Thesis for the degree of Licentiate of Engineering in Mathematical Statistics

Modeling Dependence Structures in the Electricity Price Business by Means of Copula Techniques

Hui Zheng

Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden, 2010

Modeling Dependence Structures in the Electricity Price Business by Means of Copula Techniques

Hui Zheng

Department of Mathematical Science
Chalmers University of Technology
SE-41296 Gothenburg, Sweden
Phone:+46 76 583 8765

ISSN
Master Thesis Report 2010

Printed in Sweden, 2010

Abstract

The electricity market has reconstructed a new type of financial and economic model. With the rapid growth of the financial derivatives and the security concern of the deregulated electricity market, the modeling for electricity futures price has become an important topic, especially for the practitioners. The traditional method is to calculate the linear correlation between two variables or to measure whether the two derivatives follow the same trend of movement under the assumption of joint elliptically distribution. However, the above method cannot fully describe all the dependencies between two random variables, such as lower tail, upper tail and center dependency. Therefore, we introduce the copula analysis method, which includes the description of these dependencies. The objective of this thesis is to apply the copula technique to analyze the various relevant future prices in the electricity market, then to compile the real data to investigate the intrinsic link and relevant structure between them, and finally to find the best model to simulate the real world data. The theory of this thesis is based on the Archimedean copula and elliptic copula. Firstly, a most proximate copula model, either single or mixed, is estimated based on the real data in 2D. Then, based on the same theory, the application in 2D is extended to 3D, and the 3D copula is generated by applying the similar method.

KEYWORDS : Copula, Dependence , Correlation, Electricity

Acknowledgements

I sincerely appreciate Prof. Patrik Albin for his kind supervising and valuable suggestions. Meanwhile, I am very grateful of Mr. Alain Angeralides for helping me collecting the data.

Contents

Abstract **i**

Acknowledgements **ii**

1 Introduction **3**

2 Basic Copula Functions **5**

2.1 Definition and Properties 5

2.2 Frechet Bounds 7

2.3 The Dependence Structure 7

2.4 The Linear Correlation 10

2.5 The Probability Density Function of Copulas 12

2.6 Derivation of copulas : Survival Copulas 12

2.7 Non-parameter copula :Empirical Copulas 15

2.8 Mixture Copulas 15

3 Copulas Family Examples **16**

3.1 Archimedean Copulas 17

3.2 Bivariate Copulas examples 17

3.3 Extreme Value Copulas 24

3.4 Fréchet − Hoeffding Family 26

4 Sampling Random Number for Copulas Models **27**

4.1 The Generation Method 27

4.2 The Generation Random Numbers of some examples 29

4.3 Robust Estimation of Significant 30

5 Data exploration and Investigation **31**
5.1 Data exploration 31
5.2 Find fitted bivariate copula for data sets 36
5.3 Fitting multi-dimension copula for data 38
5.3.1 pair-copula for high dimension data sets 38
5.3.2 Fit copulas for three dimension data sets 40

6 Mixture Copulas for bivariate copulas and multivariate copulas **42**
6.1 Mixture copulas for bivariate copulas 42
6.2 Mixture copulas for multivariate copulas 45

7 Conclusion **46**

Appendices **47**
.1 48
.2 48

Bibliography **48**

Chapter 1

Introduction

The deregulation of the electricity market aims to provide the consumers freedom to select their preferred electricity distributor, in order to create a more effective electricity pricing system. Under such situation, the consumption price of electricity and other types of resources are considered. The electricity traders particularly concern the electricity price at a certain timeframe T in the future, so that they can make the right transactions in advance. The dependences in the financial market have been widely studied in the past few decades. The usual method is to calculate the moving average to observe whether two derivatives follow the same trend of movement, so that the dependency between the variables could be concluded. However, sometimes these simple linear correlations are not enough to describe the entire dependency structure, such as low tail, center and up tail. Therefore, we apply a new method - using the copula model to analyze the dependency between the electricity price and other related resources, such as coal, co_2, gas and oil. I will analyze these resources in detail in my thesis. The study and application of copulas in the financial market is a very modern phenomenon. The copula model is also a relatively convenient tool in the study of the dependence structure. In statistics, copula is a function that connects the marginal distributions with the joint distribution, and the different copula functions represent different dependence structures between the variables. While simulating the copula model, the first task is to choose an appropriate copula function as well as the corresponding estimation process. Marginal distributions are considered as uniform distributions. The marginal distributions of derivative in the single electricity market may be very complicated, and they may not be easily simulated by the existing parametric models. Therefore, we should investigate a large number of single copula models, and validate their goodness with Kolmogorov-Smirnov distance, in order to find the model most proximate to the real data.

In this thesis, the dependence structure of several covariant is estimated by using the mixture copula approach in 3D. This mixed copula method can help the functions separate the concept of relative dependency degree and relative dependence structure. These concepts are constituted by two different groups of parameters - association parameters and weight parameters.

The data used in this thesis is provided by Bixia Energy Management AB. Our key concern is to find an appropriate copula model to simulate the complicated dependence structure between the electricity futures price and other fuel resource futures prices.

This thesis is organized in the following structure: Chapter 2 introduces some basic concepts and properties of copulas, the definition of mixture copula, and the comparison with linear correlation. Chapter 3 reviews the important examples of some basic copula families in 2D. Chapter 4 describes how to generate random variables by using known copula models. Chapter 5 is the statistical investigation for the real world data. Chapter 6 shows how to find a fit mixture copula model, both in 2D and 3D, which describes real data well enough. The last Chapter is the conclusion of the whole thesis.

Chapter 2

Basic Copula Functions

This chapter introduces the notion of a copula function and its probabilistic interpretation, which allows us to consider it as “dependence function” (Deheuvels, 1978). It also examines the survival copula and density notions.

Lastly, the use of copulas in setting probability bounds for sums of random variables. It collects a number of financial applications, which will be further developed in the following chapters. The examples are mainly intended to make the reader aware of the usefulness of copulas in extending financial modeling, we will exercise copulas technique in two dimensions, and extend it to three dimensions. The generalization to n dimensions is straightforward.

2.1 Definition and Properties

In the literature, the idea of a copula arose in the 19th century based on the multivariate cases of non-normality. Modern theories about copulas can be dated to about forty years ago when Sklar(1959) defined copulas and shows some of their fundamental properties: By Sklar's theorem, for a copula C,Let $(x_1, ..., x_N)$ be a random vector with cumulative distribution function

$$F(x_1, x_2, ..., x_N) = C(F_1(x_1), F_2(x_2), ..., F_N(x_N)) \tag{2.1}$$

where$F(x_1, x_2, ..., x_N)$ is the joint distribution function and $F_i(x_i)$ is the marginal distribution function of x_i. Here I denotes the interval $[0, 1]$, besides I^n defines the space $[0, 1]^n$. Therefore, it is clear that a copula is a mapping from I^n to I,that is a multivariate distribution with uniform marginal on I . From (2.1), it is obvious that the marginal dependence can be separated from the dependence structure between the variates, the that it makes interpretation C as the dependence structure of the multivariate random vector x.

Definition 2.1.1. *A map C: $I^n \rightarrow I$ is called a copula if the following conditions are satisfied:*

1. For all $u = (U_1, U_2, ..., U_N) \in I^n$

$$C(u) \geqslant 0 \tag{2.2}$$

2. for every $U_n \epsilon I$

$$C(1, 1, ...U_n, ..., 1) = U_n$$

;

3. for every U_{j2} , U_{j1} with $U_{j2} - U_{j1} \geq 0$ $\forall j$

$$C(U_{12}, U_{22}, ..., U_{N2}) - \sum_{i,j...,p/(i=j=...=p)} C(U_{1i}, U_{2j}, ..., U_{Np}) + C(U_{11}, U_{21}, ..., U_{N1}) \geq 0$$

Now we will confine to the bivariate copula, and the multivariate copula is derived straightforward.

Definition 2.1.2. *A two-dimensional copula C is a real function defined on $A \times B$,where A and B are non-empty subsets of $I = [0, 1]$, containing both 0 and 1:*

$$C : A \times B \longrightarrow \mathscr{R};$$

(i) grounded $(C(v, 0) = C(0, z) = 0)$
(ii) such that $C(v, 1) = v$, $C(1, z) = z$ for every (v,z)of $A \times B$
(iii) they are 2-increasing, for every rectangle $[v_1, v_2] \times [z_1, z_2]$ where vertices lie in $A \times B$,such that $v_1 \leqslant v_2$,$z_1 \leqslant z_2$

$$C(v_2, z_2) - C(v_2, z_1) - C(v_1, z_2) + C(v_1, z_1) \geq 0;$$

A function that fulfills Property 1 is said to be *grounded*,and also property 3 is fulfilled called *2-increasing*, see $Figure 2.1$

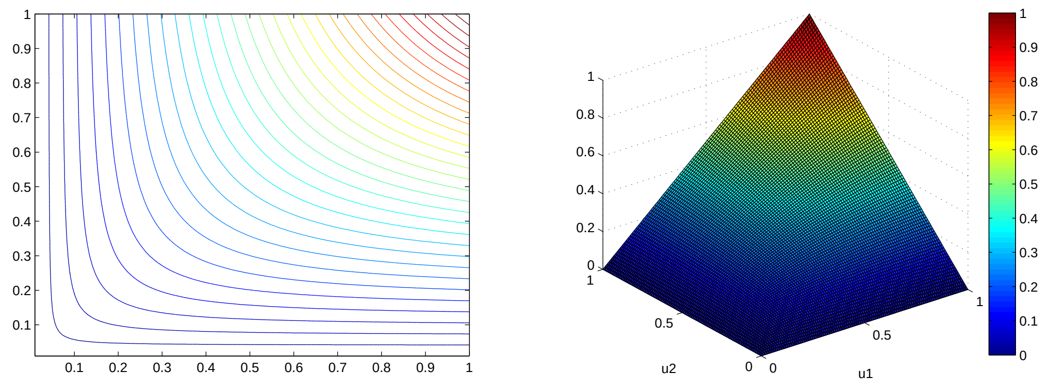

Figure 2.1: Example of a copula

2.2 Frechet Bounds

Definition 2.2.1. *The copulas $C^{-} : I^2 \longrightarrow I$ and $C^{+} : I^2 \longrightarrow I$ are given that:*
C^{-} denotes the lower bound, called the minimum copula $\longrightarrow$ $min(u, v)$; C^{+} denotes the upper bound, called the maximum copula $\longrightarrow$ $max(u + v - 1, 0)$;

For every copula C and every $(u, v) \in I^2$

$$C^{+}(u, v) \geq C(u, v) \geq C^{-}(u, v) \tag{2.3}$$

This is the Frechet-Hoeffding inequality,which refer to C^{-} as the Frechet-Hoeffding lower bound and C^{+} as the Frechet-Hoeffding upper bound, see Figure 2.2

