverse. An* Archimedean copula *with generator φ is a map C_φ defined as follows* $\forall (u, v) \in I$

$$C_\varphi(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)). \tag{2.5}$$

In order to be sure that the previous definition is well posed, i.e. the map C_φ satisfies properties from *i.* to *iii.* in Definition 1.3, next result turns out to be useful.

Theorem 2.9. *Let $\varphi : I \longrightarrow \mathbb{R}^+$ be a continous, strictly decreasing function such that $\varphi(1) = 0$, and let φ^{-1} its pseudo-inverse.*
The map C_φ given in (2.5) is a copula if and only if φ is convex.

It can be shown that the density of an Archimedean copula C_φ or briefly C is computed as function of its generator, $\forall (u, v) \in [0, 1]$

$$c_\varphi(u, v) = -\frac{\varphi^{''}(C_\varphi(u, v))\varphi^{'}(u)\varphi^{'}(v)}{(\varphi^{'}(C_\varphi(u, v))^3)} \tag{2.6}$$

where $\varphi^{'}$ exists a.e. since the generator is convex.
As for the concordance measures, Genest and Mackay (1986) demonstrated that

$$\tau^\varphi = 4 \int_I \frac{\varphi(v)}{\varphi^{'}(v)} dv + 1 \tag{2.7}$$

while, as for tail dependence parameters, they can be expressed in terms of limit involving the generator and its inverse (Joe, 1997)

Theorem 2.10. *Let C_φ be an Archimedean copula. Then*

$$\lambda_U^\varphi = 2 - \lim_{v \to 0^+} \frac{1 - \varphi{-}1(2v)}{1 - \varphi{-}1(v)}$$

$$\lambda_L^\varphi = \lim_{v \to +\infty} \frac{\varphi{-}1(2v)}{\varphi{-}1(v)}.$$

Among the popular Archimedean families we find the *Gumbel*, the *Clayton* and the *Frank* families. We will now introdce each of them, singularly, focusing on the dependence structure they model.

Gumbel family

For all $\alpha \in [1, \infty)$ we define the family of generators φ_α as follows, $\forall t \in I$

$$\varphi_\alpha(t) := (-\ln t)^\alpha.$$

The Archimedean copula family associated to the previous generators is called *Gumbel copula family* or *Gumbel-Hougaard copula family* and it is given by

$$C^\alpha(u, v) = \exp^{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{1/\alpha}}.$$

Following (2.6), its density is given by

$$c_\alpha(u, v) = \frac{[-\ln(u)]^{\alpha-1}[-\ln(v)]^{\alpha-1}}{uv\exp(-A^{1/\alpha})}\left(\frac{1}{A^{2-2/\alpha}} + \frac{\alpha - 1}{A^{2-1/\alpha}}\right) \tag{2.8}$$

where

$$A := [-\ln(u)]^\alpha + [-\ln(v)]^\alpha$$

For $\alpha = 1$ we discover the independence copula, i.e. $C^1(u, v) = \Pi(u, v)$ while for $\alpha \to \infty$ the Gumbel copula tends to the comonotone copula, i.e. $C^\alpha(u, v) \to M(u, v)$ so that the Gumbel copula interpolates between independence and perfect positive dependence.
Gumbel family comes as an example of a simple copula which has tail dependence in the up-right corner of the unit square I^2, indeed, from (2.7) and Theorem 2.10 it follows that $\forall \alpha \in [1, \infty)$

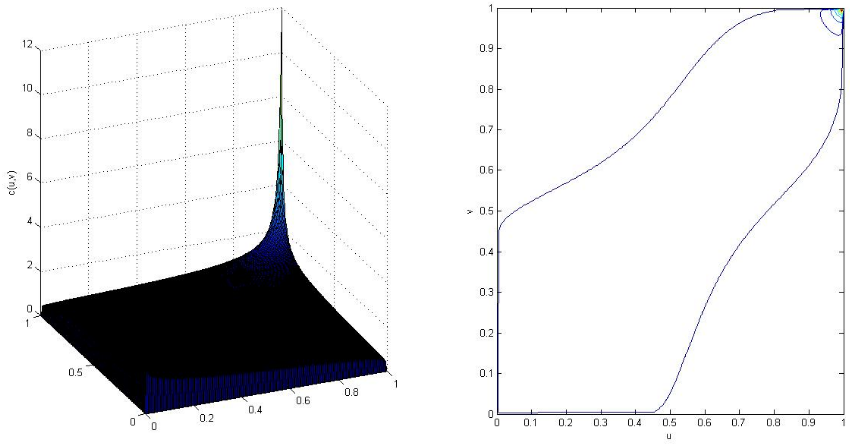
$$\tau^\alpha = 1 - \alpha^{-1} \geq 0$$

$$\lambda_U^\alpha = 2 - 2^{1/\alpha} > 0$$

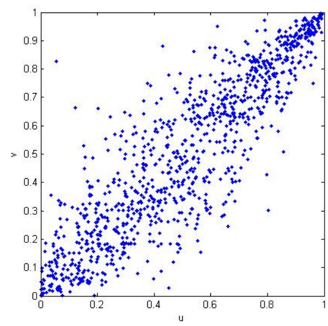
$$\lambda_L^\alpha = 0.$$

In other words, Gumbel family seems to model only positive dependence together with upper tail dependence - because of the non negativity of all the parameters- and in particular the larger α, the more C^α models a positive dependence (notice that τ^α and λ_U^α are both increasing function of α).
Figure 2.3 illustrates an example of Gumbel copula with parameter $\alpha = 1.12$.

Furthermore, Figure 2.4 plot the survival copula of Figure 2.3 and it clearly show how the upper tail dependence turns into the lower tail dependence.

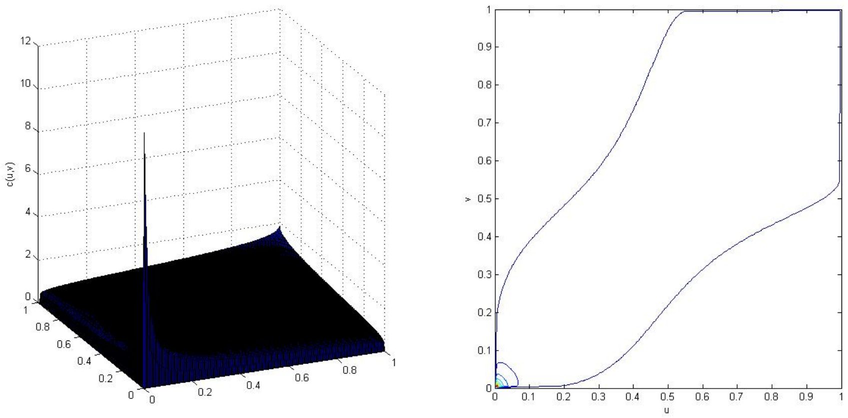


(a) Gumbel copula density $\alpha = 1.2$

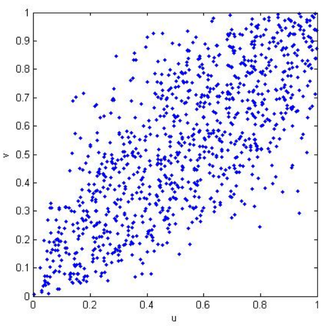


(b) Scatter plot

Figure 2.3: Gumbel family: an example.



(a) Survival Gumbel copula density $\alpha = 1.2$



(b) Scatter plot

Figure 2.4: Survival Gumbel family: an example.

Clayton copula

For all $\alpha \in [-1, 0) \cup (0, \infty) =: \Omega$ we define the family of generators φ_α as follows, $\forall t \in I$

$$\varphi_\alpha(t) := \frac{1}{\alpha}(t^{-\alpha} - 1).$$

The Archimedean copula family associated to the previous generators is called *Clayton family* and it is given by

$$C^\alpha(u, v) = (\max((u^{-\alpha} + v^{-\alpha} - 1)^{-\frac{1}{\alpha}}, 0)).$$

Following (2.6), its density is given by

$$c_\alpha(u, v) = [-1 + u^{-\alpha} + v^{-\alpha}]^{-2-\frac{1}{\alpha}} * u^{-\alpha-1} * v^{-\alpha-1} * (1 + \alpha) \tag{2.9}$$

For $\alpha = -1$ we obtain the countermonotone copula, i.e. $C^{-1}(u, v) = W(u, v)$ while, for the limits $\alpha \to 0$ we obtain $C^\alpha(u, v) \to \Pi(u, v)$, and for $\alpha \to \infty$ we obtain $C^\alpha(u, v) \to M(u, v)$.
Thus, as the Gumbel family, the Clayton one interpolates between two dependency structures, in this case from the countermonotonicity up to the monotonicity, passing through the independence.
Unlike th Gumbel family, the Clayton one comes as an example of a simple copula which has tail dependence in the low-left corner of the unit square I^2, indeed, from (2.7) and Theorem 2.10 it follows that $\forall \alpha \in \Omega$

$$\tau^\alpha = \frac{\alpha}{\alpha + 2}$$

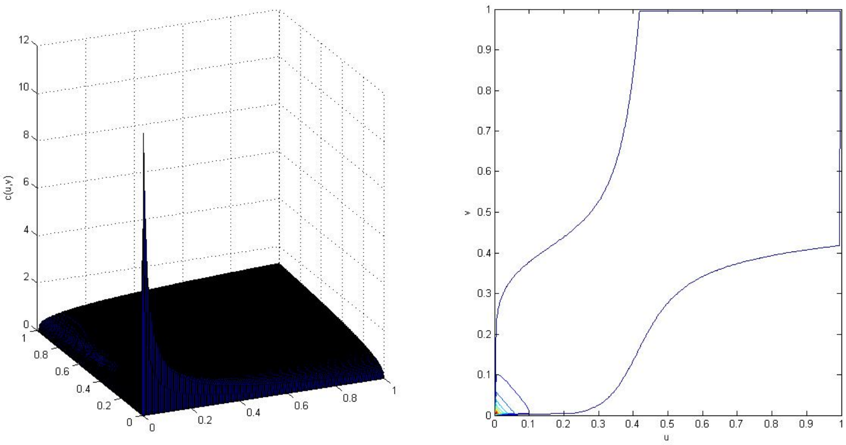
$$\lambda_U^\alpha = 0$$

$$\lambda_L^\alpha = 2^{-1/\alpha}$$

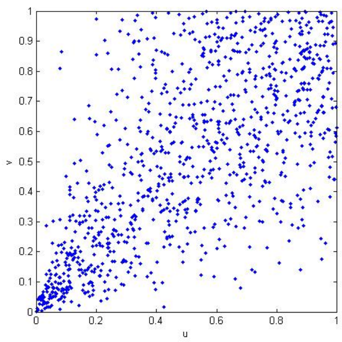
Clayton family seems to model both positive and negative dependence together with lower tail dependence and as in the previous case, the larger α the more C^α describes a positive dependence (notice that also here τ^α is an increasing function of α while λ_L^α reaches his largest values when $\alpha \to 0^-$).
Figure 2.5 shows an example of Clayton copula with parameter $\alpha = 0.31$

Furthermore, in Figure 2.6 the survival copula of Figure 2.5 is reported, and it clearly shows how the lower tail dependence turns into the upper tail dependence.

It should be noted that both for Gumbel and Clayton families the parameter do not need to be very far from 0 to model the tail dependence. In other words there are no values close to independence which model also the central part of the distribution.

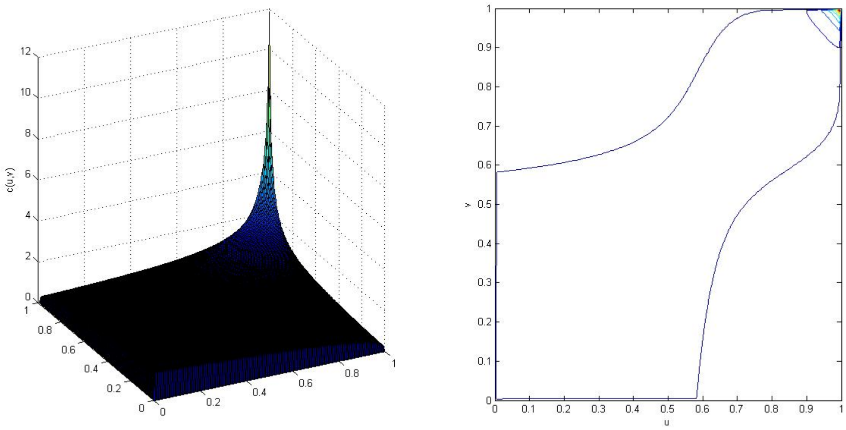


(a) Clayton copula density $\alpha = 0.31$

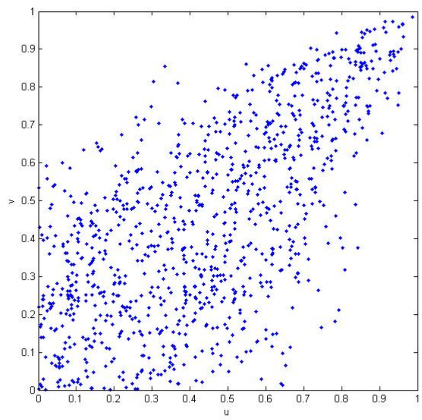


(b) Scatter plot

Figure 2.5: Clayton family: an example.



(a) Survival Clayton copula density $\alpha = 0.31$



(b) Scatter plot

Figure 2.6: Survival Clayton family: an example.

Frank copula

For all $\alpha \in (-\infty, 0) \cup (0, \infty)$ we define the family of generators φ_α as follows, $\forall t \in I$

$$\varphi_\alpha(t) := \ln(\exp^{-\alpha} - 1) - \ln(\exp -\alpha t - 1).$$

The Archimedean copula family associated to the previous generators is called *Frank family* and it is given by

$$C^\alpha(u, v) = -\frac{1}{\alpha} \ln \left(1 + \frac{(\exp^{-\alpha u} - 1)(\exp^{-\alpha v} - 1)}{\exp^{-\alpha - 1}} \right).$$

Its density is given by formula (2.10)

$$c_\alpha(u, v) = \frac{-\alpha \exp^{-\alpha u} \exp^{-\alpha v} (\exp^{-\alpha} - 1)}{[\exp^{-\alpha} - 1 + (\exp^{-\alpha u} - 1)(\exp \alpha v - 1)]^2} \tag{2.10}$$

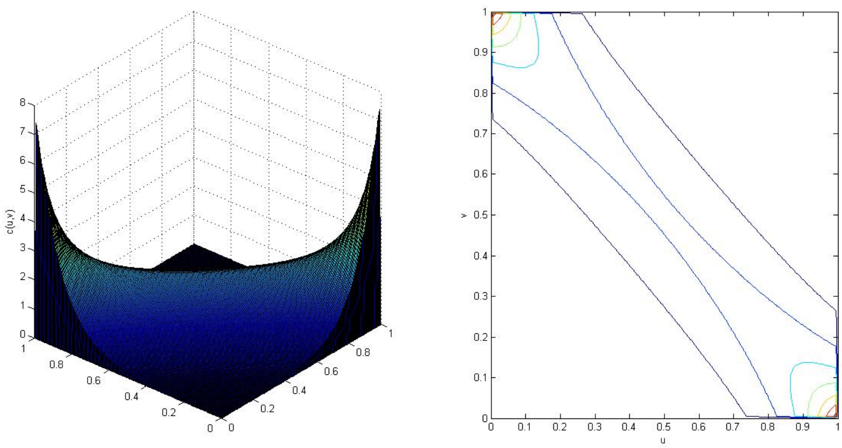
We obtain the perfect dependent copulas $W(u, v)$ and $M(u, v)$ when $\alpha \to -\infty$ and $\alpha \to \infty$ respectively, and the independent copula Π when $\alpha = 0$ thus, similarly to Clayton family, it interpolates between comonotonicity and countermonotonicity passing through independence. As for parameters of dependence we have $\forall \alpha \in \mathbb{R}\ 0$

$$\tau^\alpha = 1 + \frac{4[D_1(\alpha) - 1]}{\alpha}$$

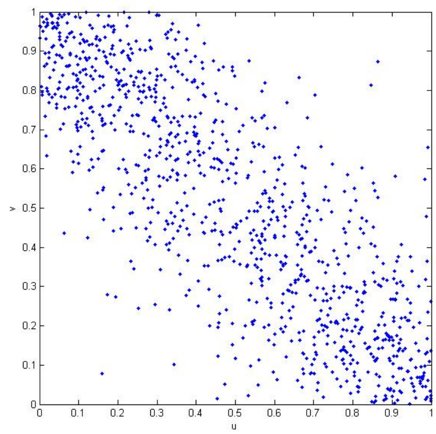
$$\lambda_U^\alpha = 0$$

$$\lambda_L^\alpha = 0$$

meaning that Frank copula models from perfect negative to perfect positive dependence as α increases, without any dependence in extreme values. Figure2.7 shows an example of Frank copula with parameter $\alpha = -8$.