2.3 The Dependence Structure

Now we talk about some features of ”positive” and ”negative” dependence: positive dependence tends to express that ”large value(small value)” of the random variables occur together,while negative dependence shows that ”large value” of one variable occurs with ”small value” of the other

Definition 2.3.1. *The copulas $\Pi : I^2 \longrightarrow I$ is given that:*

$$\Pi\,(u, v) = uv;$$

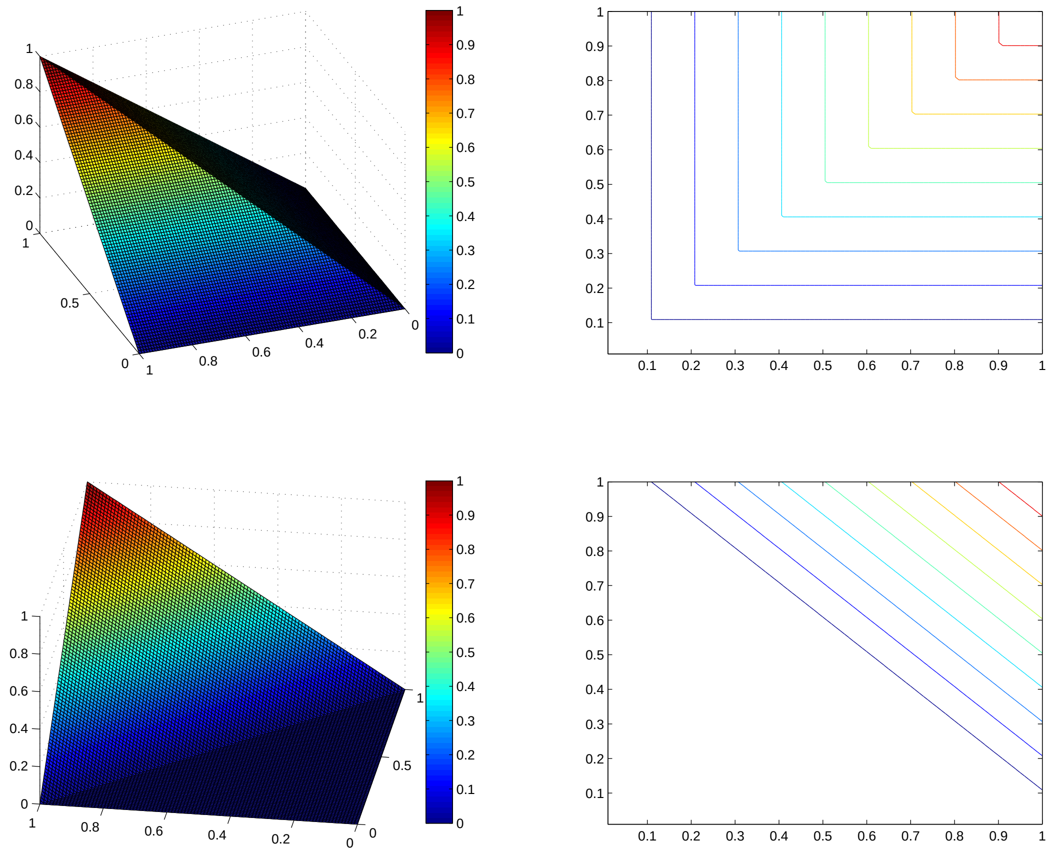

Figure 2.2: Minimum (upper figures) and maximum (lower figures)copulas

Definition 2.3.1. *Two random variables X and Y are called positively quadrant dependent if for all (x,y)*

$$P[X \leq x, Y \leq y] \geq P[X \leq x]P[Y \leq y], \tag{2.4}$$

or similarly

$$P[X > x, Y > y] \geq P[X > x]P[Y > y], \tag{2.5}$$

The notion of negative quadrant dependence is analogical reversing the inequalities in (2.4) *and* (2.5).
If X and Y have joint distribution function G, with continuous marginal distributions F_1 *and* F_2*,respectively and copula C and* (2.4) *holds so*

$$G(x, y) \geq F_1(x)F_2(y)$$

for all (x,y),then

$$C(u, c) = G(F_1{}^{-1}(u), F_2{}^{-1}(v)) \geq F_1(F_1{}^{-1}(u))F_2(F_2{}^{-1}(v))) = uv = \Pi\,(u, v)$$

for all $(u, v) = (F_1(x), F_2(y))$*, i.e.*

$$C(u, v) \geq \Pi\,(u, v) \tag{2.6}$$

These prove that Π *copula is separator of positive quadrant dependent (PQD) and negative quadrant dependent(NQD).*

Theorem 2.3.1. *Let X and Y be continuous random variables.The X and Y are independent if and only if* $C_{xy} = \Pi$

Tail dependence is an important property for cases in which this type of dependence is possible. Then, a model with tail dependence is appropriate, even though the tail dependence might be stronger than in reality. But assuming more tail dependence than necessary implies a value for riskiness which is on the safe side. So the notion of tail dependence is a method to describe the amount of extremal value dependence. One way is to measure strength of positive tail dependence, in a word, copula functions can be used to investigate tail dependence according to simultaneous booms and crashes on different markets.

Generally speaking, bivariate tail dependence looks at concordance in the tail, or extreme, values of X and Y. Geometrically, it focuses on the upper and lower quadrant tails of the joint distribution function.

Definition 2.3.1. *The computation of tail dependence is often based on the* l'Hopital rule*. *For a copula C the lower tail dependence is given by*

$$\lambda_l = \lim_{u \to 0} \frac{C(u, v)}{u}, \tag{2.7}$$

the upper tail dependence is given by

$$\lambda_u = \lim_{u \to 1} \frac{1 - 2u + C(u, v)}{1 - u}, \tag{2.8}$$

And it could be proofed that $0 \leq \lambda_u, \lambda_l \leq 1$. Whenever $\lambda_u \approx 1, \lambda_l \approx 1$, there is a strong tail dependence between variables.

$*$ *in calculus, l'Hopital's rule (also called Bernoulli's rule) uses derivatives to help evaluate limits involving indeterminate forms.*

2.4 The Linear Correlation

Linear correlation is an important statistical tool to investigate how strongly pairs of variables are related.Correlation naturally founded on the assumption of multivariate normally distributed variables, in order to describe dependencies. So, correlation analysis feature as an necessary technique to measure dependencies,take an example, returns in stock markets, in our case, it can be used to analyze dependency of electricity market variables.

The linear correlation coefficient between X and Y is

$$\rho(X;Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E[(X-E[X])(Y-E[Y])]}{\sqrt{Var(X)Var(Y)}} \tag{2.9}$$

where $Var(X)$ denotes the variance of X.
Correlation is a measure of linear dependence. So $P[Y = aX + b] = 1$,where $a \neq 0$,then $\rho(X,Y) = sign(a)$is -1 or 1.The correlation coefficient would take on any value between positive and negative one, that is

$$-1 \leq \rho \leq 1$$

.
The sign of the correlation coefficient defines the direction of the relation, either positive or negative. Specifically, a positive correlation, in the example of electricity market futures, implies that it is a larger probability for the variates to move in the same directions, which means that if one variate booms, it is a high probability that another variate may also go to climbing.

In application, if two random variables X and Y are jointly normal distributed with covariance $Cov(X,Y)$, then full of dependencies between these two variates are hold in the covariance. Which means that X is independent from $Y - Cov(X,Y)/Var(Y)X$, since

$$Cov(X, Y - \frac{Cov(X,Y)}{Var(X)}X) = Cov(X,Y) - \frac{Cov(X,Y)}{Var(X)}Cov(X,X) = 0 \tag{2.10}$$

There are two random variables $\bar{X}$ and $\bar{Y}$ have a dependence structure that can be described by correlation analysis, then $\bar{X}$ and $\bar{Y} - Cov(\bar{X},\bar{Y})/Var(\bar{Y})\bar{X}$ is independent, so in this case the Π copula should fit its empirical copula correctly.

If the copula function of two real random variables X and Y is given, their covariance is written by

$$Cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)(c(F_X(x), F_Y(y)) - 1)dxdy \tag{2.11}$$

changing $(x,y) = (F_X^{-1}(u), F_Y^{-1}(v))$ and $(dx, dy) = (du/f_X(x), dv/f_Y(y))$,

$$Cov(X,Y) = \int_0^1 \int_0^1 F_X^{-1}(u), F_Y^{-1}(v)(c(u,v) - 1)dudv \tag{2.12}$$

It is obvious that all information of copula is not included in the above covariance formula. We can conclude that copulas illustrate the complete dependence structure while covariance only is a measure of linear dependence. Therefore we can't derive c copula from given $Cov(X,Y)$ *in* (2.12).

In addition, we have $\partial^2\Pi(u,v)/\partial u\partial v = 1$, then by the (2.12) the covariance is 0, meanwhile all other relations between the two random variables involved.

2.5 The Probability Density Function of Copulas

Due to there is not too much virtual differences between all these copula functions, it is convenient to study density functions of copulas.
The density of a copula C is given by

$$c(u,v) = \frac{\partial}{\partial u \partial v} C(u,v),$$

where C is continuously differentiable for u and v.

See Figure 2.3, the two graphs of first column looks same compared with that of the second column, it is obvious that dependance structures in the probability density functions is considerably different.

2.6 Derivation of copulas : Survival Copulas

The random variables of interest represent the lifetimes of individuals or objects in some population. The probability of an individual living surviving beyond time x is given by the survival function.

For a pair(X,Y) of random variables with a joint distribution function G, the joint survival function is given by $\bar{G}(x,y) = P[X > x, Y > y]$. The margin of G are the univariate survival

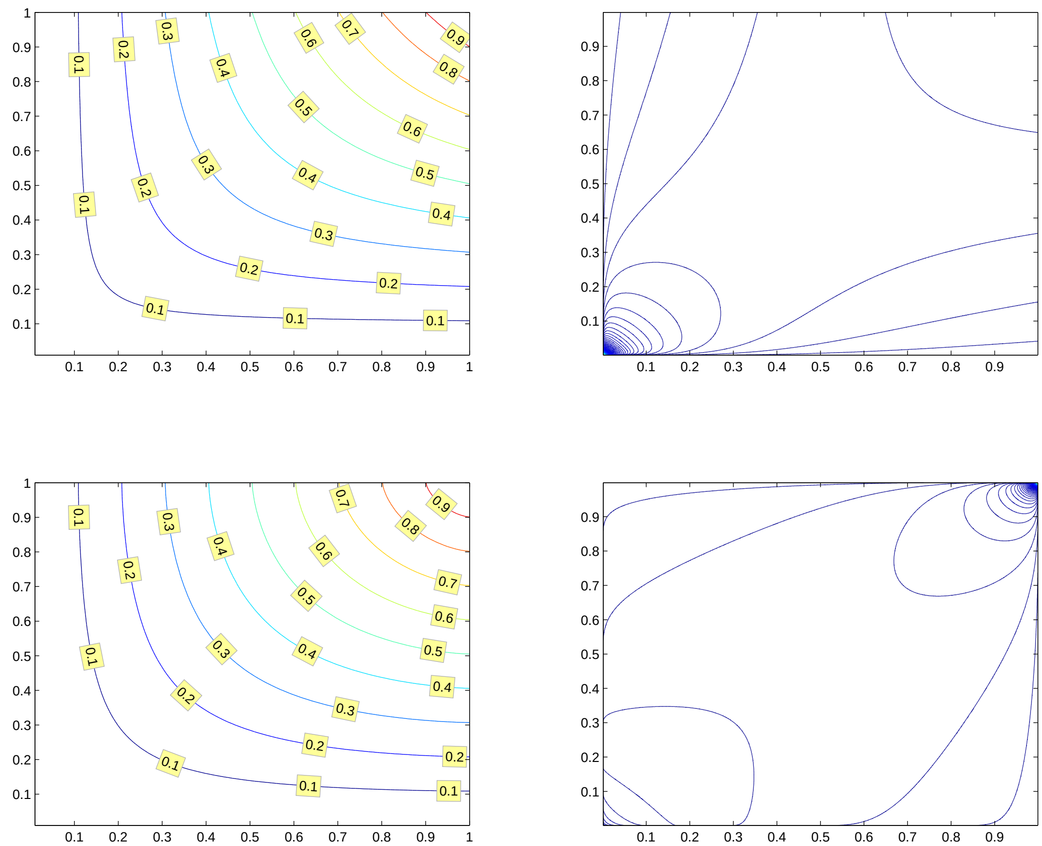

Figure 2.3: First column plots cumulate functions of these copula,while second column displays probability density functions of copula

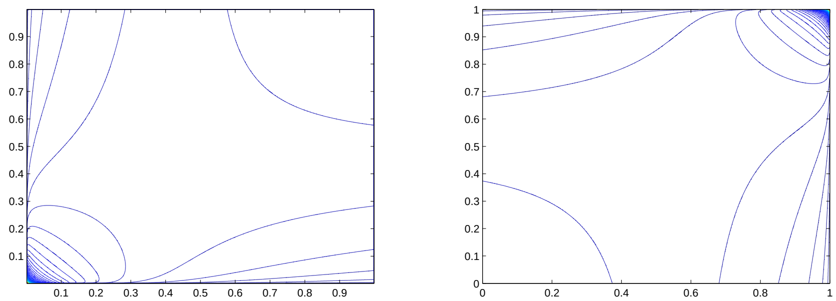

Figure 2.4: First column plots density function of copula,while second column displays its corresponding survival copula

function $\bar{F_1}$ and $\bar{F_2}$,respectively.Then we have

$$\bar{G}(x,y) = 1 - F_1(x) - F_2(y) + G(x,y) = \bar{F_1}(x) + \bar{F_2}(y) - 1 + C(1 - \bar{F_1}(x), 1 - \bar{F_2}(y))$$

so we have:

Definition 2.6.1. *A bivariate copula $C_s : I^2 \to I$ is called the survival copula of a copula C if*

$$C_s(u,v) = u + v - 1 + C(1-u, 1-v). \tag{2.13}$$

which can easily be derived that C_s is a copula if C is a copula.see Figure Also the density of the survival copula c_s and[p mlv]=cmlstat(family,x) the density of the original copula are related by

$$c_s(u,v) = c(1-u, 1-v)$$

Therefore they are mirror images about$(u,v) = (1/2, 1/2)$. Lower tail dependence in survival copula will characterize “immediate joint death”, while upper tail dependence in survival copula will characterize “long-term joint survival”.

For instance, if a copula features positive lower tail dependence means the probability of both variables being in the lower tail is relatively high; accordingly in its survival copula, its mirror image, has positive upper tail dependence and the probability of both variables being in upper tails are high.

The joint behavior of survival times can be easily modeled through copulas. It is a powerful tool to analyze the dependence structure among these data, especially because symmetric distributions are not natural candidates for these data, see Figure 2.4.

2.7 Non-parameter copula :Empirical Copulas

The empirical copula is obtained by transforming the empirical data distribution into an ”empirical copula” by warping such that the marginal distributions become uniform, in other words, it is through empirical cumulative density transform of the original data.

Definition 2.7.1. *Let* ${(x_i, y_i)^N}_{i=1}$ *represent a sample of size N from a continuous bivariate distribution. The empirical copula is given by*

$$C_e(u, v) = \frac{Number\ of\ pairs\ (x_i, y_i)\ such\ that : F_X(x_i) \leq u, F_Y(y_i) \leq v}{N}$$

and its probability density function is given by

$$c_e(u, v) = \frac{1}{N}\sum_{i=1}^{N} f(u - F_X(x_i), v - F_Y(y_i)).$$

the function f can be estimated by normal-kernel smoothing.

2.8 Mixture Copulas

Definition 2.8.1. *Let* ${C_1}^{\alpha_1}, ..., {C_N}^{\alpha_1}$ *be copulas with parameters* $\alpha_1, ..., \alpha_N$*, and* $\beta_1, ...\beta_n \geq 0$ *numbers such that* $\beta_1 + \beta_2 + ... + \beta_N = 1$ *So a mixture copula is given by*

$$C_{mix}(u, v) = \beta_1 C_1^{\alpha_1}(u, v) + ... + \beta_N C_N^{\alpha_N}(u, v). \tag{2.14}$$

Mixture models may be used in many application, and through it we could obtain more precise copula models, for example, asymmetric tail dependence.
The method to fit mixture models facilitates the separation of the concepts of *dependence degree* and *dependence structure*, and these concepts are combined with two different groups of parameters:*association parameters* α and *weight parameters* β. The association parameter in copula affects the degree of dependence,while the weight parameter in copula controls the shape of the dependence.

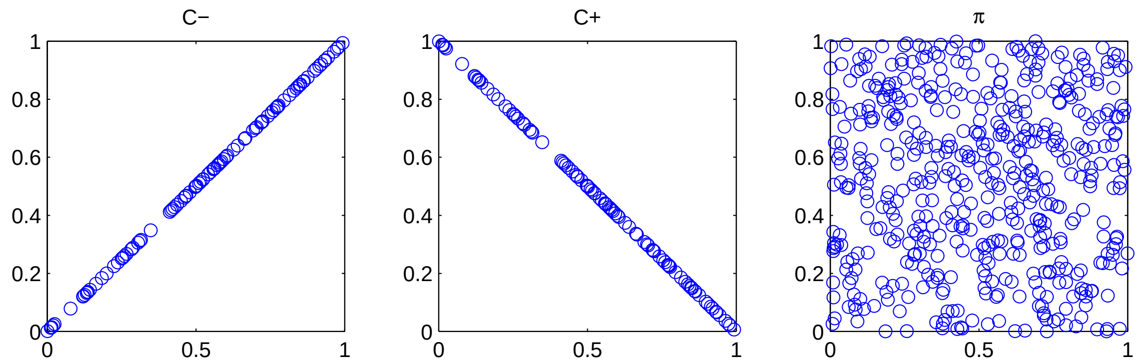

Figure 3.1: samples generated form C^{-}, Π and C^{+}

Chapter 3

Copulas Family Examples

The maximum copula C^{+} (in Section 2.3) is denoted *comonotonic* copula since it describes prefect positive dependence, while the minimum copula C^{-} (in Section 2.3) is denoted *counter-monotonic* since it describes perfect negative dependence, see Figure 3.1

The figures shows that C^{+} has maximum upper and lower tail dependence, on the contemporary C^{-} has zero upper and lower tail dependence.

In the following section, we will discuss several important classes of copulas. The tail dependence coefficients of copulas have been computed in Appendix A.

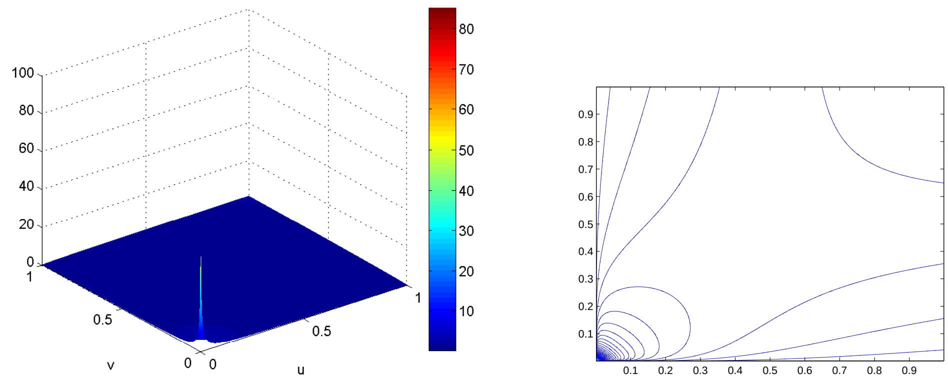

Figure 3.2: First column displays Clayton Copula density function in 3D; Second displays contour lines of Clayton Copula density function

3.1 Archimedean Copulas

The concept of archimedean copulas allows to construct a copula from a real valued function $\phi(u)$) called the generator of the copula. The copula defined via the generator ϕ is

$$C(u_1, ..., u_N) = \phi^{-1}(\phi(u_1) + ... + \phi(u_N)) \tag{3.1}$$

In order that $C(u_1, ..u_N)$ satisfies the rectangular condition we need an additional property of ϕ. For the case $d = 2$,this extra condition is convexity. And these copulas allow for a great variety of dependence structure. They have closed from expressions they are not concluded from multivariate distribution using Sklar's Theorem.

3.2 Bivariate Copulas examples

Clayton Copula For $\alpha > 0$,consider the copula,

$$C_{clayton}(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha} \tag{3.2}$$

, The Clayton copula's dependence structure is asymmetric and its tail dependence coefficients are $\lambda_U = 0$ and $\lambda_L = 2^{-1/\alpha}$, see Figure3.2. If two variates of the electricity market follow the clayton copula, they would have a larger probability of simultaneous crashing than simultaneous booming.

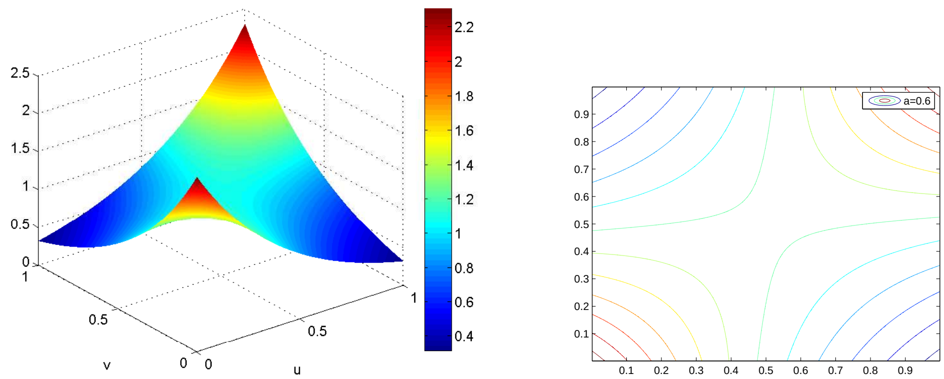

Figure 3.3: First column displays Frank Copula density function in 3D; Second column displays contour lines of Frank Copula density function

Frank Copula For $\alpha > 0$ or $0 < \alpha < 1$,consider the copula,

$$C_{frank}(u, v) = log_{\alpha}(1 + \frac{(\alpha^u - 1) + (\alpha^v - 1)}{(\alpha - 1)} \tag{3.3}$$

, The Clayton copula's dependence structure is symmetric and its tail dependence coefficients are $\lambda_U \lambda_L = 0$, see Figure3.3. If two variates of the electricity market follow the frank copula, they would have a larger probability than the Gaussian copula in the middle region,while the tails may be lighter.