(a) Frank copula density $\alpha = -8$



(b) Scatter plot

Figure 2.7: Frank family: an example.

Chapter 3

Modelling dependence with copulas

This work aims at studying the dependence structure of two stock returns S_1 and S_2 through a copula approach: starting from an iid sample of (S_1, S_2) we look for the copula model which best fit $C_{S_1 S_2}$.
The following chapter is meant firstly to better describe the problem in mathematical terms and in a general framework (X and Y are continous r.v), and secondly to point out the main steps of the analysis towards the solution of that problem.

3.1 The mathematical problem

Let $X \sim F$ and $Y \sim G$ be two *continous* random variates and let $\{(x_i, y_i)\}_{i=1}^{n}$ be an iid sample from the joint vector (X, Y).
How to trace back to C_{XY} from the sample?
Sklar's theorem states that

$$C_{XY}(u,v) = H(u,v) \qquad \forall (u,v) \in I^2,$$

with

$$H(u,v) = \mathbb{P}(F(X) \leq u,\ G(Y) \leq v)$$

and $F(X), G(Y) \sim U[0,1]$ being X and Y continous random variables.
Thus, since C_{XY} corresponds to the cumulative distribution function of $(F(X), G(Y))$, a good model for C_{XY} is a good model for H and to find a good model for H we need an iid sample from $(F(X), G(Y))$, say $\{(u_i, v_i)\}_{i=1}^{n}$.
In case F and G are completely specified (*parametric case*) it suffices to set

$$\begin{cases} u_i := F(x_i) \\ v_i := G(y_i), \end{cases}$$

but when F and G are continous and not fully known or not known at all (*non-parametric case*), an *approximate* iid sample can be obtained by setting

$$\begin{cases} u_i := \hat{F}(x_i) \\ v_i := \hat{G}(y_i), \end{cases}$$

where $\hat{F}$ and $\hat{G}$ are the empirical versions of F and G, i.e.

$$\hat{F}(x_i) = \frac{1}{n} \sum_{j=1}^{n} 1(x_j \leq x_i)$$

and similarly

$$\hat{G}(y_i) = \frac{1}{n} \sum_{j=1}^{n} 1(y_j \leq y_i).$$

Finally, the mathematical problem of modelling C_{XY} when X and Y are two continous random variates and $\{(x_i, y_i)\}_{i=1}^{n}$ is an iid sample consists in fitting the joint distribution H through $\{(u_i, v_i)\}_{i=1}^{n}$.

3.2 The analysis

In order to obtain a reasonable model for H the analysis has been divided in three main steps:

1. Estimation of H: the empirical copula of (X, Y);
2. Estimation of Λ_U^{XY}, λ_U^{XY} and Λ_L^{XY}, λ_L^{XY} - upper and lower tail functions with their limits ;
3. Model selection and *goodness of fit* test.

Step 1: Empirical copula

We define *empirical copula* of X and Y to be the empirical cdf of $(F(X), G(Y))$, which will be denoted by $\hat{C}_{XY}$,
$\forall (u, v) \in [0, 1]^2$

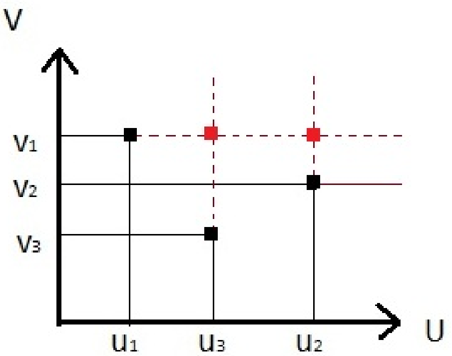
$$\hat{C}_{XY}(u, v) := \frac{1}{n} \sum_{i=1}^{n} 1(F(x_i) \leq u, G(y_i) \leq v) = \frac{1}{n} \sum_{i=1}^{n} 1(u_i \leq u, v_i \leq v).$$

Since $\hat{C}_{XY}$ is an empirical cdf on the unit square, its graph is a step function on $[0, 1]^2$ jumping both in the 'sample points' $(u_i, v_i)_{i=1}^{n}$ and in the so called

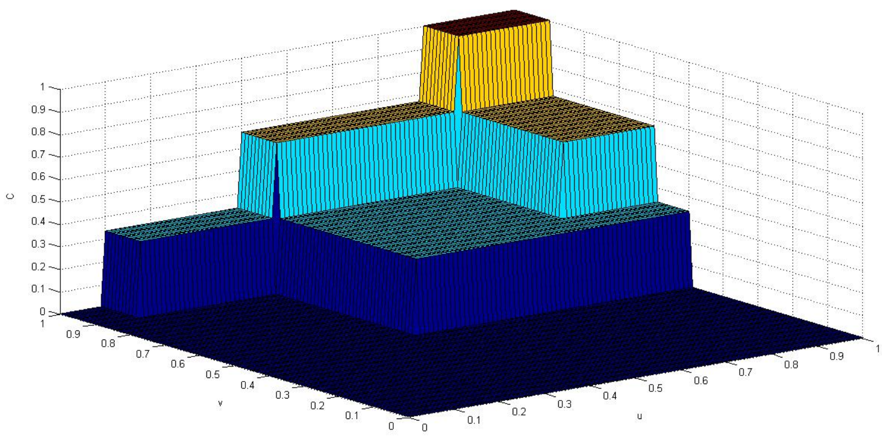
intersection points.
An *intersection point* is any point (u_i, v_j) such that

$$u_i > u_j \text{ and } v_i < v_j.$$

For $n = 3$ an example is given in Figure 3.1-(a) where the black dots represent a possible configuration of the sample $\{(u_i, v_i)\}_{i=1}^n$ while the red ones the related intersection points; 3.1-(b) displays instead the related empirical copula surface jumping exactly in those points.



(a) Intersection points.



(b) Empirical copula

Figure 3.1: Example of empirical copula.

The Glivenko - Cantelli theorem ensures that $\hat{C}_{XY}$ is a good estimator of C_{XY} -the larger n the better the estimate- and it gives initial tips about the

dependence between X and Y: according to whether $\hat{C}_{XY}$ lies above or below the product copula Π a positive or negative dependence can be deduced - contour plots are generally the best way to notice it -.
However, you may face situations in which $\hat{C}_{XY}$ neither is completely below nor even completely above Π which means that you cannot deduce a complete positive or negative dependence.
Estimating C_{XY} with its empirical version is a good way to start with, but far it is from giving detailed information about symmetry, tails behaviour so to figure out a possible model.
For that reason it could be much more helpful estimating its pdf c_{XY} with a bivariate kernel density estimator
$\forall (u, v) \in [0, 1]^2$

$$\hat{c}_{XY}(u, v) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_1, h_2} K \left(\frac{u - u_i}{h_1}, \frac{v - v_i}{h_2} \right)$$

where K is a bivariate kernel function and (h_1, h_2) is the bandwidth vector. It is rather common, and that is what we will actually do, to choose $K(u, v)$ as the gaussian kernel function $K^{GA}(u, v)$

$$K^{GA}(u, v) := \frac{1}{2\pi} \exp^{\frac{-1}{2}(u^2+v^2)} \qquad \forall (u, v) \in [0, 1]^2$$

and (h_1, h_2) as the optimal bandwidth (h_1^*, h_2^*) in the sense that

$$(h_1^*, h_2^*) = arg \min_{(h_1, h_2)} \mathbb{E} \left[\int_0^1 \int_0^1 [\hat{c}_{XY}(u, v) - c_{XY}(u, v)]^2 dudv \right].$$

Step 2: Tail dependence functions

Although $\hat{c}_{XY}$ is a qualitative aid in finding out what happens in the tails of H, more accurate clues are obtained by estimating upper and lower tail functions Λ_U^{XY} and Λ_L^{XY}.
According to Theorem 2.6 reasonable estimators are:

$$\hat{\Lambda}_U^{XY}(q) = 2 - \frac{1 - \hat{C}_{XY}(q, q)}{1 - q} \qquad \forall q \in [0, 1] \tag{3.1}$$

$$\hat{\Lambda}_L^{XY}(q) = \frac{\hat{C}_{XY}(q, q)}{q} \qquad \forall q \in [0, 1]. \tag{3.2}$$

Estimators (3.1) and (3.2) are expected to be respectively a decreasing and an increasing function from $[0, 1]$ to $[0, 1]$, reaching the maximum value 1 in $v = 0$ for (3.1) and in $v = 1$ for (3.2).
If their trend looks symmetrical, i.e $\hat{\Lambda}_U^{XY}(q) = \hat{\Lambda}_L^{XY}(1 - q)$, a symmetry in the tails could be hypotized and their limits λ_U^{XY} and λ_L^{XY} can reasonably thought equal.
However, when (3.1) and (3.2) do not give evidence of symmetry at all, equality in the the tails can still hold true. A great estimate of λ_U^{XY} and λ_L^{XY} comes from the followiong procedure:
simulate N times from a fully known copula C_θ with upper and lower tail parameters λ_U and λ_L;
for each simulation k define $\hat{\Lambda}_U^k$ and $\hat{\Lambda}_L^k$ as in (3.1)-(3.2) and compute the two optimal points

$$u_U^k = arg \min_{q \in I(1^-)} \left\{ |\hat{\Lambda}_U^k - \lambda_U| \right\} \tag{3.3}$$

$$v_L^k = arg \min_{q \in I(0^+)} \left\{ |\hat{\Lambda}_L^k - \lambda_L| \right\} \tag{3.4}$$

where $I(1^-)$ is a left neighbourhood of 1 and $I(0^+)$ a right neighbourhood of 0;
Finally, if $v_U := \frac{1}{N} \sum_{k=1}^N v_U^k$ and $v_L := \frac{1}{N} \sum_{k=1}^N v_L^k$ are the mean value of (3.3) and (3.4) then we can estimate as follows:

$$\hat{\lambda}_U^{XY} := \hat{\Lambda}_U^{XY}(v_U)$$

$$\hat{\lambda}_L^{XY} = \hat{\Lambda}_L^{XY}(v_L).$$

Step 3: Goodness of fit

The previous two steps are a rough descriptive analysis to point at a family model $\mathfrak{C}_\theta$ the underlying copula C_{XY} might belong to. The model $\mathfrak{C}_\theta$ can be either a single family - e.g. the Gaussian one, an Archimedean one - or a convex combination of different families - e.g. Gumbel and Clayton, Frank and t Student -.
Once a family model $\mathfrak{C}_\theta$ has been selected, the third and last step focuses on testing the statistical hypotesis

$$\begin{cases} H_0: & C_{XY} \in \mathfrak{C}_\theta \\ H_1: & C_{XY} \notin \mathfrak{C}_\theta. \end{cases}$$

The test will be performed in three different stages:

- Parameter estimation

- Distance computation
- P-value approximation.

In the *parameter estimation* stage C_{XY} is assumed to belong to $\mathfrak{C}_\theta$, so we estimate θ with $\hat{\Theta}^{ML}$, the Maximum Log-Likelyhood estimator defined in (3.5),

$$\hat{\Theta}^{ML} = arg \max_{\theta \in \Omega} \sum_{i=1}^{n} \ln(\mathfrak{c}_\theta(u_i, v_i)) \tag{3.5}$$

where $\mathfrak{c}_\theta(u, v)$ is the pdf of $\mathfrak{C}_\theta(u, v)$.
If the model is a one-parameter copula family, the parameter θ is a scalar value ($\Omega \subset \mathbb{R}$) so (3.5) gives back the scalar value $\hat{\theta}^{ML}$; in case of a convex combination $\mathfrak{C}_\theta = p\mathfrak{C}_\alpha + (1-p)\mathfrak{C}_\beta$, the parameter $\theta = (p, \alpha, \beta)$ is a triple ($\Omega \subset \mathbb{R}^3$) so (3.5) gives back the triple $\hat{\theta}^{ML} = (\hat{p}, \hat{\alpha}, \hat{\beta})$.

Once θ has been estimated we have found the best model for C_{XY} in $\mathfrak{C}_\theta$, say $C_{\hat{\theta}^{ML}}$. Next step consists in testing the hypotesis

$$\begin{cases} H_0: & C_{XY} = C_{\hat{\theta}^{ML}} \\ H_1: & C_{XY} \neq C_{\hat{\theta}^{ML}}. \end{cases}$$

The literature offers several methods to test such a goodness of fit hypotesis, and almost all of them agree on rejecting H_0 if $\hat{C}_{XY}$ distances "a lot" from the fitted one $C_{\hat{\theta}^{ML}}$. In other words, if we define a distance $\mathfrak{D}$ between the two copulas, the rejection region RR will be defined

$$RR := \{\mathfrak{D} > k\}$$

where k is a real parameter whose value depends on the test significance level. In this work $\mathfrak{D}$ has been selected as the Kolmogorov-Smirnov distance

$$KS := \sup_{(u,v)\in[0,1]^2} \left\{ |\hat{C}_{XY}(u, v) - C_{\hat{\theta}^{ML}}(u, v)| \right\}$$

and the *distance computation* stage takes care of computing it.
Similarly to the one dimentional framework, a numerical method to compute the right (and not approximated) K-S distance consists in computing that distance exactly in the "jumping points" and then take the max.
In symbols, if by changing notation $C_{\hat{\theta}} := C_{\hat{\theta}^{ML}}$ and if $\{(u_j, v_j)\}_{j=1}^{m}$ represents the intersection points, then

$$\begin{aligned} KS_1 \;=\; & \max_{i=1\ldots n} \big(\; |\hat{C}(u_i, v_i) - C_{\hat{\theta}}(u_i, v_i)|, |\hat{C}(u_i^-, v_i) - C_{\hat{\theta}}(u_i^-, v_i)| \\ & |\hat{C}(u_i^-, v_i^-) - C_{\hat{\theta}}(u_i^-, v_i^-)|, |\hat{C}(u_i, v_i^-) - C_{\hat{\theta}}(u_i, v_i^-)| \; \big) \end{aligned}$$

$$\begin{aligned} KS_2 &= \max_{j=1\dots m}\big(\ |\hat{C}(u_j,v_j)-C_{\hat{\theta}}(u_j,v_j)|,|\hat{C}(u_j^-,v_j)-C_{\hat{\theta}}(u_j^-,v_j)| \\ &\qquad |\hat{C}(u_j^-,v_j^-)-C_{\hat{\theta}}(u_j^-,v_j^-)|,|\hat{C}(u_j,v_j^-)-C_{\hat{\theta}}(u_j,v_j^-)|\ \big) \end{aligned}$$

$$KS_3 = \max_{i=i\dots n}\big(\ |\hat{C}(1,v_i)-C_{\hat{\theta}}(1,v_i)|,|\hat{C}(u_i,1)-C_{\hat{\theta}}(u_i,1)|\ \big)$$

and finally

$$KS = \max\{KS_1, KS_2, KS_3\}\,.$$

To visulaise where the max value is reached and generally speaking where the two functions mainly differs, it is rather useful to plot the objective function $|\hat{C}(u,v)-C_{\hat{\theta}}(u,v)|$ on $[0,1]^2$ on the unit square $[0,1]\times[0,1]$.

How to establish now whether to reject H_0 or not? The *P-value approximation* stage aims at approximating P-value and comparing it with the most common significance levels α. Two are the limit cases: if the P-value is smaller than any values of α, H_0 is rejected and conversely, if it is larger than any α, H_0 cannot be rejected and we use to say that there is no statistical evidence of rejecting null hypotesis.
In our case to decide whether the value KS is "too big" or not the bootstrap method has been performed to approximate the P-value defined as

$$P-value := \mathbb{P}(K-S > KS|H_0 \text{ true})$$

where KS is the K-S distance evaluated in $\{(u_i,v_i)\}_{i=1}^n$, while $K-S$ is the statistic.. The bootstrap method is an iterative method which performs in each iteration $k=(1,2\dots N)$ - with N "large enough" - the following steps

1. Generate a sample of size n from $C_{\hat{\theta}}$, say $(u_i^k,v_i^k)_{i=1}^n$;

2. Supposing that $(u_i^k,v_i^k)_{i=1}^n$ comes from the copula family $\mathfrak{C}_\theta$, estimate the parameter θ as in the *estimation* part, so to have $\hat{\theta}_k^{ML}$ and then $C_{\hat{\theta}_k^{ML}}$;

3. Define $\hat{C}_k$ as the empirical cdf of $(u_i^k,v_i^k)_{i=1}^n$ and compute the KS distance as previousely done, so to have KS_k.