Gumbel Hougaard Copula For $\alpha \geq 1$,consider the copula,

$$C_{gumbel}(u, v) = e^{-((-log(u))^{\alpha} + (-log(v))^{\alpha})^{1/\alpha}} \tag{3.4}$$

, For $\alpha = 1$, the formula (3.4)reduces to $\Pi(u, v)$,which is independent copula. The Frank copula's dependence structure is asymmetric and its tail dependence coefficients are $\lambda_U = 2 - 2^{1/\alpha}$ and $\lambda_L = 0$,see Figure3.4. Since its upper tail is heavier than the lower tail(from the expression we know that the larger is α, the heavier is the upper tail), if two variates of the electricity market follow the clayton copula, they would have a larger probability of simultaneous booming than simultaneous crashing.

Ali-Mikhail-Haq Copula It is also can be AMH copula,for the only parameter α,$0 < \alpha < 1$.

$$C_{AMH}(u, v) = \frac{uv}{1 - \alpha(1 - u)(1 - v)} \tag{3.5}$$

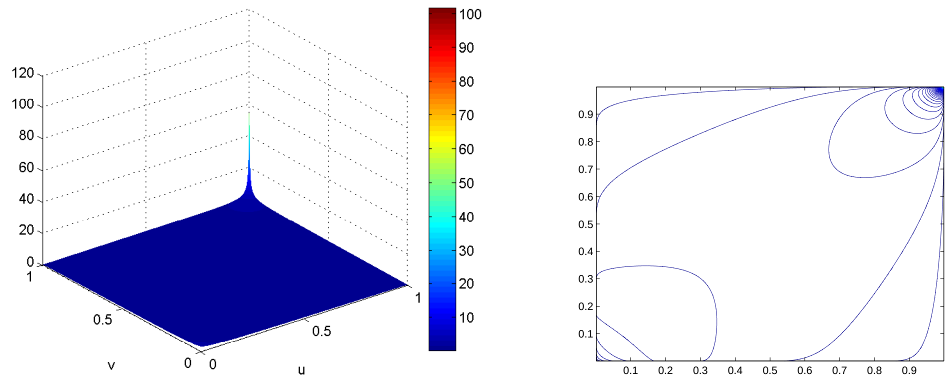

Figure 3.4: First column displays Gumbel Hougaard Copula density function in 3D; Second column displays contour lines of Gumbel Hougaard Copula density function

See figure 3.5 clearly shows that a AMH dependence structure is asymmetric and the lower tail is heavier than upper tail. Besides, its tail dependence coefficient $\lambda_U \lambda_L = 0$, so we can conclude that its upper tail is light tail and its lower tail is light tail or normal tail.

Joe Copula For $\alpha \geq 1$,consider the copula,

$$C_{Joe}(u,v) = 1-((1-u)^{\alpha}+(1-v)^{\alpha}-(1-u)^{\alpha})(1-v)^{\alpha(}1/\alpha) \tag{3.6}$$

, For $\alpha = 1$, the formula (3.4)reduces to $\Pi(u,v)$,which is independent copula. The Frank copula's dependence structure is asymmetric and its tail dependence coefficients are $\lambda_U = 2-2^{1/\alpha}$ and $\lambda_L = 0$,see Figure3.6. Since its upper tail is heavier than the lower tail(from the expression we know that the larger is α, the heavier is the upper tail), if two variates of the electricity market follow the clayton copula, they would have a larger probability of simultaneous booming than simultaneous crashing.

Gaussian Copula One of the most usually used copula,in particularly for finance modeling, is the bivariate *Gaussian copula*.Its formula is:

$$C_{Gaussian}(u,v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} exp\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\} \frac{dxdy}{2\pi\sqrt{1-\rho^2}} = \phi_\rho(\phi^{-1}(u), \phi^{-1}(v)) \tag{3.7}$$

Here ϕ^{-1} is the inverse probability distribution function of the standard normal distribution, meanwhile ϕ_ρ is joint distribution function of a standard bivariate Gaussian with the correlation coefficient ρ,which is the sole parameter of Gaussian copula ,and yield in $-1 < \rho < 1$.

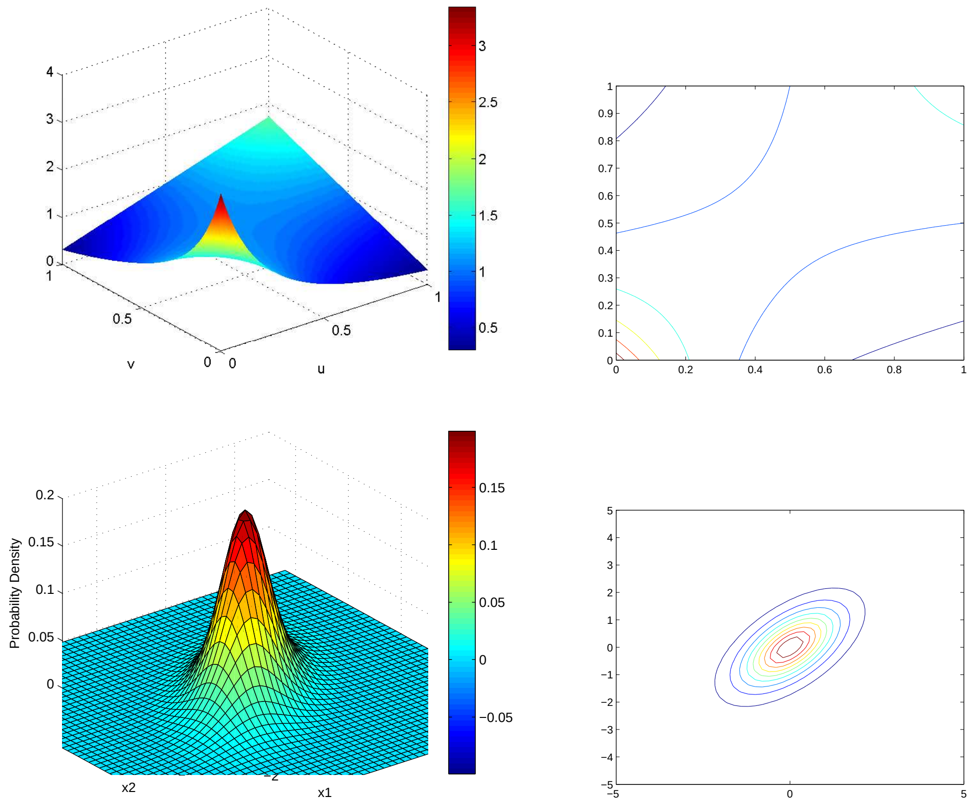

Figure 3.5: First column displays AMH density function in 3D; Second column displays contour lines of density function; first row shows copula density function , second row shows joint distribution with standard normal marginal distribution

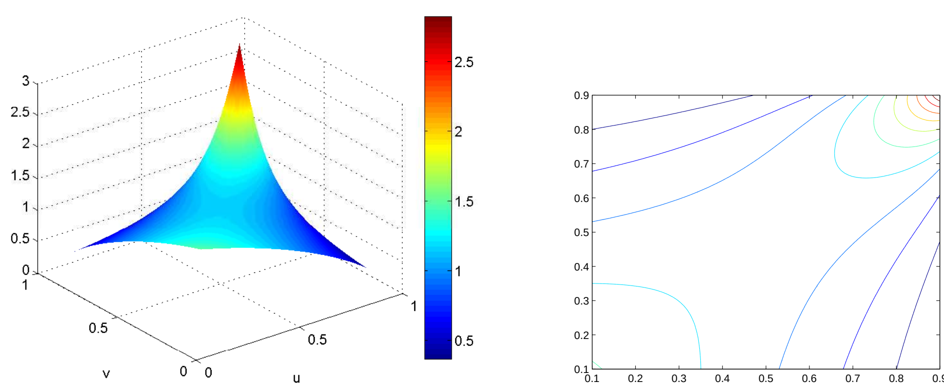

Figure 3.6: First column displays Joe Copula density function in 3D; Second column displays contour lines of Joe Copula density function

See Figure 3.7 clear displays that a Gaussian dependence structure is symmetric,If two variates of the electricity market follow the Gaussian copula, they would have the same probability of simultaneous crashing than simultaneous booming.Morever, its tail dependence coefficient $\lambda_U = \lambda_L = 0$, so we say that the tail of Gaussian copula is normal tail.

Indeed, a gaussian dependence structure with $\rho > 0$ means that the variants are positive quadrant dependent, similarly, if $\rho < 0$ the variants are negative quadrant dependent.

For example, if we take $\rho < 0$ for the electricity future price, it displays that there is larger probability for variants to move in the reserve way, which means that if one variate boom, the other variate shows higher probability to crash.

NOTE: The variates of Gaussian copula are just multivariate normal distribution if marginal are normal distributed.

t Copula The bivariate Student t' s copula, $C_t^{\rho,\upsilon}$, where υ is degrees of freedom, defined as

$$C_t(u,v) = \int_{-\infty}^{t_\upsilon^{-1}(u)} \int_{-\infty}^{t_\upsilon^{-1}(v)} (1 + \frac{x^2 - 2\rho xy + y^2}{\upsilon(1-\rho^2)}) \frac{dxdy}{2\pi\sqrt{1-\rho^2}} = t_{\rho,v}(t_\upsilon^{-1}(u), t_\upsilon^{-1}(v)) \tag{3.8}$$

Here t_υ^{-1} is the inverse probability distribution function of the student t distribution,and the correlation coefficient ρ,which is one parameters of Student t's copula ,and yield in $-1 < \rho < 1$.Student' s copula, with the corresponding picture for the Gaussian copula (Figure 3.8)It is easy to notice that Student t copula shows more observations in the tails than the Gaussian one.Morever, its tail dependence coefficient $\lambda_U = \lambda_L = 0$, so we say that the tail of Student t's copula is nearly normal tail.