The entire procedure gives back as output an N dimensional vector of distances $(KS_1, KS_2, \dots KS_N)$ each of them computed in each iteration.
Recalling that large values of K-S statistic lead to a rejection of H_0 Stute showed that under appropriate regularity conditions an approximate P-value for the test is given by

$$P-value = \frac{1}{N}\sum_{k=1}^{N} 1(KS_k > KS).$$

The validity of that procedure stems from the fact that under H_0 and as $n \to \infty$, the sequence $(KS, KS_1, KS_2 \dots KS_N)$ converges weakly to a vector $(KS^*, KS_1^*, KS_2^* \dots KS_N^*)$ of mutually independent and identically distributed random variables, thus $(KS, KS_1, KS_2 \dots KS_N)$ can approximately be viewed ad an iid sample from K-S.

Motivation of choices.

In the *Parameter estimation* stage we have decided to estimate the parameter θ with the MLE because it is a powerful estimator for all its nice properties, but any other estimation could have been used instead.
As for the *Distance computation* there are pros and cons in choosing the Kolmogorov distance. The KS statistic is known to be most sensitive around the median of the distribution and relatively insensitive to deviations in the tails, and as we are above all interested in the tails behaviour, that one seems to be the main drawback. However, it can be shown that KS distance is a distribution-free statistc, so no matter which copula model we are testing, under H_0 the KS distribution does not change. This is actually a optimal property from a numerical point of view when computing P-value because we can approximate P-value only for those models $C_{\hat{\theta}^{ML}}$ performing the least distance.

Chapter 4

Data analysis and results

As roughed out at the beginning of the previous chapter, we wish to study the dependence of two stock returns S_1 and S_2 by looking for a good copula model of (S_1, S_2).
The method developped in *Chapter 3* is applicable only when dealing with continous random variables and stock returns are generally supposed to be stochastic processes - $S_1 = \{S_1(t)\}_{t\in T}$ and $S_2 = \{S_2(t)\}_{t\in T}$ -.
For that reason, a rather common way to bypass this problem is working with the so called *log-returns* X and Y associated to S_1 and S_2. In fact, under a specific assumption made on stocks dynamic - and we will see in a while which is -, X and Y are continous random variables and C_{XY} can be modelled exactly as the method in *Chapter 3* suggests.
The following Chapter is firstly ment to introduce the concept of *log-returns* and to show how the assumptions let them be continous random variates, and secondly to perform the analysis on log-returns of real data.

4.1 The assumption

The only assumption made is that $S_1(t)$ and $S_2(t)$ are two *geometric brownian motions*, that is
$\exists\ \mu_1, \mu_2 \in \mathbb{R}$ (drifts) and $\sigma_1, \sigma_2 \in \mathbb{R}^+$ (volatilities) such that

$$dS_1(t) = S_1(t)\left(\mu_1 dt + \sigma_1 dW_1(t)\right) \tag{4.1}$$

$$dS_2(t) = S_2(t)\left(\mu_2 dt + \sigma_1 dW_2(t)\right), \tag{4.2}$$

where $W_1(t)$ and $W_2(t)$ are two Wiener processes.
By applying Ito formula it can be easily checked that

$$S_1(t) = S_1(0)e^{(\mu_1 - \frac{1}{2}\sigma_1^2)t + \sigma_1 W_1(t)}$$

$$S_2(t) = S_2(0)e^{(\mu_2 - \frac{1}{2}\sigma_2^2)t + \sigma_2 W_2(t)}$$

are solutions of equations (4.1) and (4.2).
Fixed a time amplitude δt, the *log-returns* of $S_1(t)$ and $S_2(t)$ are defined as

$$X := ln\left(\frac{S_1(t+\delta t)}{S_1(t)}\right) = \left(\mu_1 - \frac{1}{2}\sigma_1\right)\delta t + \sigma_1\left(W_1(t+\delta t) - W_1(t)\right)$$

$$Y := ln\left(\frac{S_2(t+\delta t)}{S_2(t)}\right) = \left(\mu_2 - \frac{1}{2}\sigma_2\right)\delta t + \sigma_2\left(W_2(t+\delta t) - W_1(t)\right).$$

Being $(W_i(t+\delta t) - W_i(t))$ Wiener process increments for $i = 1, 2$
$(W_i(t+\delta t) - W_i(t)) \sim N(0, \delta t)$, thus

$$X \sim N\left(\left(\mu_1 - \frac{1}{2}\sigma_1\right)\delta t, \sigma_1^2 \delta t\right)$$

$$Y \sim N\left(\left(\mu_2 - \frac{1}{2}\sigma_2\right)\delta t, \sigma_2^2 \delta t\right).$$

Therefore, if $\{(s_1(i), s_2(i))\}_{i=1}^n$ is a time series from (S_1, S_2) of amplitude δt, i.e. $t_i - t_{i-1} = \delta t$, the set $\{(x_i, y_i)\}_{i=1}^{n-1}$ defined as follows

$$x_i := ln\left(\frac{s_1(i)}{s_1(i-1)}\right) = \left(\mu_1 - \frac{1}{2}\sigma_1\right) + \sigma_1\left(W_1(i) - W_1(i-1)\right)$$

$$y_i := ln\left(\frac{s_2(i)}{s_2(i-1)}\right) = \left(\mu_2 - \frac{1}{2}\sigma_2\right) + \sigma_2\left(W_2(i) - W_2(i-1)\right)$$

is an iid sample from (X, Y), where $X \sim N((\mu_1 - \frac{1}{2}\sigma_1)\delta t, \delta t)$ and equivalently $Y \sim N((\mu_2 - \frac{1}{2}\sigma_2)\delta t, \delta t)$.
In conclusion: just assuming the stock prices S_1 and S_2 to be geometric brownian motions, the associated log-returns X and Y are both gaussian random variates of which an joint iid sample is available once a joint time series of the two prices is given.
Furthermore, since the purpose of the work is far from estimating the drifts (μ_1, μ_2) and the volatilities (σ_1, σ_2) in equations (4.1)-(4.2), X and Y will be thought as continous and not fully specified random variables. By doing that we are supposed to treat the non-parametric framework according to the notations in *Chapter 3*.

4.2 Data analysis

The two stocks aim of the study are both traded in the Italian market: S_1 is issued by the insurance company *Azimut* (AZM.MI) whereas S_2 is issued by the bank *Banca Generali* (BGN.MI).
The whole analysis is based on a daily time series $\{(s_1(i), s_2(i))\}_{i=1}^{n}$ which spans 5 years starting from the 1^{st} of January 2008 up to the 1^{st} of January 2013 ($n = 1295$)[1]. Red and blue lines in Figure 4.1 represent *Azimut* and *Banca Generali* price trajectories all over the 5 years.

Figure 4.1: *AZM* returns VS *BGN* returns.

The returns trend clearly hints a positive dependence between the stocks in the sense that whenever one of the two prices goes up or down, so does the other one.
When computing log-returns $\{(x_i, y_i)\}_{i=1}^{n-1}$, the same positive dependence can be observed from their plot in Figure 4.2 - still red and blue lines for X =AZM log-returns and Y =BGN log-returns respectively-.

According to the notations previousely introduced, the analysis will be now addressed to find a good model for H, the joint distribution of $(F(X), G(Y))$. Dealing with a non parametric case, the samples $\{u_i\}_{i=1}^{n}$ and $\{v_i\}_{i=1}^{n}$ of (U, V) should only *approximately* be distributed as a uniform on $[0, 1]$. In fact their empirical cfd plotted in Figure 4.3 look like a uniform cdf except for a particularly marked jump in the middle of the interval.

[1]The time series has been downloaded from the website *www.yahoofinance.it*.

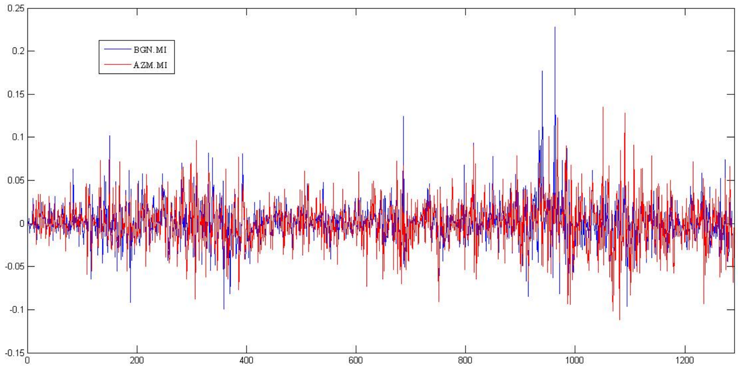
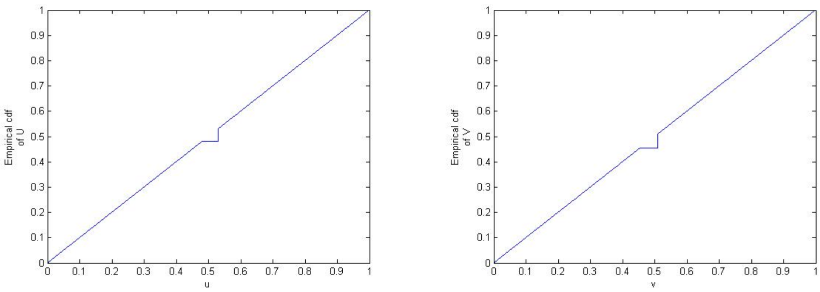


Figure 4.2: *AZM* log-returns VS *BGN* log-returns.



(a) *Empirical cdf of U.* (b) *Empirical cdf of V.*

Figure 4.3: Empirical distribution.

A formal proof of non uniformity of the two samples is given by small P-values returned by the Kolmogorv-Smirnov test.
The reason why the uniform hypotesis is missing comes from the fact that log-returns X and Y are not continous random variates. There is indeed a great amount of zero log-returns meaning that in the five analysed years time there are groups of at least two consecutive days in which returns are stationairy for both stocks.
However, despite the lack of uniformity, the analysis has been carried on without modifying the dataset.

Step 1: Empirical copula

Figure 4.4 displays the level curves of the empirical copula $\hat{C}_{XY}$ compared with the Frchet-bounds copulas ones, M and W, and product copula ones Π.

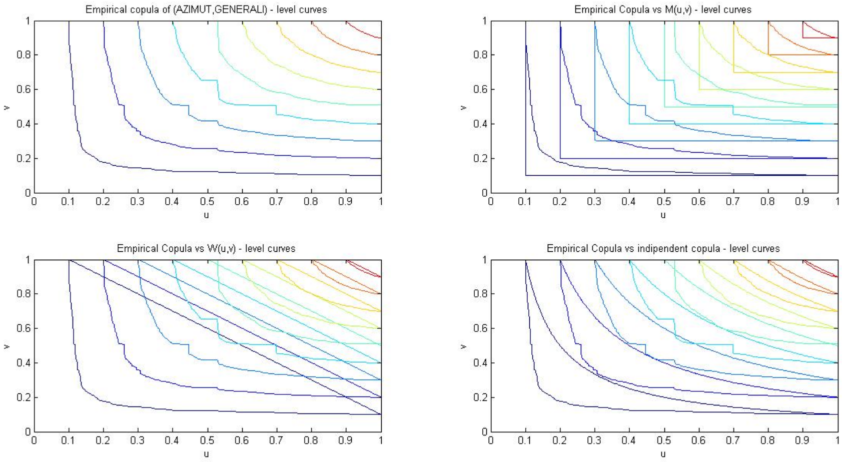


Figure 4.4: Empirical copula VS main copulas contours.

As expected, contour plots gives evidence that $\hat{C}_{XY}$ is bounded by M and W and that $\hat{C}_{XY}$ completely lies above Π which confirms the positive dependence already glimpsed in the previous graphs.
To have a more clear and complete overview, Figure 4.5 shows the surface of $\hat{C}_{XY}$ in (a) and the normal kernel density estimator $\hat{c}_{XY}$ with its level curves in (b)-(c). The latter two graphs point out a clear symmetry of the distribution and marked spikes in the tails.

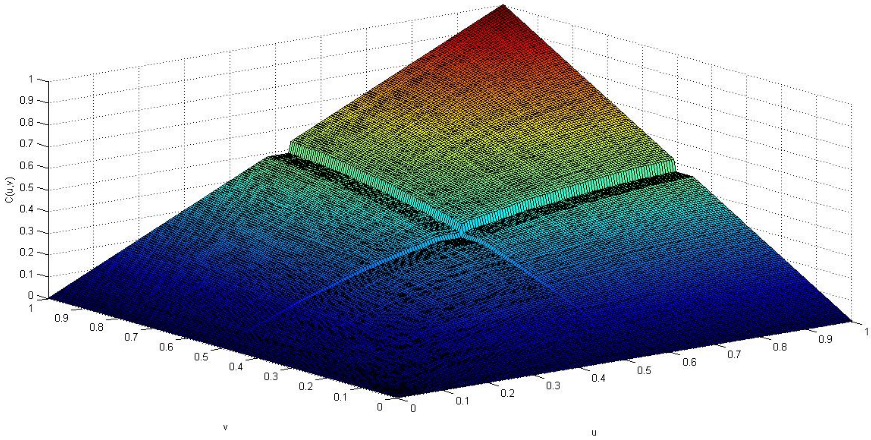
Step 2: Upper-Lower tail

In order to have more detailed clues on the tails, Figure (4.6) illustrates empirical upper and lower tail functions estimated as in (3.1)-(3.2).

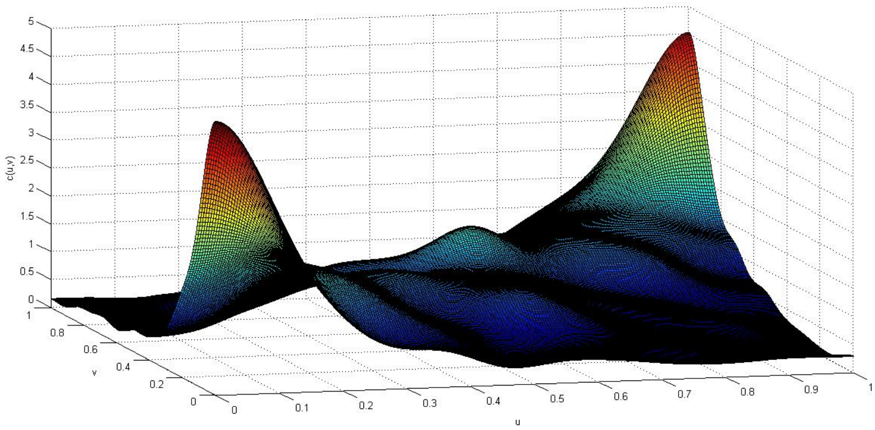
The green curves are average curves which are smoother than $\hat{\Lambda}_U$ and $\hat{\Lambda}_L$ and they are obtained by a bootstrap method: After s bootstrap resampling from $\{(u_i, v_i)\}_{i=1}^n$, say $\left\{(u_i^k, v_i^k)\right\}_{i=1}^n$, we obtain s different couple of trajectories, $\hat{\Lambda}_U^k$ and $\hat{\Lambda}_L^k$ from which we derive the average curves.
Smoothing the estimated functions $\hat{\Lambda}_U$ and $\hat{\Lambda}_L$ is a great way to better see possible symmetries in the tails and to better understand the limit behaviour.
Upper and lower tail parametes have been estimated from those curves according to *Chapter 3*

$$\hat{\lambda}_U = 0.3570$$

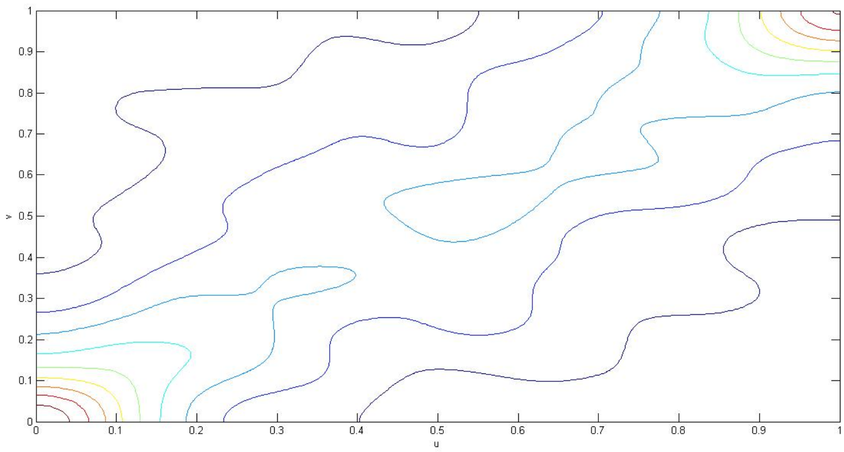
$$\hat{\lambda}_L = 0.3503.$$



(a) *Empirical copula surface*

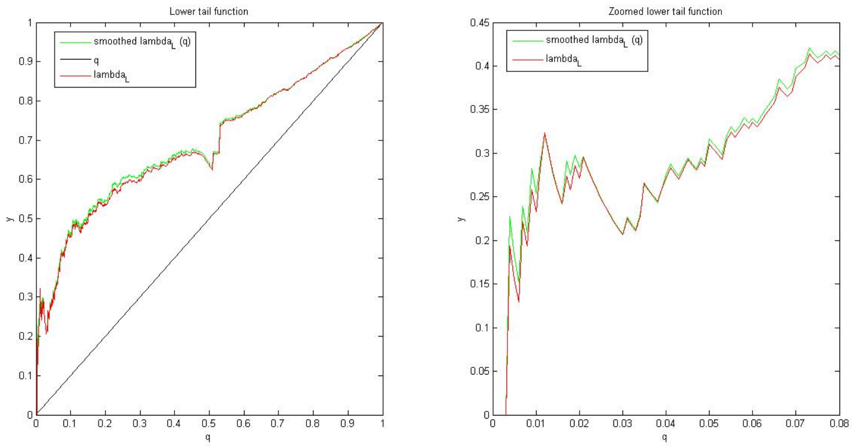


(b) *Kernel density estimator.*

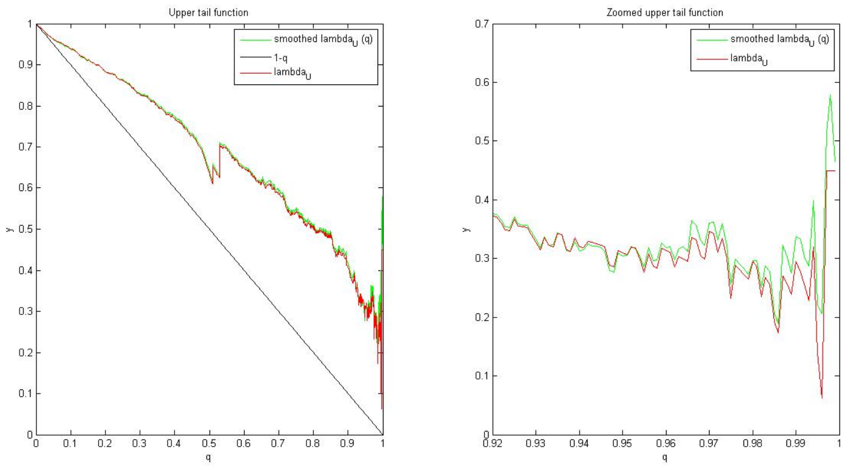


(c) *Kernel density estimator contours.*

Figure 4.5



(a) *Empirical lower tail function.*



(b) *Empirical upper tail function.*

Figure 4.6: Empirical tail functions

Both plots and estimates agree on a possible symmetry in tails with a reasonably chance of equality for upper and lower tail parameters.

4.2.1 Step 3: Goodness of fit

ML analysis
Among all the copula families introduced in *Chapter 2*, five reasonable models matching the features come up, like symmetry of the distribution and symmetry in the tails with equal limits, have been selected and four out of those are convex combinations:

1. $\mathfrak{C}_\theta = Student(\rho, n)$ - with $\theta = (\rho, n)$;

2. $\mathfrak{C}_\theta = pStudent(\rho, n) + (1-p)Frank(\alpha)$ - with $\theta = (p, \rho, n, \alpha)$;

3. $\mathfrak{C}_\theta = pClayton(\alpha) + (1-p)SurvivalClayton(\beta)$ - with $\theta = (p, \alpha, \beta)$;

4. $\mathfrak{C}_\theta = pClayton(\alpha) + (1-p)Gumbel(\beta)$ - with $\theta = (p, \alpha, \beta)$;

5. $\mathfrak{C}_\theta = pGumbel(\alpha) + (1-p)SurvivalGumbel(\beta)$ - with $\theta = (p, \alpha, \beta)$.

For each model $\mathfrak{C}_\theta$ the parameter θ has been estimated with the MLE and the related upper and lower tail coefficients have been computed.
All the results are reported in Table 4.1.

Family model	$\hat{\theta}^{ML}$	λ_U	λ_L
1. *Student*	$(0.5926, 5.5327)$	0.2399	0.2399
2. $Student + Frank$	$(0.7567, 0.5016, 5.2650, 10)$	0.1493	0.1493
3. $Clayton + SurvClayton$	$(0.4608, 2.0031, 0.8855)$	0.2465	0.326
4. $Clayton + Gumbel$	$(0.3347, 2.3869, 1.5173)$	0.2801	0.2503
5. $Gumbel + SurvGumbel$	$(0.4968, 1.3835, 2.1932)$	0.1737	0.3162

Table 4.1: *ML* parameter estimation

To compare the estimated models with the dataset, Figure 4.7 shows the kernel density estimator against all the fitted models,

and Figure 4.8 plots $\hat{\Lambda}_U$ and $\hat{\Lambda}_L$ against upper and lower tail functions of each model.

From the level curves the first two models, the Student t one and the convex combination of Student t and Frank, seem to fit quite well the data

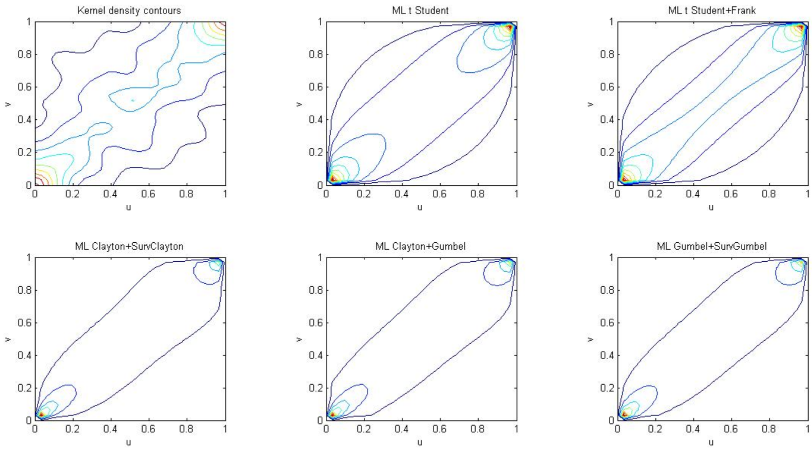


Figure 4.7: Kernel density VS *MLE* models contours.

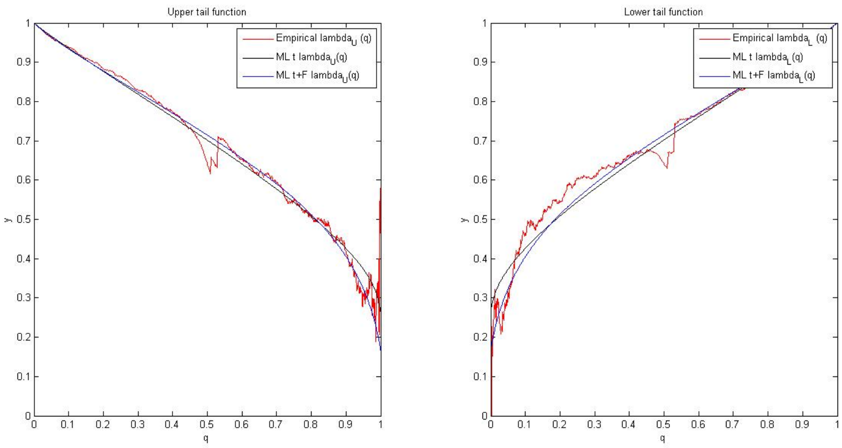
either in the central part and on the tails of the pdf, whereas the last three estimates lack the fit in the central part of the distribution.
As for the tails, the firts two estimated models have upper and lower tail functions rather close to the empirical ones, despite the limits are a bit lower; the other estimated models are instead a bit far from the empirical curves and for that reason we are pushed to discard them at this step.
However, to decide which model to test Table 4.2 reports the Kolmogorov-Smirnov distances for each model

Fitted copula model	KS distance
1. $Student$	0.0539
2. $Student + Frank$	0.0589
3. $Clayton + SurvClayton$	0.0539
4. $Clayton + Gumbel$	0.0546
5. $Gumbel + SurvGumbel$	0.0541

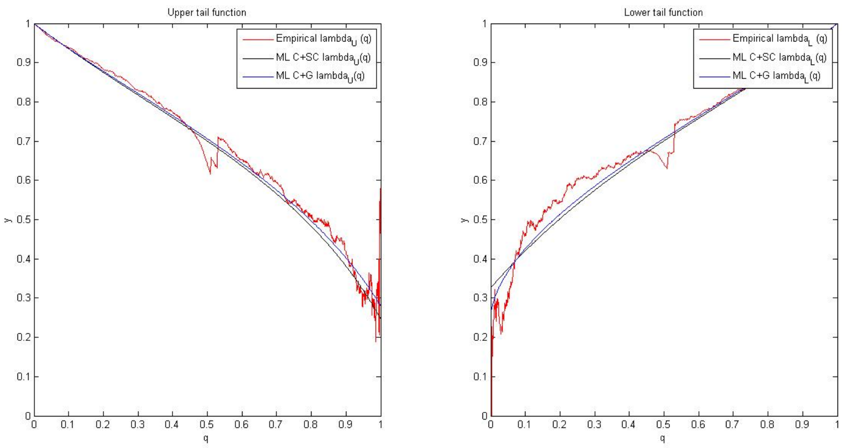
Table 4.2: K-S distance

and the five 3D-plots on the unit square in Figure 4.9 represent the differences $\left|\hat{C}_{XY}(u,v) - C_{\hat{\theta}^{ML}}(u,v)\right|$ for each model.

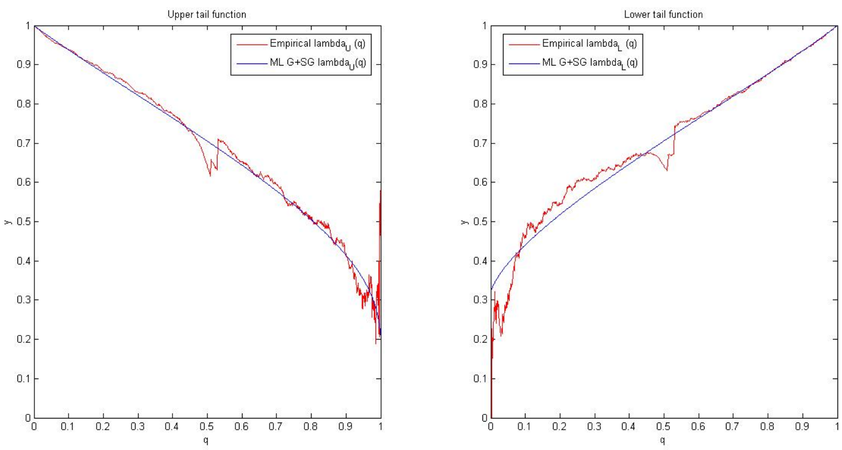
All the fitted cdfs distance almost the same from the empirical copula and the order of that distance is about 0.5 which is rather high. The models performing the least distance $KS = 0.0539$ are the Student t and the convex combination between the Clayton and the survival Clayton with, whereas the



(a)

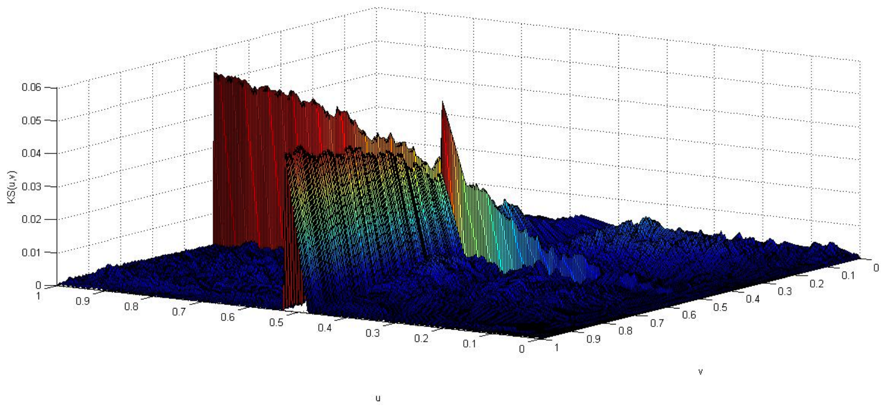


(b)

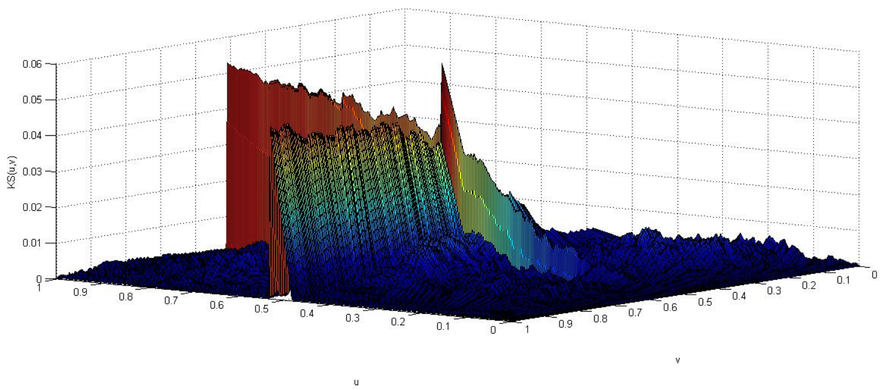


(c)

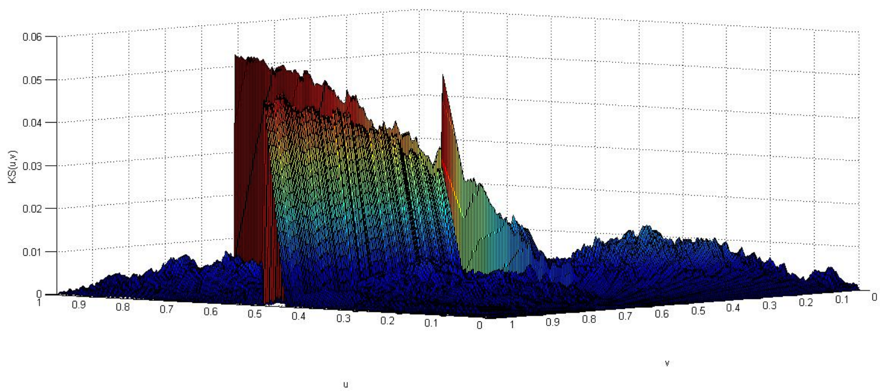
Figure 4.8: Empirical VS ML upper and lower tail functions.