HH1 Copula For the parameter $\alpha_1 > 0$,$\alpha_2 \geq 1$, HH1 is defined as:

$$C_{HH1}^{\alpha_1,\alpha_2}(u,v) = (1 + ((u^{-\alpha_1} - 1)^{\alpha_2} + (v^{-\alpha_1} - 1)^{\alpha_2})^{1/\alpha_2})^{-1/\alpha_1} \tag{3.9}$$

The HH1 copula's dependence structure is asymmetric, and it characterize both tails. The

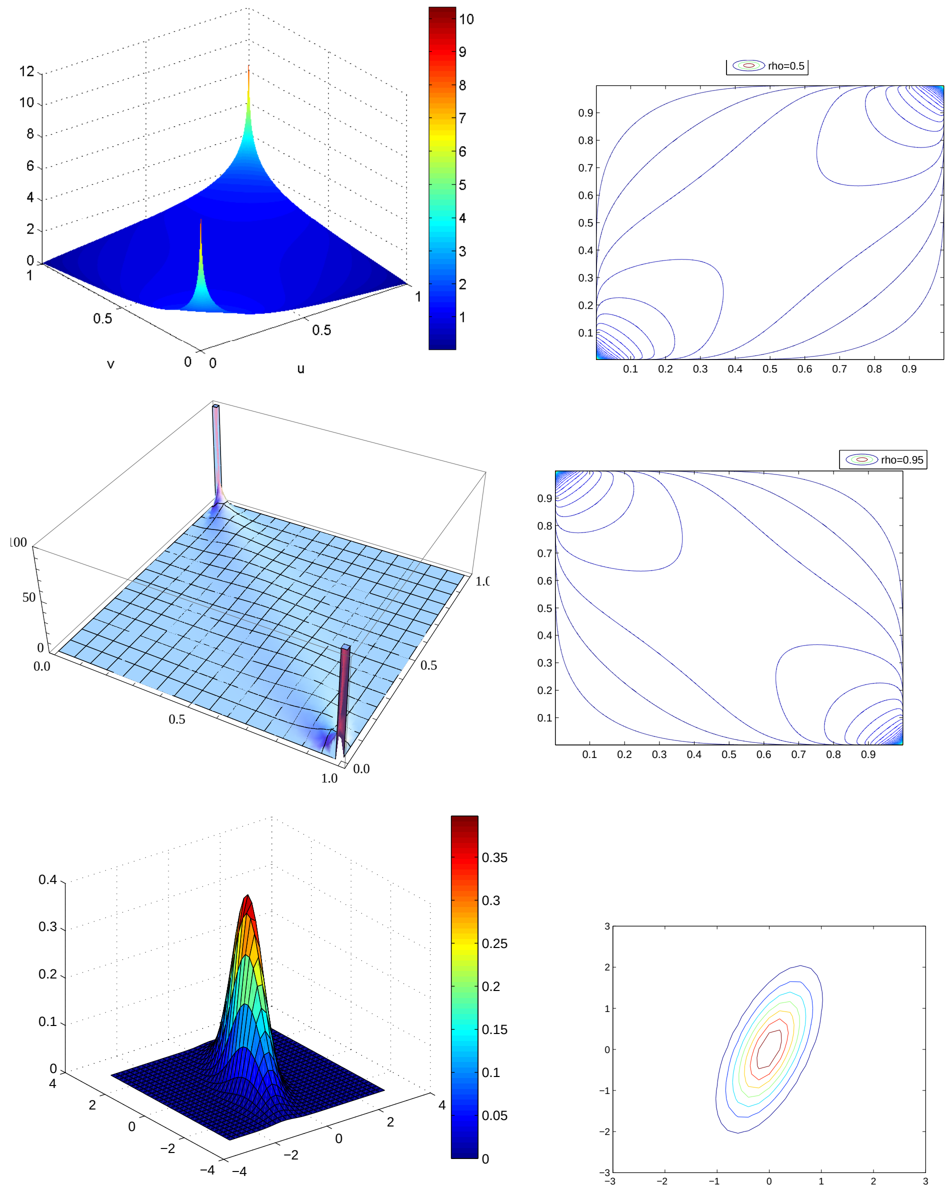

Figure 3.7: First column displays density function in 3D with ; Second column displays contour lines of density function ; first row shows Gaussian copula density function with rho=0.95, second row shows Gaussian copula density function with rho=-0.95, third row shows joint distribution with standard normal marginal distribution

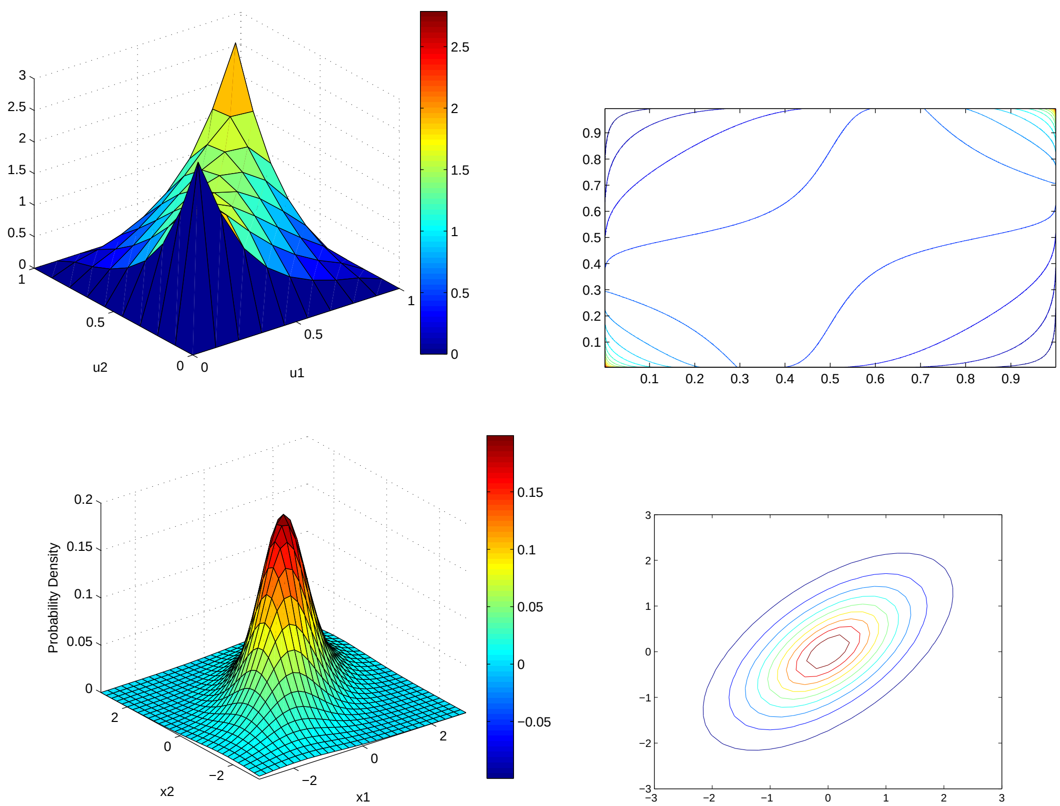

Figure 3.8: First column displays t copula density function in 3D; Second column displays contour lines of density function; first row shows t copula density function , second row shows joint distribution with standard normal marginal distribution

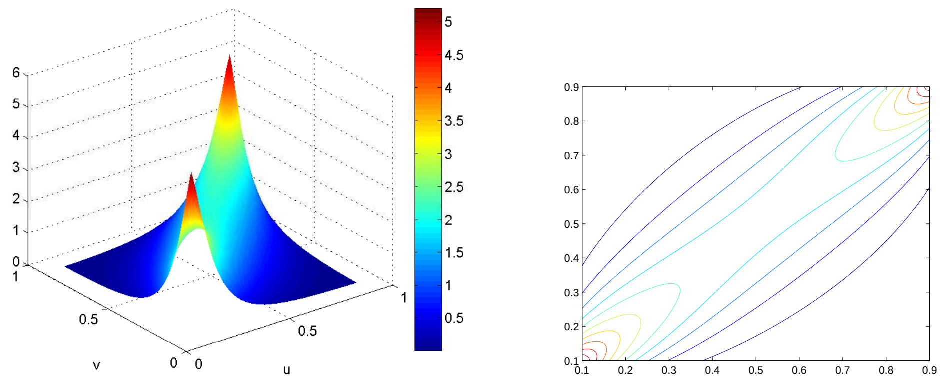

Figure 3.9: First column displays HH1 Copula density function in 3D; Second column displays contour lines of HH1 Copula density function

density function figure of copula is centralized closely to the line $u = v$, see Figure 3.9. Meanwhile, the tail dependence coefficients are $\lambda_U = 2 - 2^{1/\alpha_2}$ and $\lambda_L = 2^{1/\alpha_1\alpha_2}$.

Galambos Copula For $\alpha \geq 0$,consider the copula,

$$C_{galambos}(u, v) = uve^{-((-log(u))^{-alpha}+(-log(v))^{-alpha})^{-1/\alpha}} \tag{3.10}$$

see Figure 3.10 displays that the Galambos copula's dependence structure is asymmetric and its tail dependence coefficients are $\lambda_U = 2^{-1/\alpha}$ and $\lambda_L = 0$.

3.3 Extreme Value Copulas

A copula is named to be an extreme value copula if for all $s > 0$ the scaling property.$C(u^s, v^s) = (C(u, v))^s$ for $\forall u, v \in I$.

Extreme Value copula are max-stable, which means if $(X_1, Y_1), (X_2, Y_2), ...(X_N, Y_N)$ are independent identically distribution random pairs from an Extreme Value copula C. and $M = \{X_1, X_2, ..., X_N\}$ and $W = \{Y_1, Y_2, ..., Y_N\}$, so the copula for (M,W) is also a copula, consider the following formula:

$$C_{ExtremeValue}(u, v) = e^{log(uv)A(log(u)log(v)/log(uv))} \tag{3.11}$$

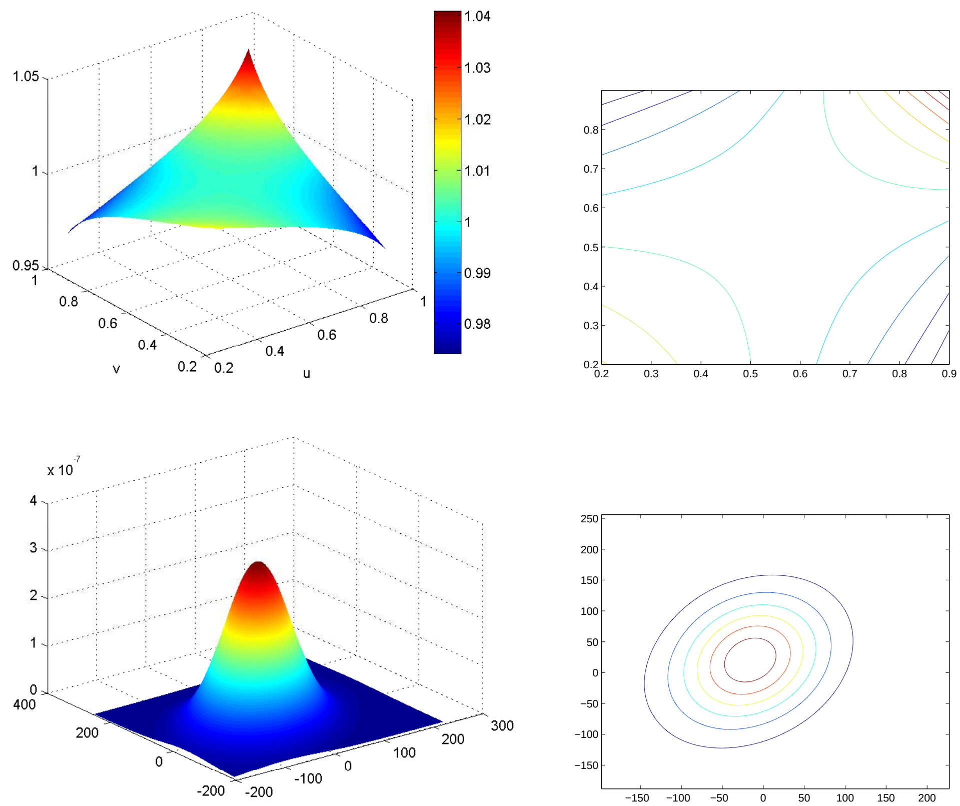

Figure 3.10: First column displays density function in 3D; Second column displays contour lines of density function; first row shows Galambos copula density function , second row shows joint distribution with standard normal marginal distribution

,

where the function A is the dependence function.

3.4 Fréchet − Hoeffding Family

Definition 3.4.1. *Let* $\alpha,\beta \in I$ *with* $\alpha + \beta \leq 1$ *and define*

$$C^{\alpha,\beta}(u, v) = \beta M(u, v) + (1 - \alpha - \beta)\Pi(u, v) + \alpha W(u, v), \tag{3.12}$$

this expression with two unknown parameter is called a Frechet Hoeffding Family.

Now set the correlation of the Frechet Hoeffding Family is

$$\rho = \alpha - \beta$$

.With the condition $\alpha = 0$, it is negative quadrant dependence, while the condition $\beta = 0$, it is positive quadrant dependence.

Chapter 4

Sampling Random Number for Copulas Models

After understand the definition and important properties of several copulas.In this chapter we will present the method how to generate random variables whose dependence structure is define by a copula.

4.1 The Generation Method

The joint density function is bounded, $f(x, y) \leq N$,then a random variate (X,Y) from the this dependence structure may be created as follow:

1.Generate two uniform distributed random variables ε, ζ from a certain box domain;

2. generate a uniform variable,ϵ from $[0, N]$;

3. if f(ε, ζ)≤ ϵ,accept (ε, ζ) to the data sets, otherwise go back to 1step.

But C(u,v)may not bounded, in this case we may change the uniform marginal distribution, for example as Gaussian copula, consider two random variables with normal marginal distributions,then

$$F(x, y) = C(\phi(x), \phi(y))$$

.

and its density function is given by :

$$f(x,y) = \frac{\partial}{\partial u \partial v} F(x,y) = \frac{\partial}{\partial u \partial v} C(\phi(x), \phi(y)) = c((\phi(x), \phi(y))\phi(x)\phi(y),$$

,

where $d\phi(x)/dx = \phi(x)$.

Simulation in the n dimension conditions is given as follow:

1.Generate n independent random variates $z = (z_1, z_2, ..., z_n)$ from N(0, 1);

2. Set $u_i = \phi(x_i)$ with i = 1, 2, . . . , n and where ϕ denotes the univariate standard normal distribution function;

3.$(u_1, ..., u_n)^* = (F_1(t_1), ..., F_n(t_n))^*$ where F_i denotes the *ith* margin.

The Student T copula is also similar to estimate.

Another method to generate random variables from a chosen copula is formulated by using a conditional approach (conditional sampling). Let us explain this concept in a easy way, under the assumption of that a bivariate copula in which all of its parameters are known. The task is to generate pairs (u, v) of observations of [0, 1] uniformly distributed r.v.s U and V whose joint distribution function is C. To reach this, we will cite the conditional distribution.

Let c_u define the conditional distribution function for the random variables V at the given value u of U,

$$c_u(V) = Pr(V \leq v \mid U = u)$$

From(2.1), and because the density function of a uniform distribution constantly equal to one, we have

$$c_u(V) = Pr[V \leq v \mid U = u] = \int_{-\infty}^{v} \frac{f(y,u)}{f_X(U)} dx = \int_{v}^{0} c(u,v) dy = \frac{\partial}{\partial x} C(x,y) \mid_{x=u} = C_u(u,v) \quad (4.1)$$

where $C_u(u,v)$ is the partial derivative of the copula C. We know that $c_u(V)$ is a non-decreasing function and exists for all $v \in I$.

With the result(4.1), we have the following conditional distribution method to generate the data, as follows:

1. Generate a uniformly distribution random variables ($\eta \in I$);

2. Generate $(V \mid U = \eta)$ from the conditional copula;

3. Now $(\eta, (U \mid V = \eta))$ will be a random variate with the expected distribution.

From above bivariate generation processing, the general procedure in a multivariate setting is as follows:

1. Define $C_i = C(F1, F2, ..., F_i, 1, 1, ..., 1)$ for $i = 2, 3, ..., n$.

2. Generate F_1 from the uniform distribution U(0, 1).

3. Next, generate F_2 from $C_2(F_2 \mid F_1)$.

4. More generally, generate F_n from $C_n(F_n|F_1, ..., F_{n-1})$.

4.2 The Generation Random Numbers of some examples

Example A Also for the sake of intuition, we assume the parameter ρ of Gaussian is 0.5, we used the general method to generate random numbers u_1, u_2, u_3, N=300, see Figure 4.1 on the left

Example B For the sake of intuition, we assume the parameter α of Clayton is 2.88, we used the above conditional copula method to generate paired random numbers N=300, see Figure 4.1 on the right

In Figure 4.2 , the generated paired random numbers are well fit the Clayton copula model by comparing to its 3-D contour lines.

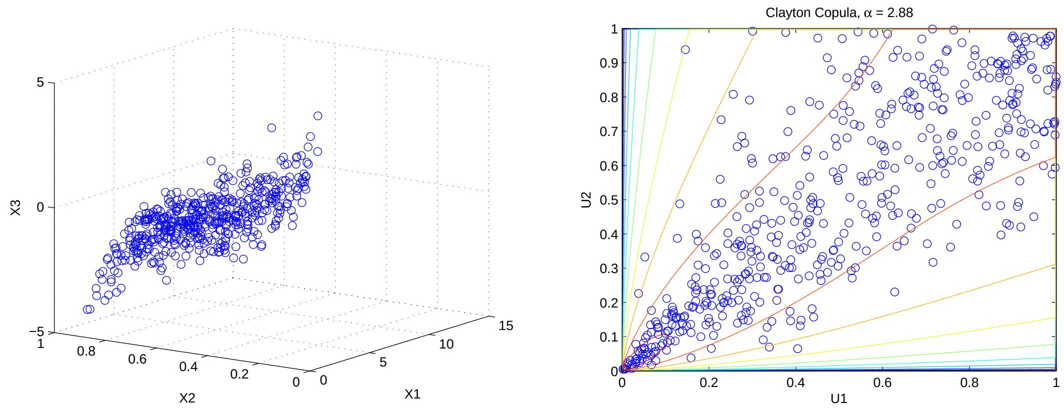

Figure 4.1: First column displays generated Gaussian Copula in 3D; Second displays generation of random variables from clayton copula

4.3 Robust Estimation of Significant

The bootstrap method would be used to test the goodness of fit the models.
The details is as follow:

1. Generate pairs (η, ξ) from model of data sets sample size;

2. find empirical copula function from the generated data;

3. calculate the maximum distance(KS) of the empirical copula to the model copula (See Appendix);

4. repeat the above procedure 1000 times;

5. plot the empirical distribution function of the observed maximum distances.

Finally, using the empirical distribution function of maximum distance, estimate p-values for the maximum distance statistics may be found.

Chapter 5

Data exploration and Investigation

Data sets process electricity futures and related fuel future returns. And all these electricity futures are affected by these fuel prices, when modelling the price process an asset on a true market, normally the logreturns of the asset are considered as i.i.d. According to the distribution of electricity future price, the investment policy of electricity price is reflected in the structure of its future price. So in order to avoid crashes, the futures should display negative dependence, on the contrary, to reach maximum profit, the futures should display positive dependence.

5.1 Data exploration

To investigate the different dependence structures of electricity futures of years 2007-2010 on the real-world markets,long term data of electricity future price is provided by **SKM Market Predictor** and help from Alain Angeralides as an electricity price trader at Bixia Energy Management AB.

Since the length of very future price is different, moreover, pair of them must be picked on a same day,I divide them into two group , group one is electricity future price in 2011(ENYOR) , Coal future price and Co2 emission price; the other is the same electricity future price in 2011(ENYOR) , Gas future price and Oil future price. Which is also for the sake of fitting three dimension copulas for the data sets.

First I fit a bivariate copula for these two group data, that means there are four pairs variates; then I use the pair-copula method to fit a three dimension copula.

Bachelier-Samuleson Black-Scholes Model

In many model of stock return in the financial market, Bachelier-Samuleson Black-Scholes

Model is most used, which gives the stock value at time t as the solution to the stochastic differential equation

$$dS(t) = (\mu + \frac{\sigma}{2})S(t)dt + \sigma S(t)dB_t, \tag{5.1}$$

where B_t is brownian motion.

The solution to (5.1) is

$$S(t) = S(0)e^{\mu t + \sigma B_t}, \tag{5.2}$$

where $S(0)$ is the return value at the starting time, μ is the drift coefficient and $\sigma^2 >$ is the volatility.

Consider the logreturn, $X(t)$, of the return value, the data set becomes driven easily by the increments of a Brownian motion:

$$X(t) = log(S(t + \Delta)) - log(S(t)) = log(e^{\mu(t+\delta)+\sigma B_{t+\delta} - \mu t - \sigma\ B_t}) = \mu\delta + \sigma(B_{t+\delta} - B_t)$$

,

where δ is time interval between sampling points. Therefor, for the Bachelier-Samuleson Black-Scholes Model the logreturn of stock values are the increments of an Brownian motion, and they stationary and independent.