(a) *ML* estimated Student

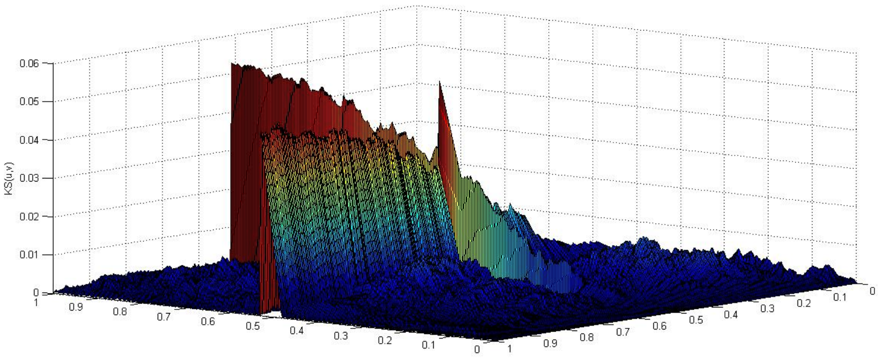


(b) *ML* estimated Student+Frank

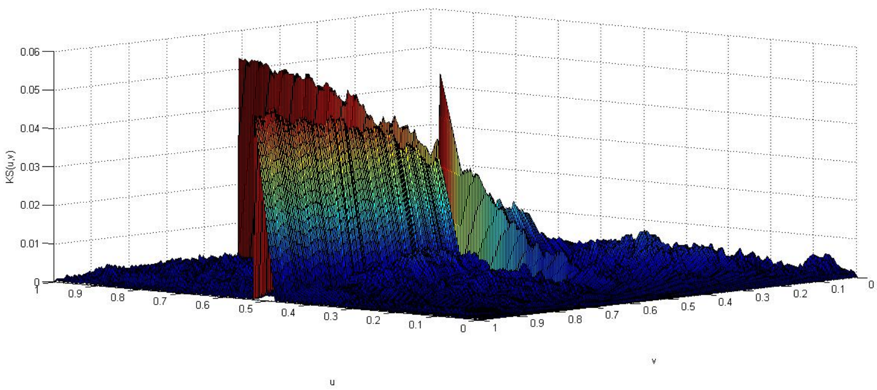


(c) *ML* estimated Clayton+SurvClayton

Figure 4.9: Kolmogorv-Smirnov distances.



(a) *ML* estimated Clayton+Gumbel



(b) *ML* estimated Gumbel+SurvGumbel

Figure 4.10: Kolmogorv-Smirnov distances.

one performing the largest distance $KS = 0.0589$ is the convex combination between Student t and Frank which was on the contrary a possible good fit. Anyway, since the approximated P-value is zero in the least distance case - the Student t -, all the five models are rejected.
It can be said more, the 3-D plots reveal that each model reaches the highest distance exactly where the empirical copula shows the discontinouity because X and Y are not continous: the lack of uniformity is the reason of a falsified analysis.

K-S minimizing analysis

A possible way to circumvent the problem of too large KS values could be estimating the parameter θ by minimizing the K-S distance.
In other words, assuming $C_{XY} \in \mathfrak{C}_\theta$, the K-S distance between C_{XY} and the generic element of $\mathfrak{C}_\theta$, say C_θ, is

$$KS(\theta) := \max_{(u,v)\in[0,1]^2} \left| \hat{C}_{XY}(u,v) - C_\theta(u,v) \right|.$$

We estimate θ with the value performing the least K-S distance, i.e.

$$\hat{\theta} = arg \min_{\theta\in\Omega} KS(\theta),$$

where Ω is the domain of θ.
Table 4.3 summarize the new estimates

Family model	$\hat{\theta}$	λ_U	λ_L
1. *Student*	$(0.4999, 2.5)$	0.3488	0.3488
2. *Student + Frank*	$(0.4672, 0.2, 7.5001, 7.492)$	0.0031	0.0031
3. *Clayton + SurvClayton*	$(0.4116, 3.7154, 0.01)$	0	0.3415
4. *Clayton + Gumbel*	$(0.8140, 0.3685, 3.4410)$	0.1445	0.1241
5. *Gumbel + SurvGumbel*	$(0.3164, 3.9808, 1.01)$	0.2562	0.009

Table 4.3: *K-S minimizing* parameter estimation.

whereas Figure 4.11 compares the kernel density level with the fitted densities,

and Figure 4.12 compares the empirical tail curves with the fitted ones.

Despite the new fitted models perform the least K-S distances -see Table 4.4- they are further from the data than the *ML* ones, both in middle of the distribution and in the tails, except model 2.

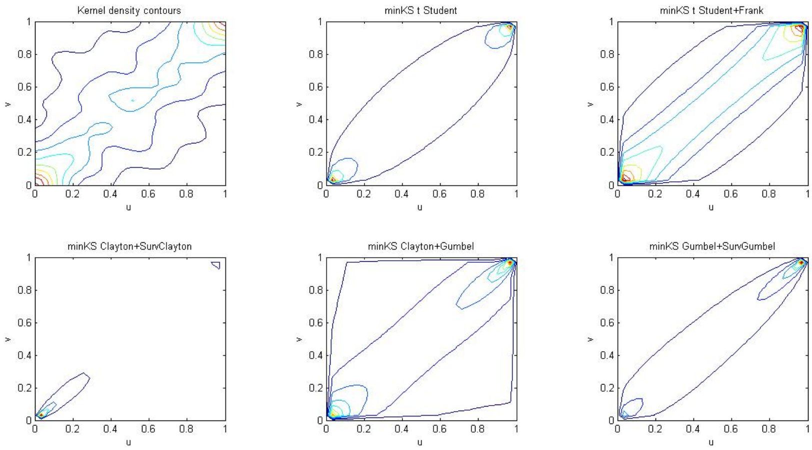


Figure 4.11: Kernel density VS *minKS* models contours.

Fitted copula model	KS distance
1. $Student$	0.0534
2. $Student + Frank$	0.0419
3. $Clayton + SurvClayton$	0.0505
4. $Clayton + Gumbel$	0.0514
5. $Gumbel + SurvGumbel$	0.0520

Table 4.4: least K-S distance

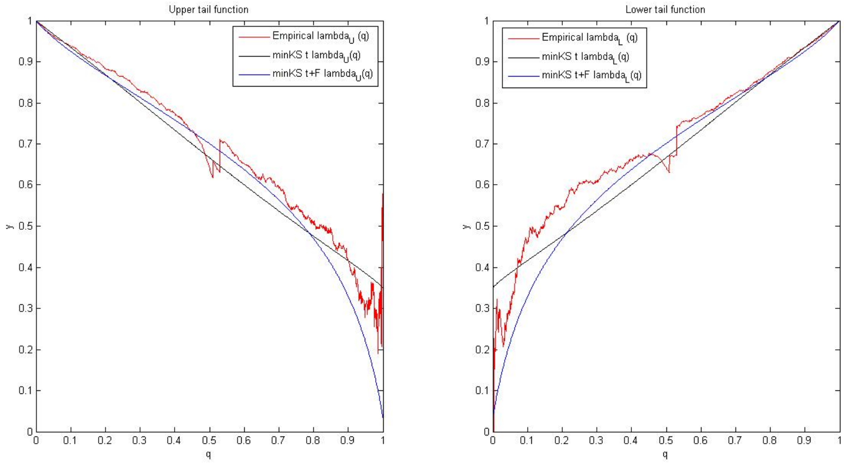
Smaller distances and worse fitting are easily justified if we take a look at Figures 4.13-4.14.

where each model actually performs lower distances than the ML case but to the detriment of tails. Thus, since we are more interested in modelling the tails than the central part of the distribution, the KS minimizing estimation method will not be taken into account in the rest of the work.

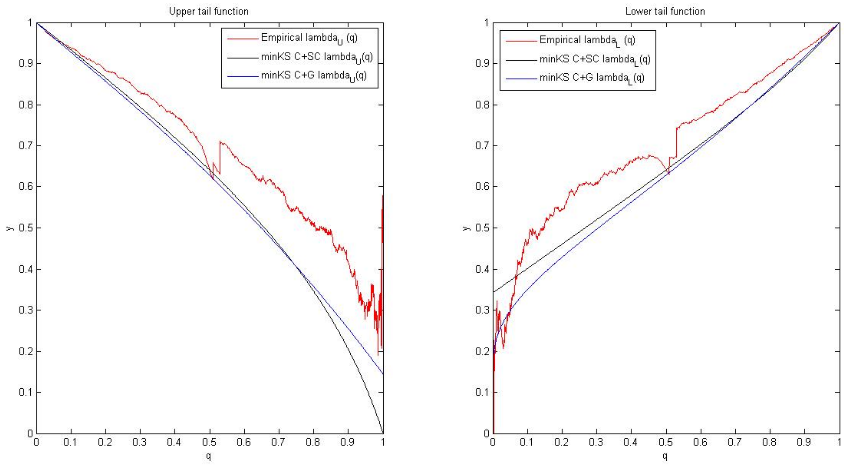
As the analysis is falsified because of too many zero log-returns, two possible options to better study the problem are:

1. Perturbing the return dataset with a noise distributed as uniform on $[-0.005, 0.005]$;

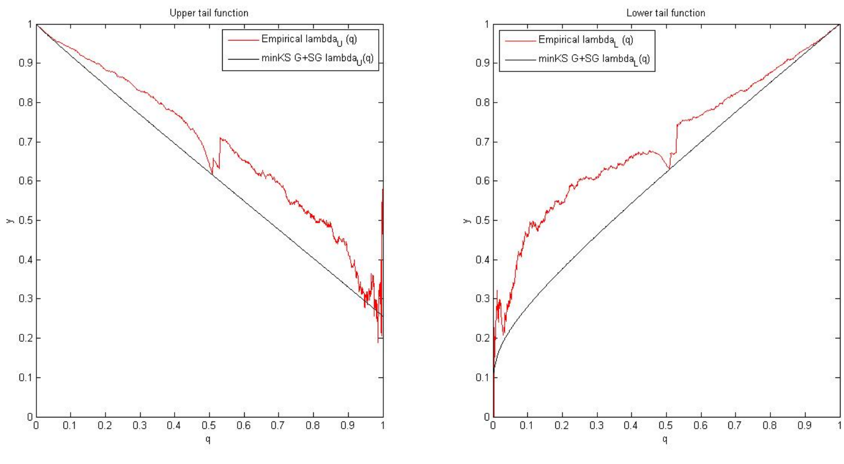
2. Removing the jointly zero couples from the log-returns dataset.



(a)

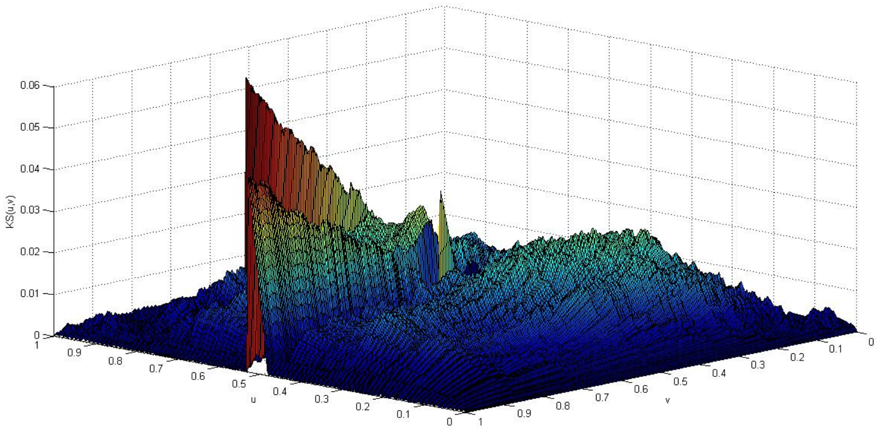


(b)

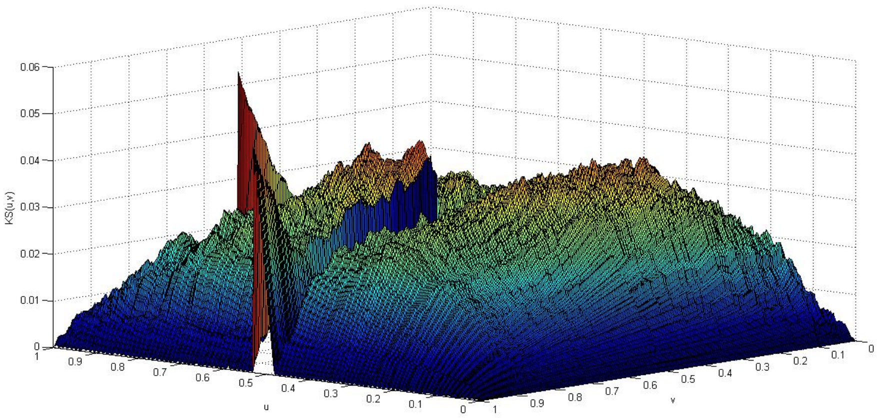


(c)

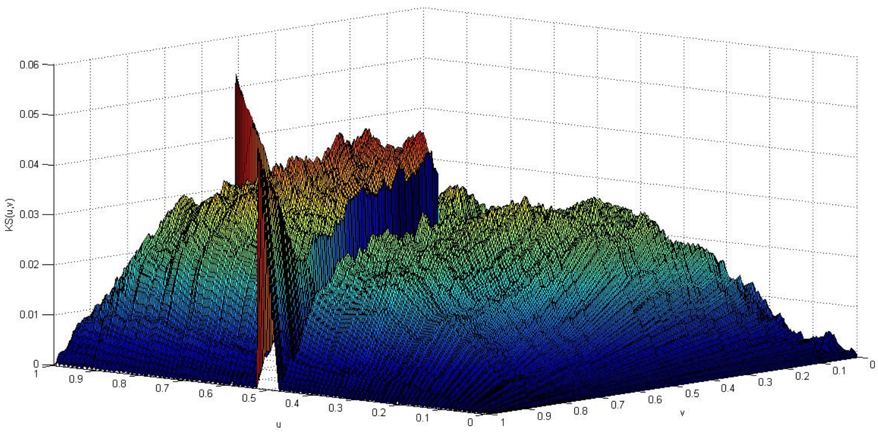
Figure 4.12: Empirical VS *minKS* upper and lower tail functions.



(a) *minKS* estimated Student

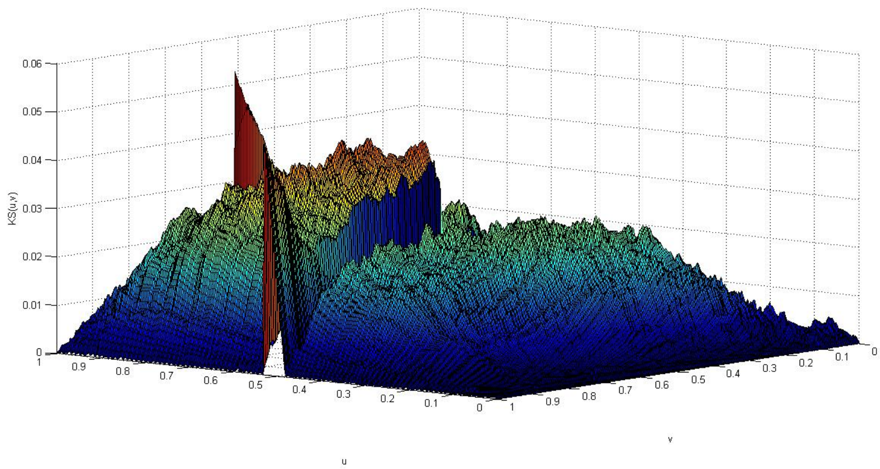


(b) *minKS* estimated Student+Frank

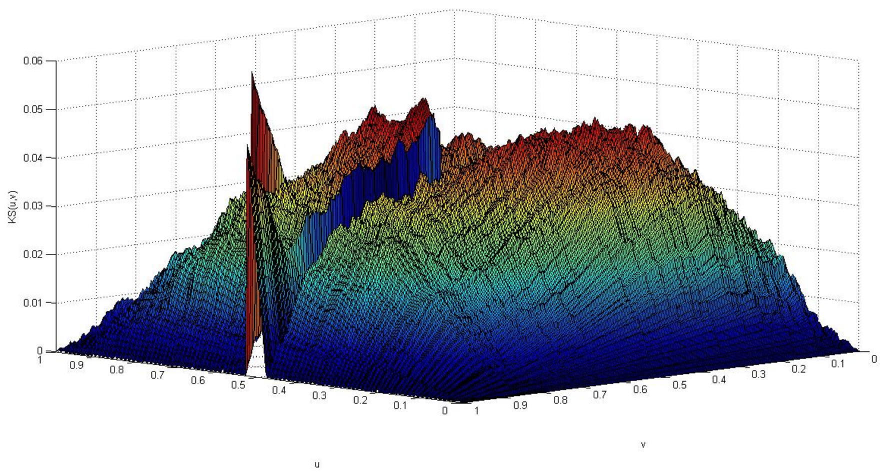


(c) *minKS* estimated Clayton+SurvClayton

Figure 4.13: Kolmogorv-Smirnov distances.