Filtering of Data

By examining data sets it can be checked that the volatility is not constant, so Bachelier-Samuleson Black-Scholes Model is appropriate. To avoid this property of a financeial market we introduce a devolzatilization of data set. One may generalize this Bachelier-Samuleson Black-Scholes Model by(5.1) but substituting brownian motion by noise process a Levy process L_t,i.e.

$$dS(t) = (\mu + \frac{\sigma_t}{2})S(t)dt + \sigma_t S(t)dL_t, \tag{5.3}$$

The logreturn X_t of the electricity future price,if σ_t moves slowly compared to L_t,

$$X_t = log(S_t) - log(S_{t-\Delta}) = log(e^{\mu\zeta + \sigma_t L_t - \sigma_{t-\Delta} L_{t-\Delta}}) = \mu\Delta + \sigma_t C_t \tag{5.4}$$

,

where δ is time interval between sample point. The devolatlized logreturns A_t is a random walk independent of the time changing volatility

$$A_t = \frac{X_t - \mu\Delta}{\sigma_t} \tag{5.5}$$

By devolatilizing the changing volatility, the data is made independent of market change. see Nadaraya(1964) and Watson(1964).

Statistical Investigation of Data Sets We firstly present basic statistics of the logreturn series of electricity futures and fuel(coal, gas, oil, co_2 allowance) futures of Bixia Energy Energy Management AB.

logreturn series	mean	covariance
ENOYR.	7.5015e-05	2.3503e-04
COAL.	7.9514e-04	3.6500e-04
CO2.	5.3195e-04	6.6712e-04
GAS.	2.6422e-04	2.3856e-04
OIL.	2.6788e-04	3.3383e-04

Table 5.1: Statistics for above five logreturn series in the period from 08-Jan-2007 to 25-Oct-2010 with N=907

From Table 5.1, we calculate the first two empirical moments of the logreturns of electricity futures and other four fuel futures.

From observation of Figure 5.1 & 5.2, we see that all of these data sets have non-normal distribution and most of them display heavy tails except CO_2.

Also from Figure 5.3, it is obvious that the logreturns are time depemdent, the data is in crashed dependency structure,and also with negative dependent of variates. Which is means that

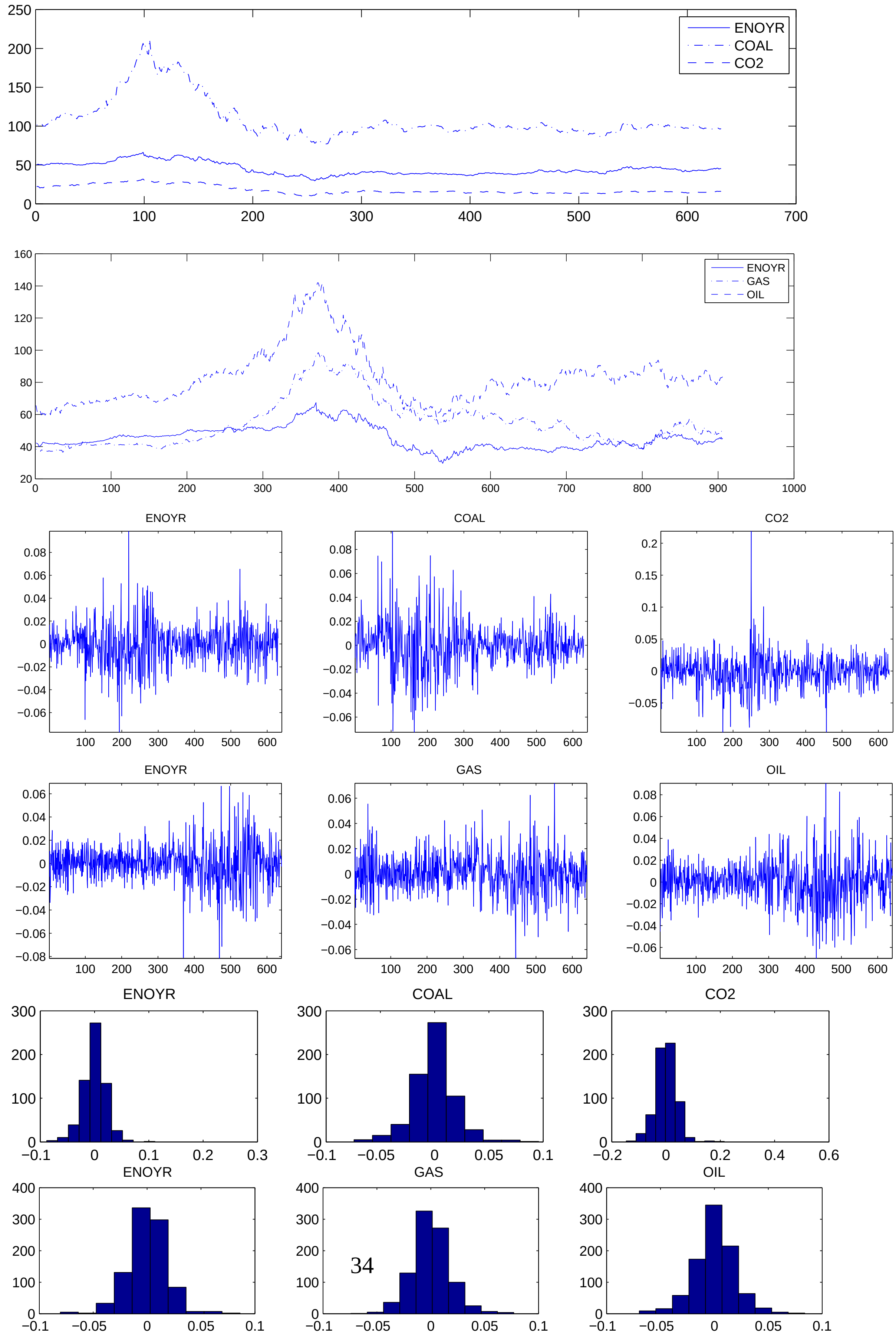

Figure 5.1: At top stock prices from 08-Jan-2007 to 25-Oct-2010 ,in the middle logreturn values in the same time period, at the bottom plots histogram of logreturns of data sets

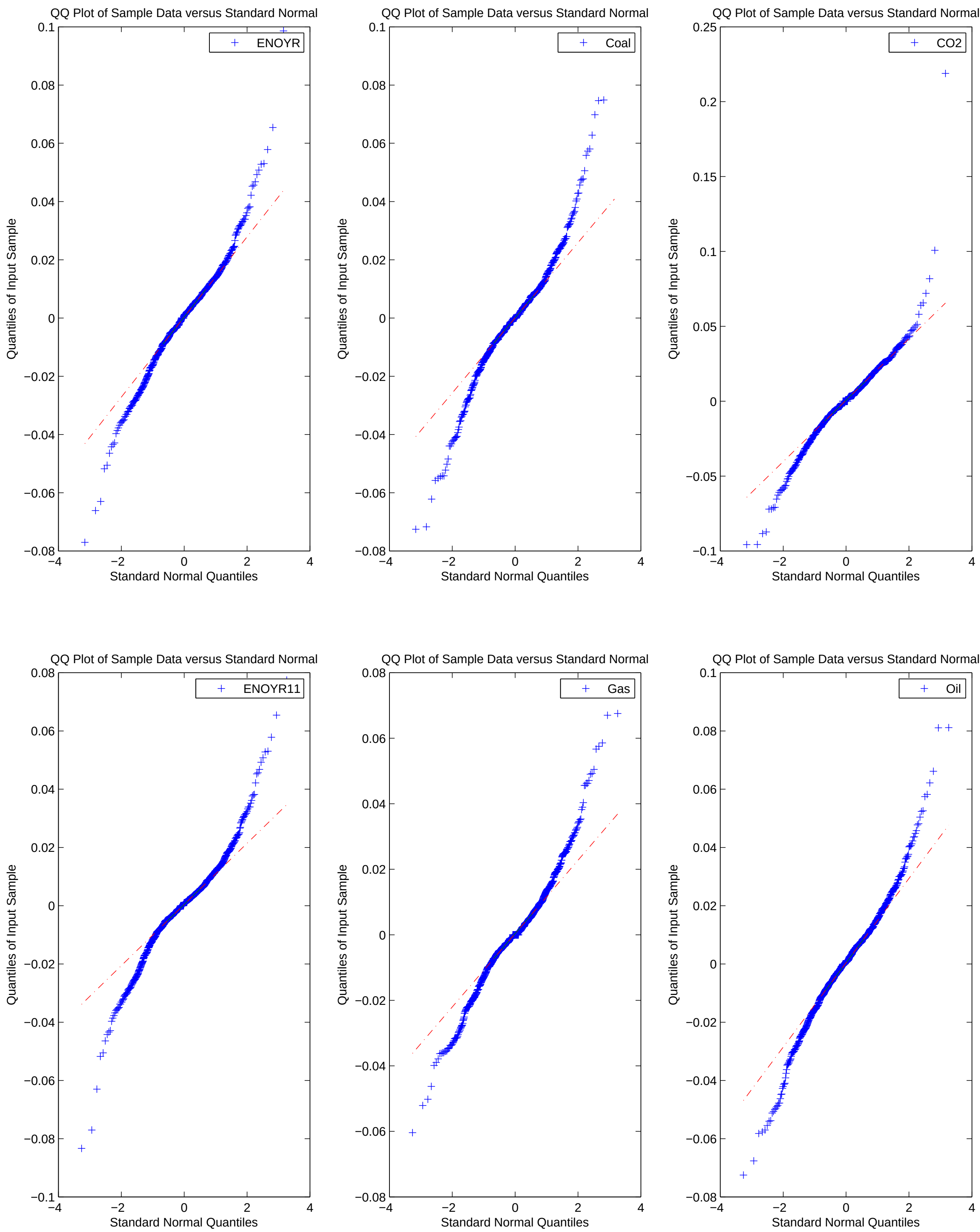

Figure 5.2: Normal quantile-quantile plot of logreturns of data sets

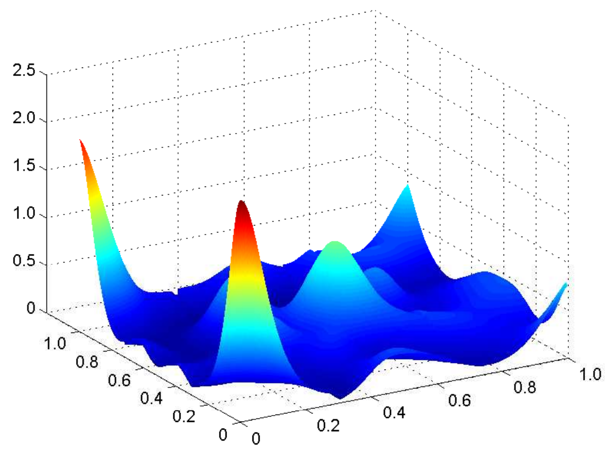

Figure 5.3: Copula density function of time dependence structure for electricity future price in 2011

data sets in not individual independent, this means that the method to use maximum likelihood to estimate the parameter of copula may not be accurate.