(a) *minKS* estimated Clayton+Gumbel



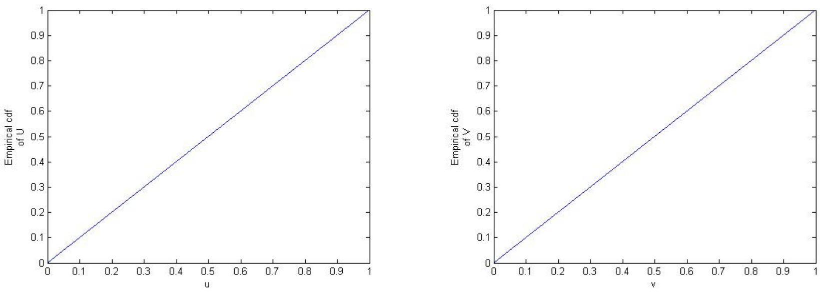
(b) *minKS* estimated Gumbel+SurvGumbel

Figure 4.14: Kolmogorv-Smirnov distances.

The final part of the work consists in studying the two cases and in comparing the risults.

4.2.2 Uniform noise - dataset

Since the downloaded stock returns are rounded to two decimals, they could be stationary in days because of rounding errors. It is therefore reasonable to perturb the dataset with a noise distributed as a uniform on $[-0.005, 0.005]$ and then re-perform the analysis.
Empirical distributions of the new datasets $\{u_i\}_{i=1}^{n-1}$ and $\{v_i\}_{i=1}^{n-1}$ are reported in Figure (4.15). From the graph they look like cdfs of a uniforms on $[0, 1]$, and the Kolmogorov Smirnov test confirms it indeed.



(a) *Empirical cdf of U* (b) *Empirical cdf of V*

Figure 4.15: Empirical distribution - uniform noise.

Step 1: Empirical copula

The little perturbations have just “regualirezed” the dataset - in the sense that the empirical copula is now smoother then before - without changing the dependence structure. In fact, the contour plot in Figure 4.16 suggests the same positive dependence we had before

and Figure 4.17 confirms that there is still symmetry also in the tails.

Step 2: Upper-Lower tail

To be more meticulous in the tail bahaviour, Figure 4.18

and limit estimates

$$\hat{\lambda}_U = 0.3575$$

$$\hat{\lambda}_L = 0.3646$$

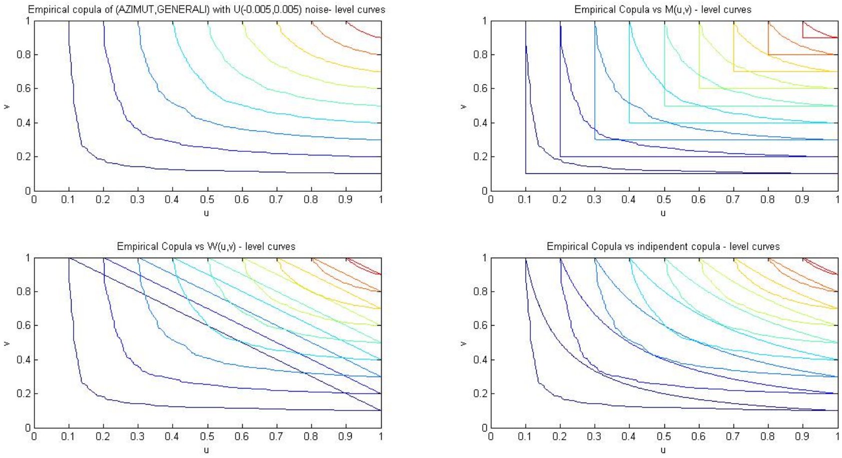


Figure 4.16: Empirical copula VS main copulas contours - uniform noise.

confirm again the hypotesis of symmetry in tails funtions, hence equality in their limits.

Step 3: Goodness of fit

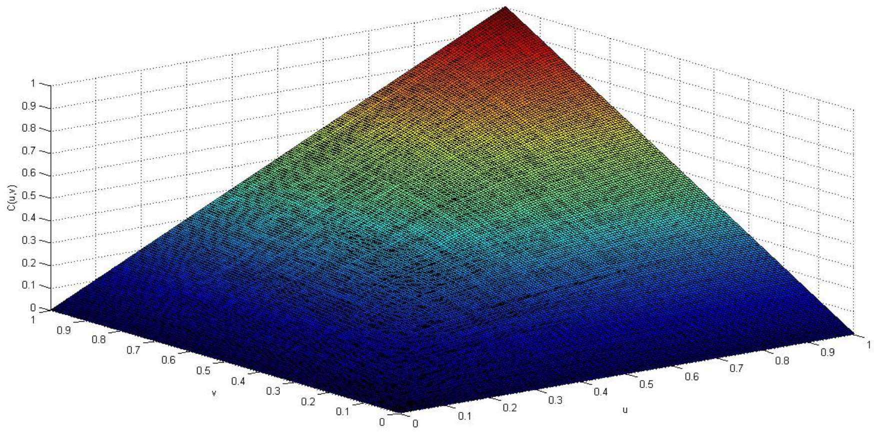
Taking into account the same family models as before, Table 4.5 reports their ML estimated parameters.

Family model	$\hat{\theta}^{ML}$	λ_U	λ_L
1. $Student$	$(0.5913, 5.4225)$	0.2434	0.2434
2. $Student + Frank$	$(0.7603, 0.5013, 5.1580, 10)$	0.2014	0.2014
3. $Clayton + SurvClayton$	$(0.4639, 1.9888, 0.8777)$	0.2438	0.3274
4. $Clayton + Gumbel$	$(0.3418, 2.3292, 1.5143)$	0.2761	0.2538
5. $Gumbel + SurvGumbel$	$(0.4923, 1.3777, 2.1855)$	0.1704	0.3182

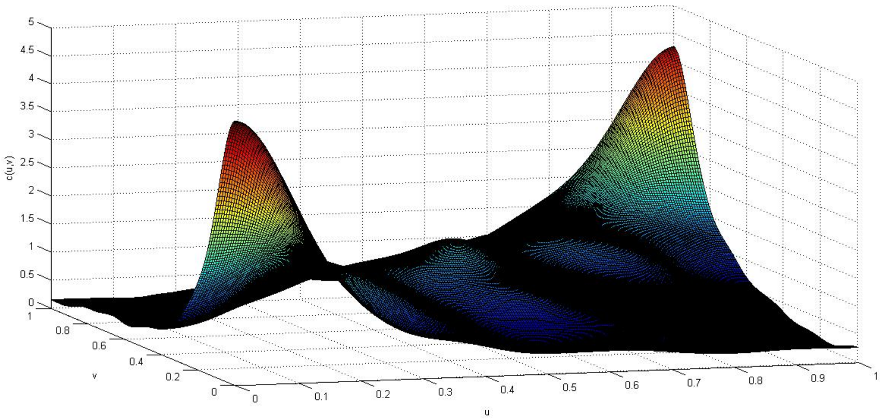
Table 4.5: *ML* parameter estimation - uniform noise.

From Figure 4.19 and Figure 4.20, again the first two models are the candidate ones to be the best fit, but before concluding the analysis we first need to have a look at the KS distances reported in Table 4.6.

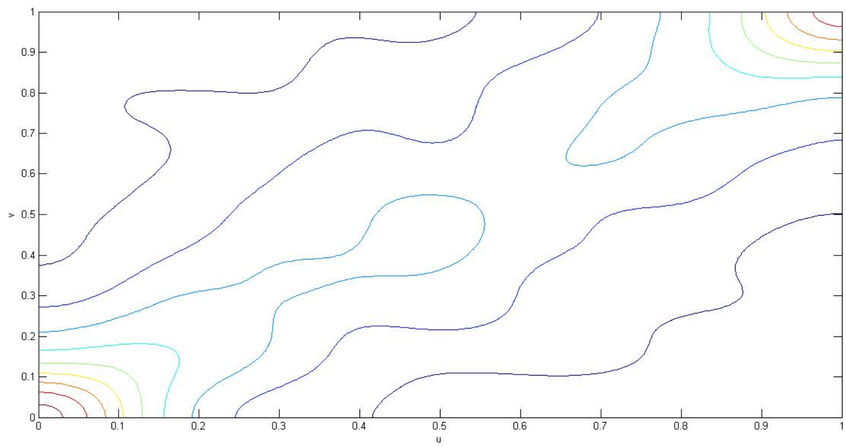
Distances are manifestly smaller than the first dataset and the least ones are reached by the Student t copula ($KS = 0.0126$) and the two Archimedean



(a) *Empirical copula surface*

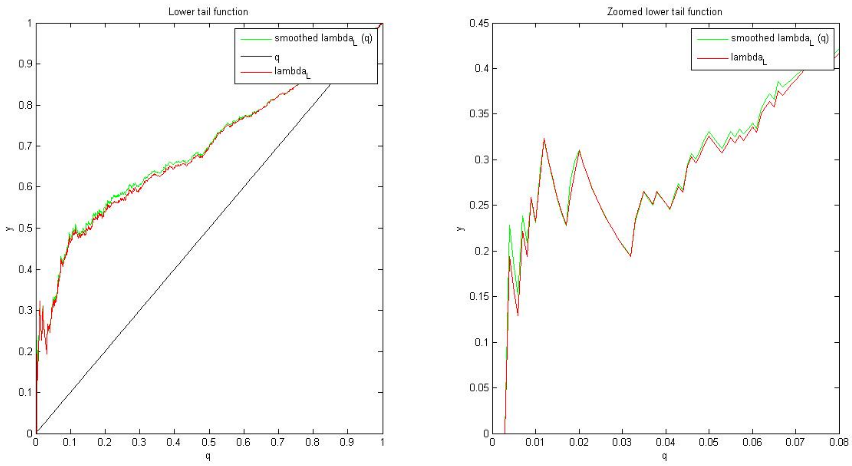


(b) *Kernel density estimator*

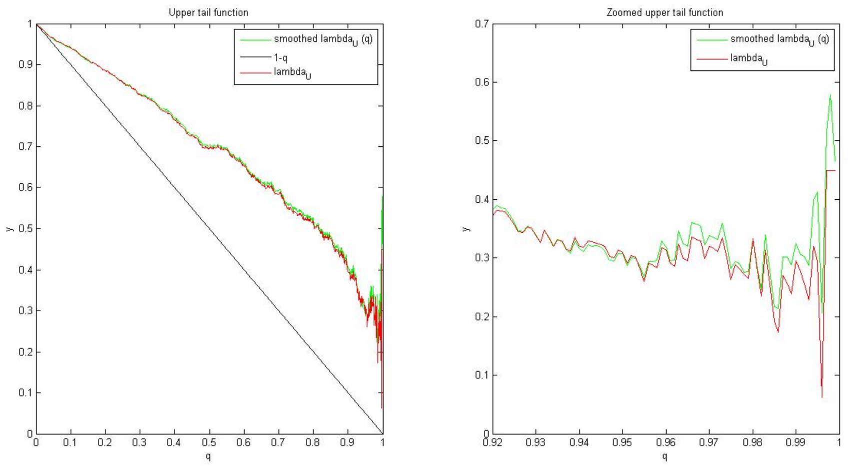


(c) *Kernel density estimator contours*

Figure 4.17: Empirical copula - uniform noise.



(a) *Empirical lower tail function*



(b) *Empirical upper tail function*

Figure 4.18: Empirical tail functions - uniform noise.

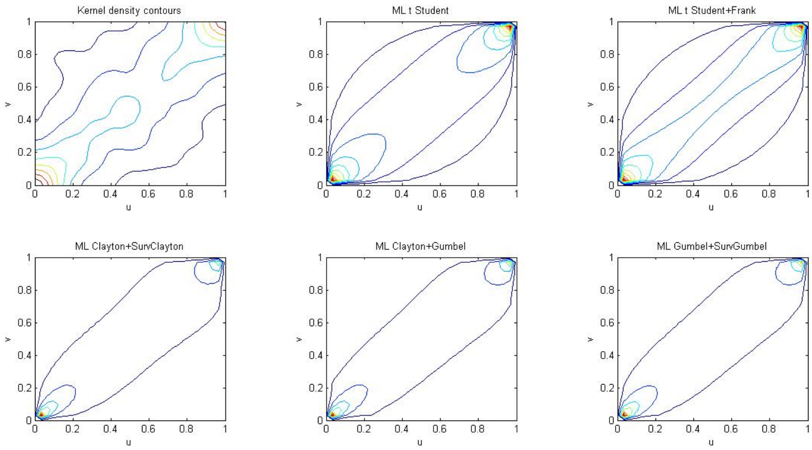


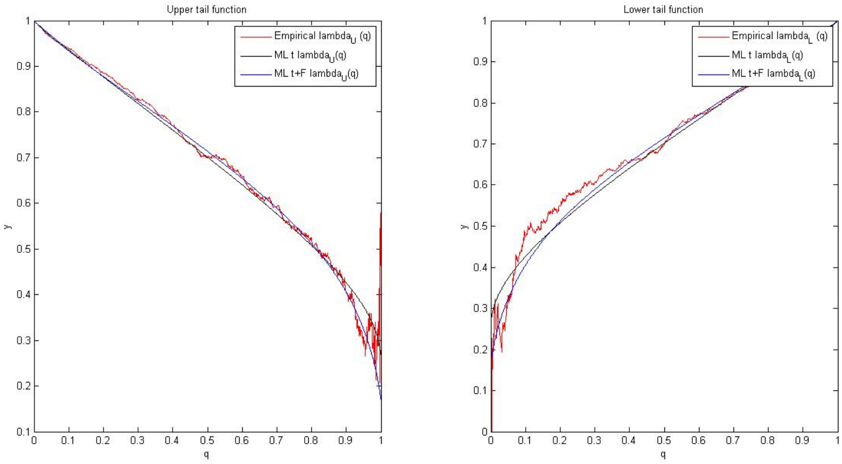
Figure 4.19: Kernel density VS *MLE* models contours - uniform noise.

Fitted copula model	KS distance
1. $Student$	0.0126
2. $Student + Frank$	0.0170
3. $Clayton + SurvClayton$	0.0139
4. $Clayton + Gumbel$	0.0144
5. $Gumbel + SurvGumbel$	0.0136

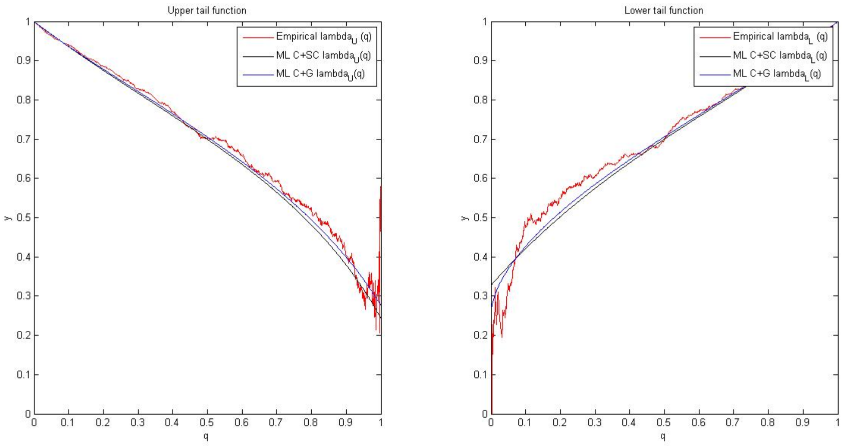
Table 4.6: K-S distances - uniform noise

convex combination Gumbel+SurvivalGumbel and Clyton+SurvivalClayton ($KS = 0.0136$ and $KS = 0.0139$); the largest distance is instead reached by the Frank+Student copula ($KS = 0.017$) and this fact is discordant with what deduced in the estimated contour plots.
Anyway, to decide which model to test and which p value to compute, let's take a look at Figures 4.21-4.22.

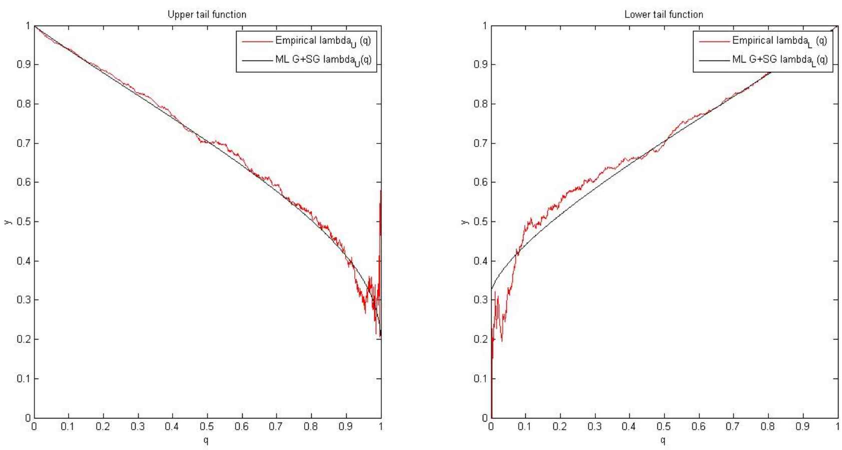
Contrary to what we were expecting, the *Frank+Student* estimate seems to be the best model to test because it performs the best fit in the tails (the maximum distance is actually reached in the middle of the distribution) then any other model do.
However it should be notice that if testing the Frank+Student copula the related pvalue is sufficiently high to not reject the model, then all the other one could not be rejected, being $KS = 0.017$ the maximum distance, but all



(a)

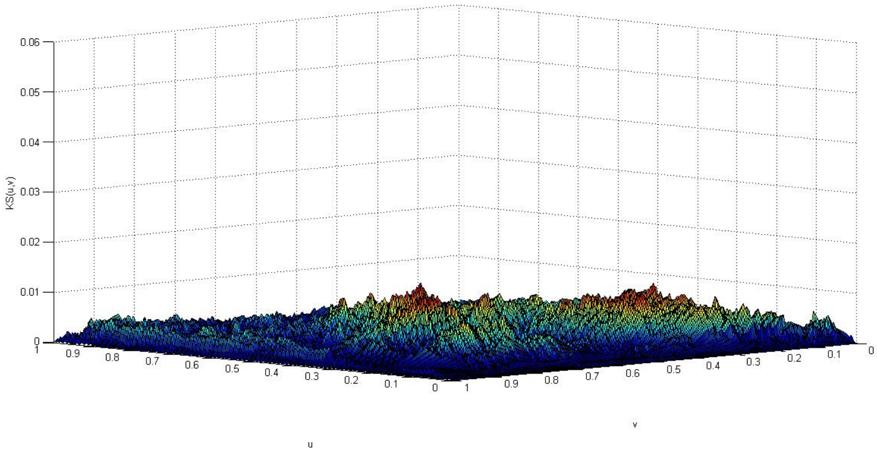


(b)

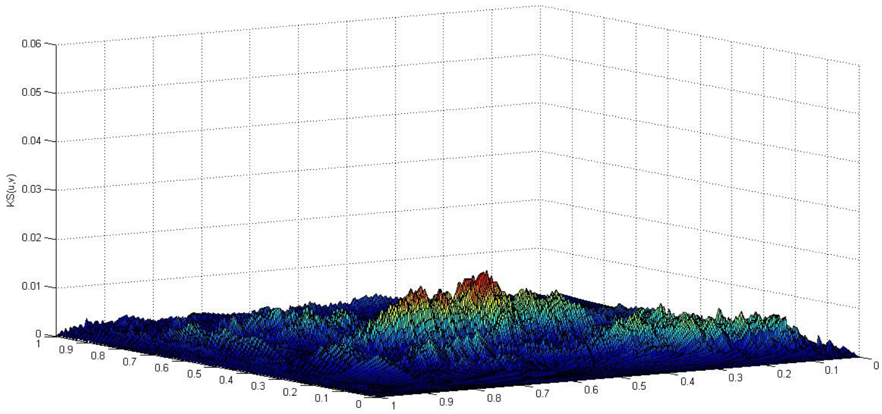


(c)

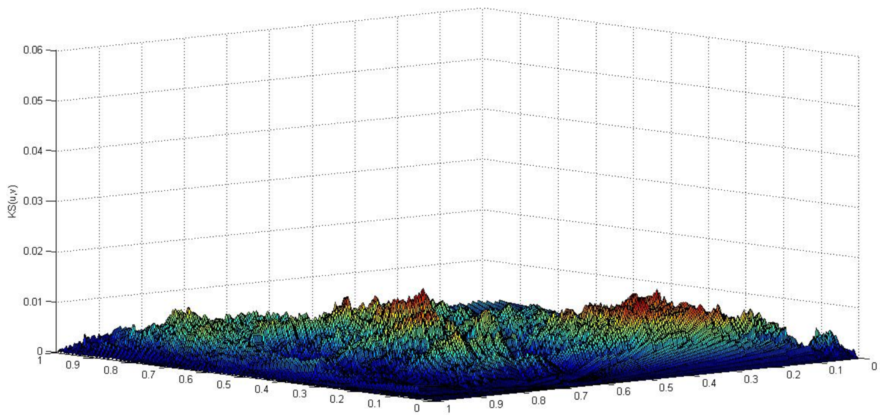
Figure 4.20: Empirical VS *ML* Upper and lower tail functions - uniform noise.



(a) *ML* estimated Student

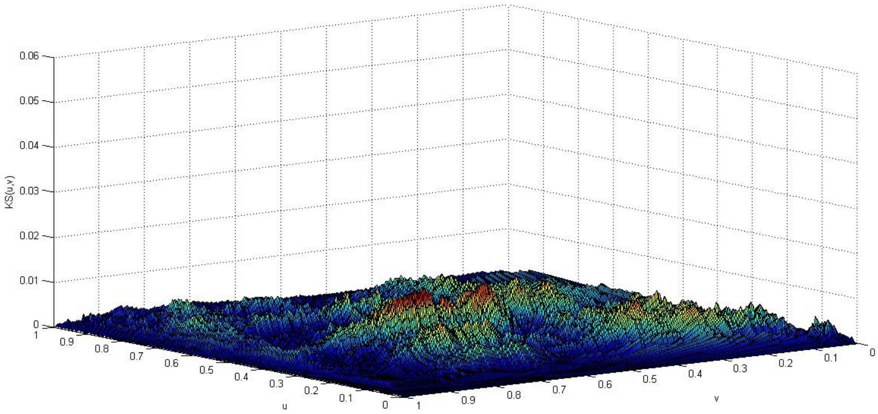


(b) *ML* estimated Student+Frank

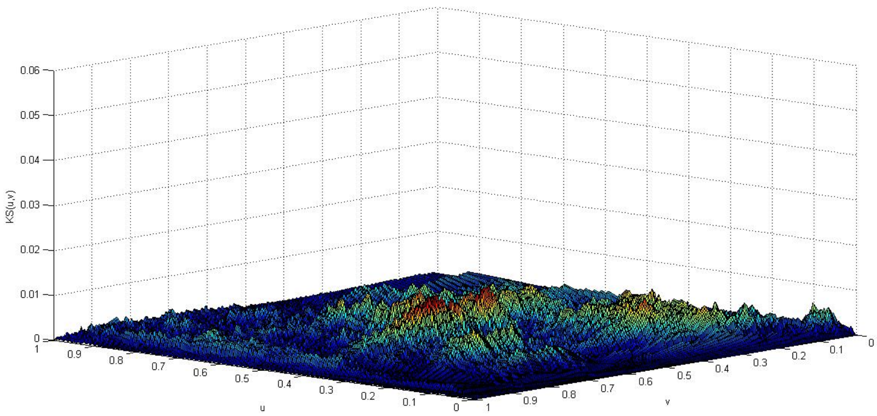


(c) *ML* estimated Clayton+SurvClayton

Figure 4.21: Kolmogory-Smirnov distances - uniform noise.



(a) *ML* estimated Clayton+Gumbel



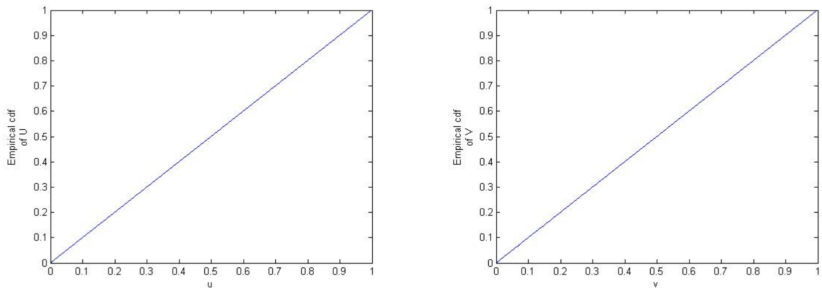
(b) *ML* estimated Gumbel+SurvGumbel

Figure 4.22: Kolmogorv-Smirnov distances - uniform noise.

the analysis done motivates anyway the choice of Frank+Student.

4.2.3 No joint zeros - dataset

An alternative way to bypass the problem of too many zero log-returns consists trivially in removing them from the dataset. By doing that we are essentially interpreting the stationary days as days in which there have been no trading of the two stocks, hence negligible days for the analysis.
Similarly to the uniform noise case, the *no joint zeros* dataset $\{(u_i, v_i)\}_{i=1}^{n-1}$ has margins manifestly uniform on $[0, 1]$ - see their empirical cdfs in 4.23-,



(a) *Empirical cdf of U* - no joint zeros.

(b) *Empirical cdf of V* - no joint zeros.

Figure 4.23

and the Kolmogorov-Smirnov test confirms it.

Step 1: Empirical copula

The new "regularized" dataset, again, keeps the same dependence structure as the original one but with a smoother empirical copula. In fact, Figures 4.24-4.25 still suggest positive dependence and symmetry.

Up to now there are no valuable differences between the two alternative approaches -uniform noise approach and no joint zeros approach-, which are simply two different ways to regularize the dataset.

Step 2: Upper-Lower tail

Same remarks hold true for the tails behaviour: in Figure 4.26 the empirical tail functions are of course slightly smoother than the first two datasets, but the trend is basically equal.

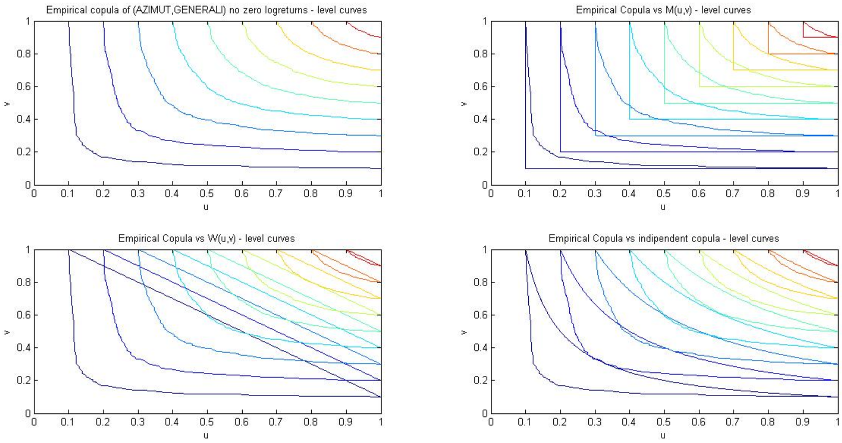


Figure 4.24: Empirical copula VS main copulas contours -
no joint zeros.

As a consequence, also the estimated limits should not be far from the previous estimates, actually

$$\hat{\lambda}_U = 0.333$$

$$\hat{\lambda}_L = 0.3411.$$

Hence, there is still symmetry in the tails and equality in the limits.

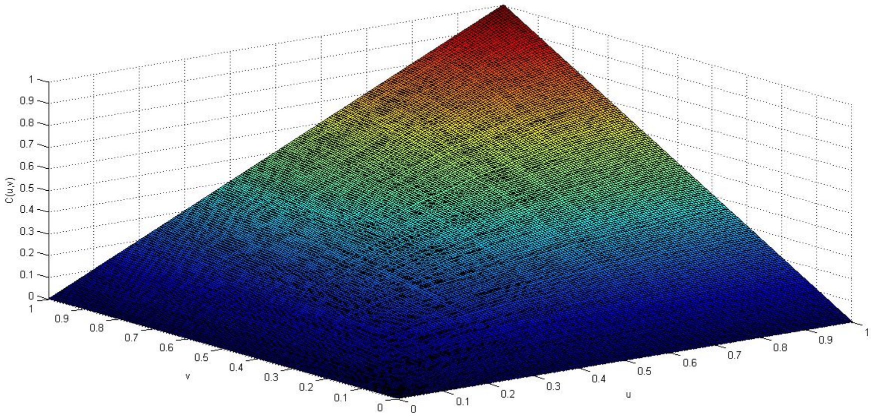
Step 3: Goodness of fit

In Table 4.7, the ML estimated parameters for the five models are reported.

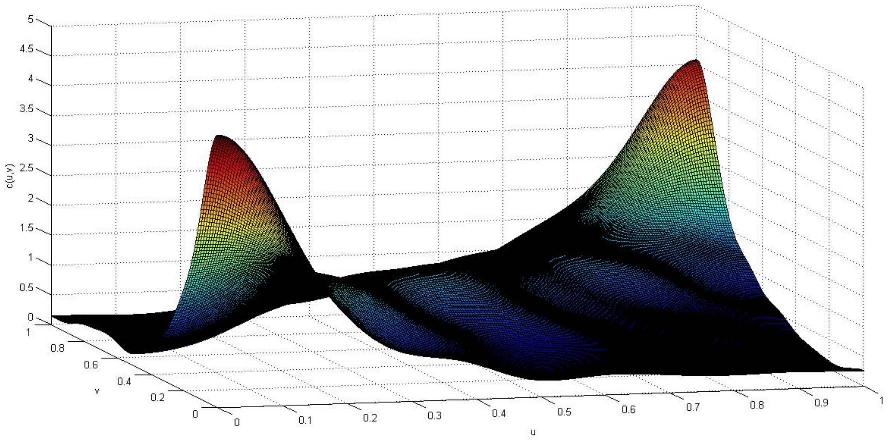
Family model	$\hat{\theta}^{ML}$	λ_U	λ_L
1. *Student*	$(0.6113, 5.7687)$	0.2434	0.2434
2. *Student* + *Frank*	$(0.7010, 0.5086, 5.0633, 9.3314)$	0.1465	0.1465
3. *Clayton* + *SurvClayton*	$(0.4941, 2.0577, 0.9285)$	0.2658	0.3528
4. *Clayton* + *Gumbel*	$(0.3703, 2.3979, 1.5459)$	0.2734	0.2773
5. *Gumbel* + *SurvGumbel*	$(0.4471, 1.3973, 2.1763)$	0.16	0.3455

Table 4.7: *ML* parameter estimation -
no joint zeros.

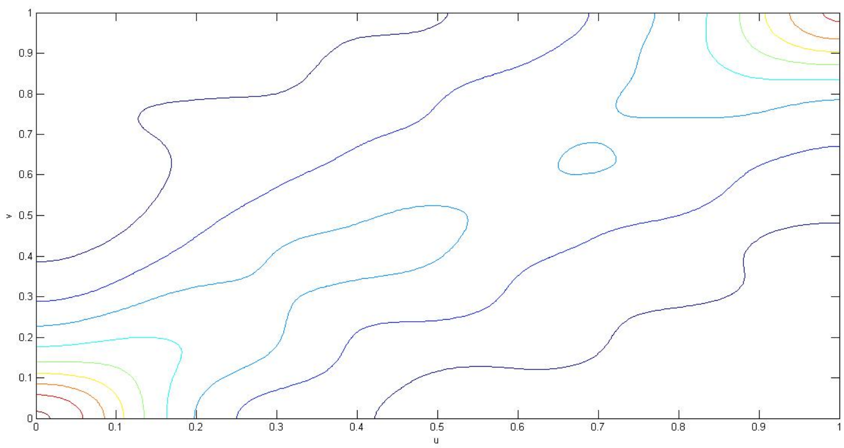
Figure 4.27 shows as usual their densities against the kernel estimator, whereas 4.28 shows the fitted tail functions against the empirical ones.



(a) *Empirical copula surface*

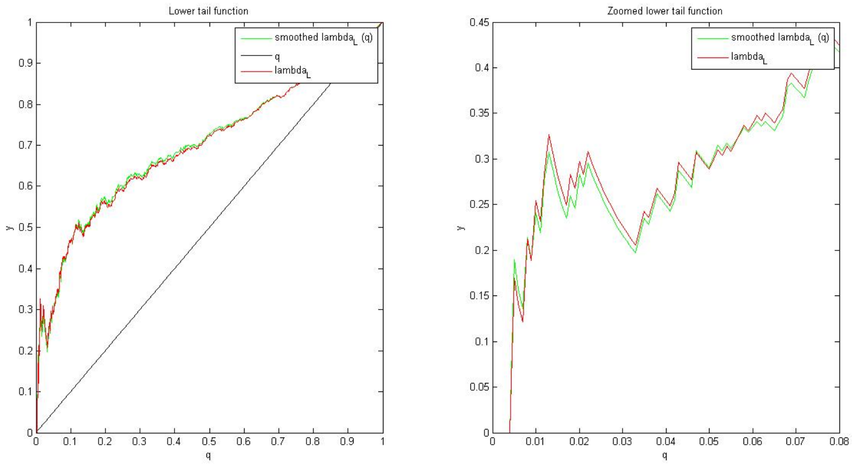


(b) *Kernel density estimator - no joint zeros.*

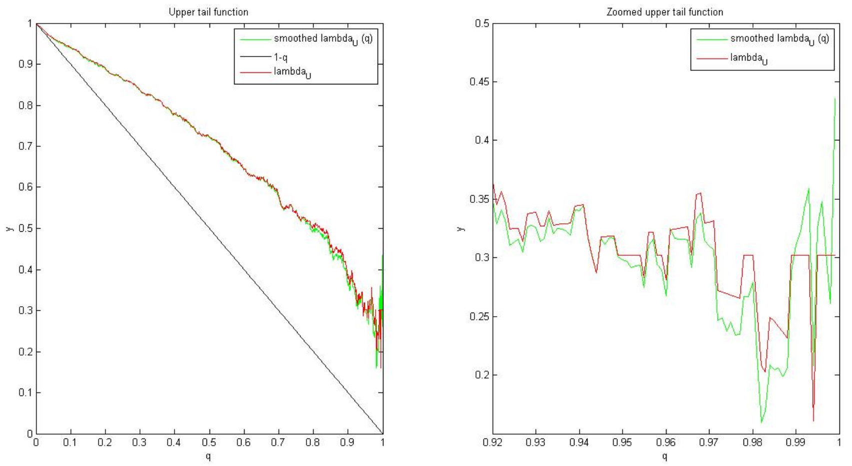


(c) *Kernel density estimator contours - no joint zeros.*

Figure 4.25



(a) *Empirical lower tail function - no joint zeros.*



(b) *Empirical upper tail function - no joint zeros.*

Figure 4.26

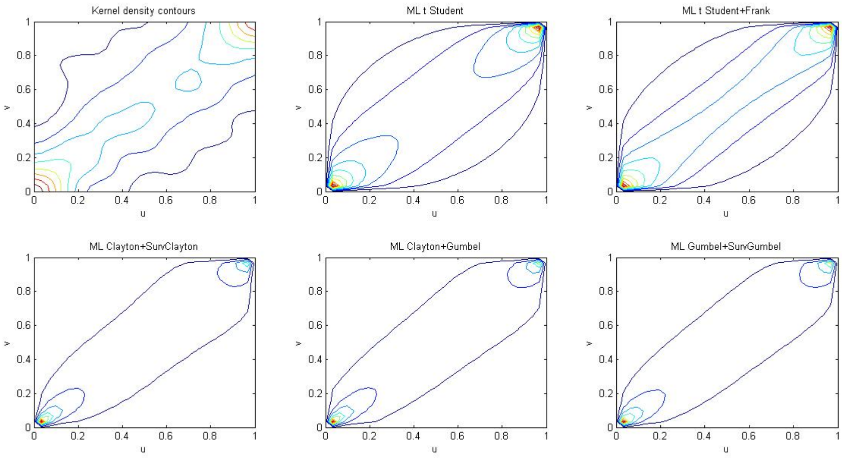


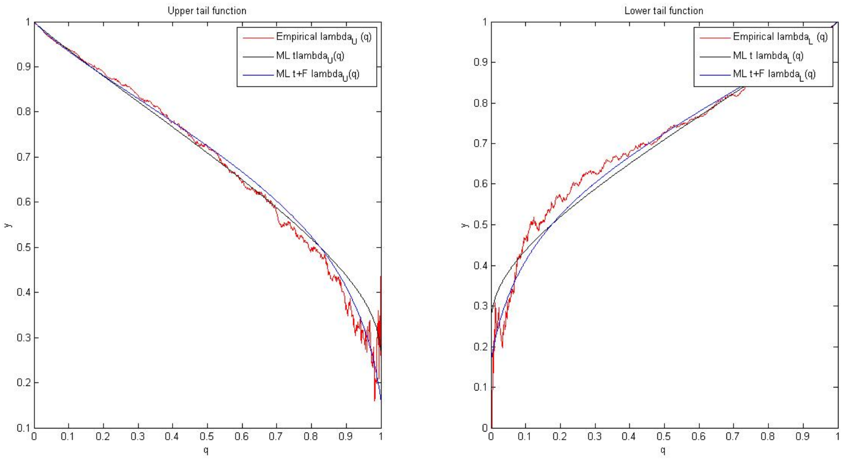
Figure 4.27: Kernel density VS *MLE* models contours -
no joint zeros.

The first two models fit better the whole distribution in the contour plot, but in the tails ones the last three are closer to data, in particulas model 3 and 4. The latter two are also the ones performing the least KS distance as it can be seen from Table 4.8 and precisely, if we sort the five models by their KS distance we find that model 4 performs the least one $KS = 0.0132$ then model 5 and 2 with $KS = 0.0137$ and $KS = 0.0138$ while the greatest distances are reached by model 3 and 1 with $KS = 0.0161$ and $KS = 0.0166$.

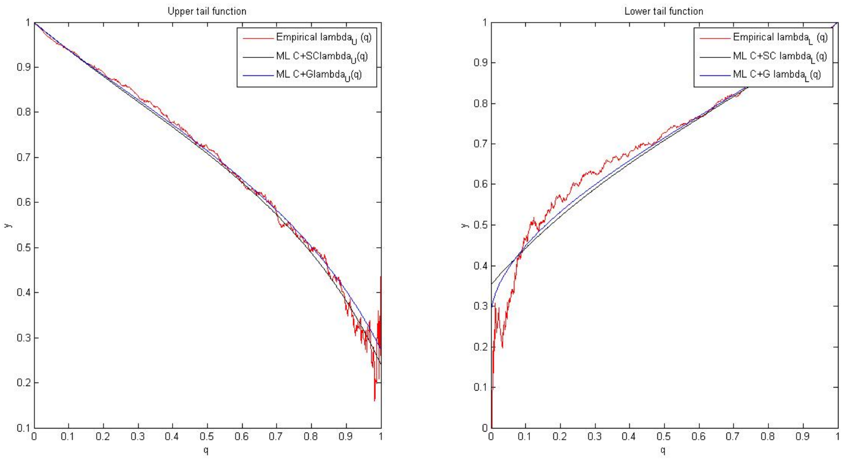
Fitted copula model	KS distance
1. $Student$	0.0166
2. $Student + Frank$	0.0138
3. $Clayton + SurvClayton$	0.0161
4. $Clayton + Gumbel$	0.0132
5. $Gumbel + SurvGumbel$	0.0137

Table 4.8: K-S distances -
no joint zeros.

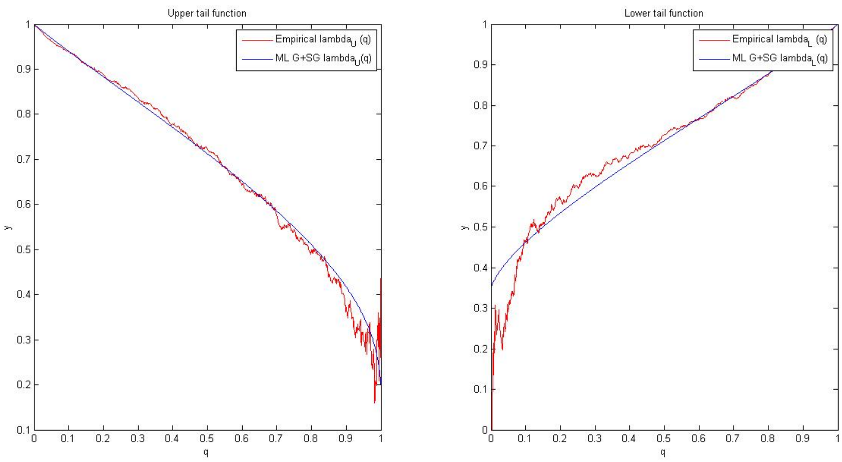
Notice that there is not such a difference between the closest and the furthest model in terms of KS order $0.0132 \to 0.0166$ and for that reason it is more valuable have a look at the regions where the maximum distance is reached. In fact, Figures 4.29-4.30 give evidence that the three



(a)



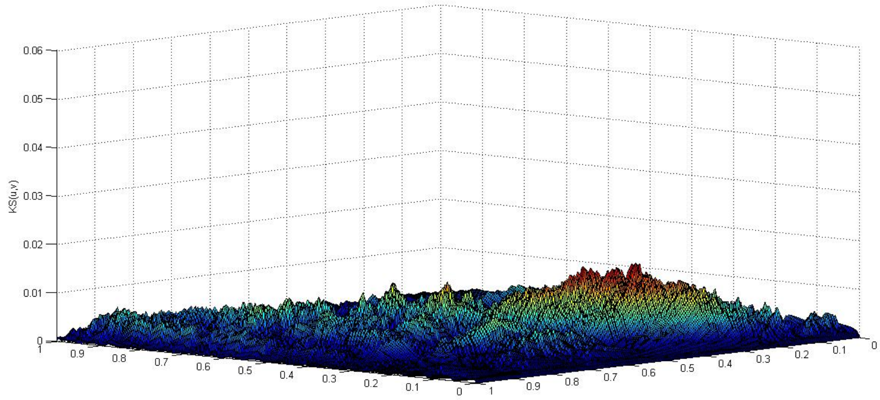
(b)



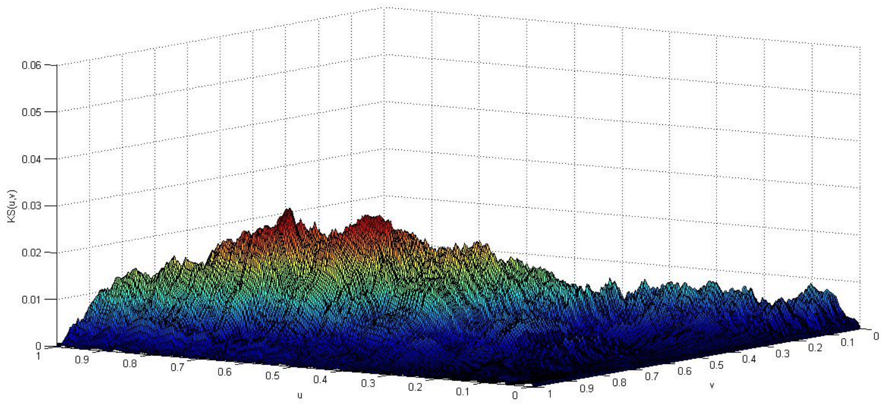
(c)

Figure 4.28: Empirical VS *ML* upper and lower tail functions - no joint zeros.

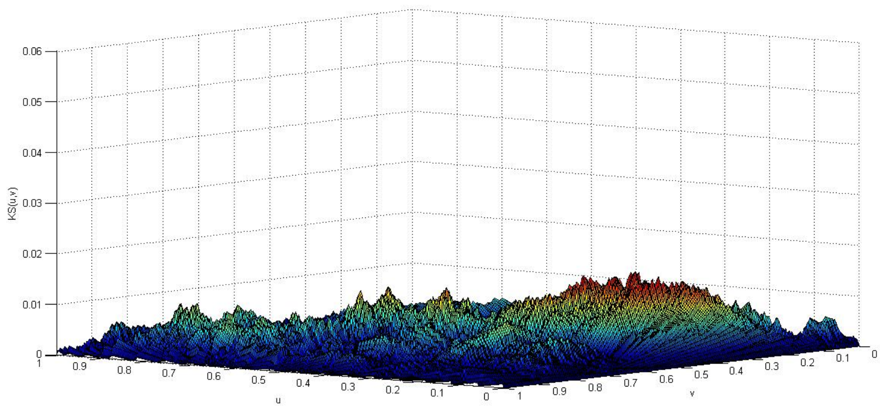
Archimedean convex combinations distances most in the down-right corner of the square where the highest spikes are located. The same holds true for the Frank+Student copula, but not for the Student copula which, despite the highest KS distance, distances most in the up-left corner of the square.
In practical terms, if we accept to model data with one among the previous copulas but not the Student one we consciusely underestimating probabilities of default, whereas if we model with the Student we areunderestimating probabilites of booming, which is reasonably less risky.
For that reason, again we convey to compute the pvalue in the greatest KS distance, which is the Student copula.



(a) *ML* estimated Student

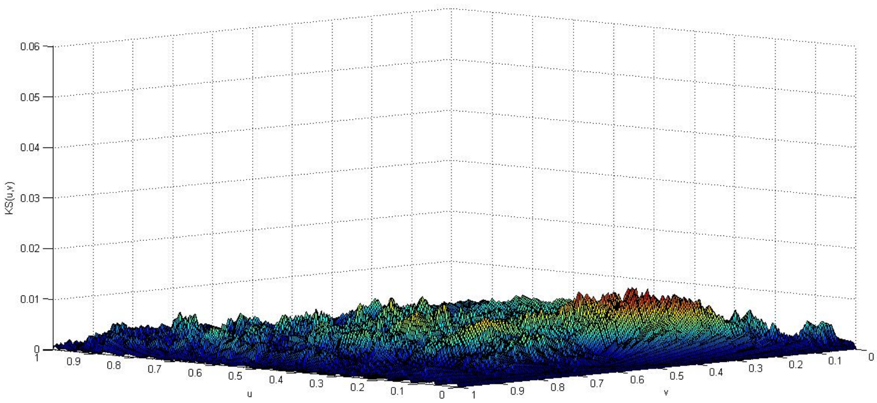


(b) *ML* estimated Student+Frank

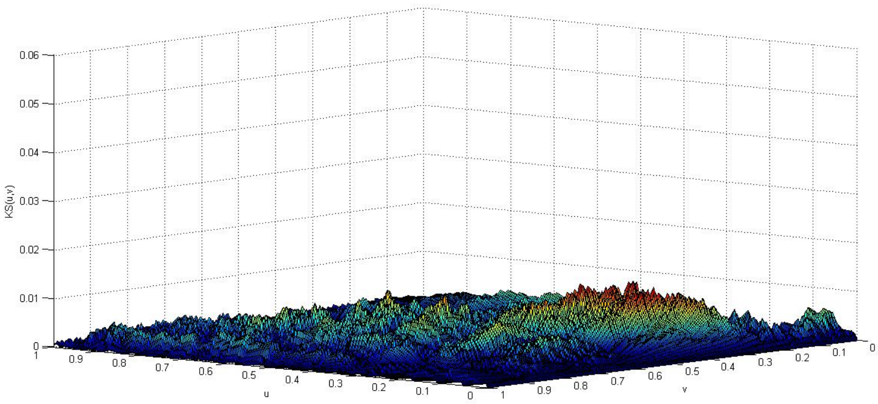


(c) *ML* estimated Clayton+SurvClayton

Figure 4.29: Kolmogorv-Smirnov distances - no joint zeros.



(a) *ML* estimated Clayton+Gumbel



(b) *ML* estimated Gumbel+SurvGumbel

Figure 4.30: Kolmogorv-Smirnov distances - no joint zeros.

Bibliography

[1] Roger B. Nelsen (1999). *An Introduction to Copulas.* 2nd edition, Springer.

[2] U. Cherubini, E. Luciano, W. Vecchiato (2004). *Copula Methods in Finance.* Wiley.

[3] B. Schweizer, E. F. Wolff (1981). "'On Nonparametric Measures of Dependence for Random Variables'", *Ann. Statist. Vol. 9, No. 4, 879-885.*

[4] H. Joe (1997). *Multivariate Models and Multivariate Dependence Concepts.* Chapman and Hall.

[5] S.Poon, M. Rockinger, J. Tawn (2004). "'Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications'", *The Review of Financial Studies, Vol. 17, No. 2, 581-610.*

[6] C. Genest, L-P. Rivest (1993). "' Statistical Inference Procedures for Bivariate Archimedean Copulas'", *Journal of the American Statistical Association*, Vol. 88, Issue 423, 1034-1043.

[7] J-D. Fermanian (2004). "'Goodness-of-fit tests for copulas'", *Journal of Multivariate Analysis 95*, 119-152.

[8] C. Genest, B. Rmillard, D. Beaudoin (2007). "'Goodness-of-fit tests for copulas: A review and a power study'", *Insurance: Mathematics and Economics 44*, 199213.

[9] C. Genest, B. Rmillard (). "'Validity of the Parametric Bootstrap for Goodness-of-fit testing in Semiparametric Models'", *Ann. de lInstitut Henri Poincar - Probabilits et Statistiques*, Vol. 44, No. 6, 10961127.