5.2 Find fitted bivariate copula for data sets

The numerical method is to calculate the minimum Kolmogorov-Smirnov distance between fitted copulas and empirical copula, which of the distance value is save in Table 5.2,the five devolatilized logreturn series of electricity future price of 2011 (ENYOR), Coal, CO_2(emission price), Gas and Oil, labeled as F_i for $i \in 1, 2, 3, 4, 5$. Now we look back for the dependence structure for pair $F_1\&F_2$,..., $F_1\&F_4$. Electricity and Coal ,Gas, Oil all have higher crash dependence than boom dependence, excepet Electricity and Co_2 are higher boom dependence.See the empirical density function in 3D and its cumulative function contour shown in the Figure 5.4

For the various copulas and each pair of logreturn series, by calculating the minimum KS distance between the empirical copulas and theoretical models, the copula's parameters are searched. And Figure 5.5 plots of KS distance of copulas compared with the empirical copulas show that Gumbel Survival copula is larger than other copulas, which means there is a lower tail dependence in our data set.

The p-value should be calculated to see whether Gumbel survial copula is the suitable cop-

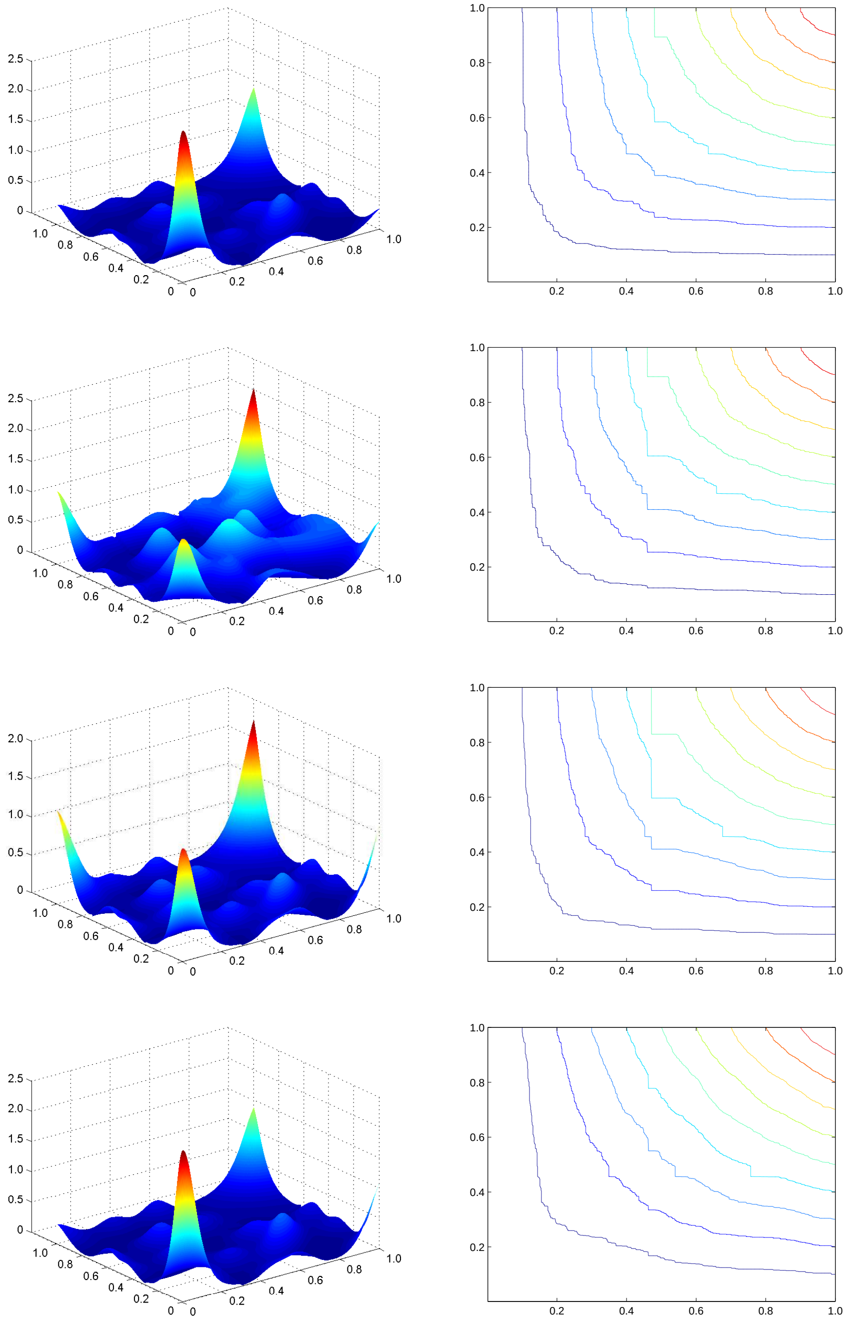

Figure 5.4: Fist row displays empirical density copula of $F_1\&F_2$(ENYOR and Coal), the left contours corresponding empirical cumulate function; Second row displays empirical density copula of $F_1\&F_3$ (ENYOR and CO2), the left contours corresponding empirical cumulate function; Third row displays empirical density copula of $F_1\&F_4$ (ENYOR and Gas), the left contours corresponding empirical cumulate function; Forth row displays empirical density copula of $F_1\&F_5$ (ENYOR and Oil), the left contours corresponding empirical cumulate function

Copula	ENYOR&COAL	ENYOR&CO_2	ENYOR&GAS	ENYOR&OIL
Gaussian	$\rho = 0.807$, $D = 0.048$	$\rho = 0.429$, $D = 0.042$	$\rho = 0.328$, $D = 0.053$	$\rho = 0.642$, $D = 0.051$
Clayton	$\alpha = 1.948$, $D = 0.057$	$\alpha = 1.545$, $D = 0.046$	$\alpha = 0.973$, $D = 0.049$	$\alpha = 0.856$, $D = 0.052$
Clayton Survival	$\alpha = 0.972$, $D = 0.061$	$\alpha = 1.457$, $D = 0.054$	$\alpha = 1.362$, $D = 0.064$	$\alpha = 1.221$, $D = 0.059$
Frank	$\alpha = 2.457$, $D = 0.051$	$\alpha = 1.753$, $D = 0.052$	$\alpha = 1.524$, $D = 0.048$	$\alpha = 2.390$, $D = 0.063$
Frank Survival	$\alpha = 2.457$, $D = 0.051$	$\alpha = 1.657$, $D = 0.051$	$\alpha = 1.524$, $D = 0.048$	$\alpha = 2.40$, $D = 0.063$
AMH	$\alpha = 0.898$, $D = 0.058$	$\alpha = 0.917$, $D = 0.049$	$\alpha = 0.872$, $D = 0.050$	$\alpha = 0.946$, $D = 0.053$
AMH Survival	$\alpha = 0.903$, $D = 0.062$	$\alpha = 0.900$, $D = 0.065$	$\alpha = 1.000$, $D = 0.060$	$\alpha = 0.916$, $D = 0.061$
Gumbel	$\alpha = 2.438$, $D = 0.057$	$\alpha = 2.712$, $D = 0.054$	$\alpha = 1.306$, $D = 0.042$	$\alpha = 1.672$, $D = 0.052$
Gumbel Survival	$\alpha = 2.645$, $D = 0.043$	$\alpha = 2.60$, $D = 0.049$	$\alpha = 1.581$, $D = 0.050$	$\alpha = 2.306$, $D = 0.048$
Joe	$\alpha = 1.257$, $D = 0.054$	$\alpha = 1.116$, $D = 0.057$	$\alpha = 1.357$, $D = 0.058$	$\alpha = 1.897$, $D = 0.055$
Joe Survival	$\alpha = 1.306$, $D = 0.049$	$\alpha = 1.004$, $D = 0.055$	$\alpha = 1.420$, $D = 0.053$	$\alpha = 1.669$, $D = 0.051$
Galambos	$\alpha = 0.916$, $D = 0.056$	$\alpha = 0.971$, $D = 0.062$	$\alpha = 0.957$, $D = 0.053$	$\alpha = 0.857$, $D = 0.058$
Galabos Survival	$\alpha = 0.930$, $D = 0.063$	$\alpha = 0.892$, $D = 0.0514$	$\alpha = 0.780$, $D = 0.065$	$\alpha = 0.910$, $D = 0.058$

Table 5.2: Minimal KS distances and copula parameter value for paired data

ula. During the calculation of p-values for these four pairs,F_1&F_2 ,..., F_1&F_4, we found that Gumbel Survival is not a good model , since all the p-value are smaller than 5%

5.3 Fitting multi-dimension copula for data

We are going to show how to do modeling for multivariate data with a series of simple constructing modules named pair-copula. This constructing method displays a new way to build complex multivariate correlation model, which is similar to classic rank model; the key here is to build certain structure models with conditional independent simple structures, which is based on the Joe, Bedford and Cooke's work.

5.3.1 pair-copula for high dimension data sets

A pair-copula model is proposed to model the high-dimensional dependency structure of financial market returns, for example: in my thesis, is electricity future price and its corresponding future price.

Considering a vector $X = (X_1, ..., X_d)$ has the follow joint density distribution function ([Bedford and Cooke, 2002])

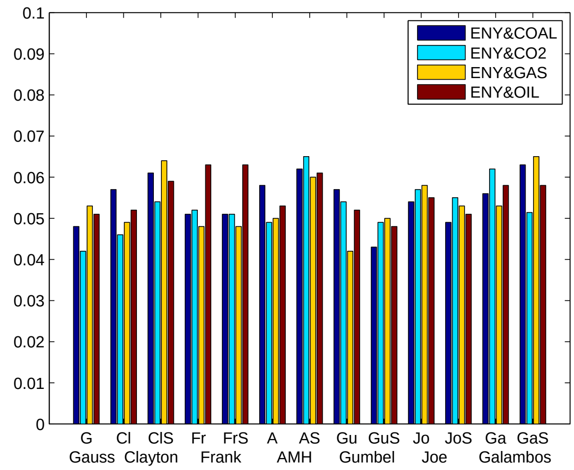

Figure 5.5: KS distance of copulas compared with the empirical copulas

$$f(x_1, ..., x_d) = f(x_d) * f(x_{d-1} \mid x_d) * f(x_{d-2} \mid x_{d-1}, x_d) * ... f(x_1 \mid x_2, ..., x_d); \tag{5.5}$$

According to [Sklar, 1959], every multivariate distributions F with marginal densities $F_1(x_1), ..., F_d(x_d)$ can be written as

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)) \tag{5.6}$$

Therefore, the copula in (5.6) can be written as

$$C(u_1, ..., u_d) = F^{-1}(F_1(u_1), ..., F_d(u_d)) \tag{5.7}$$

The general formula is given by :

$$f_{j|i}(x_j \mid x_i) = \frac{f(x_i, x_j)}{f_i(x_i)} = c_{ij}(F_i(x_i), F_j(x_j)) * f_j(x_j) \tag{5.8}$$

which is easy handy to draw to high dimension.

Now we set up 3-dimensional density function. Any such function can be written in this form:

$$f(x_1, x_2, x_3) = f_1(x_1) * f_{2|1}(x_2 \mid x_1) * f_{3|1,2}(x_3 \mid x1, x2) \tag{5.9}$$

where the factorization is unique up to a relabeling of the variables.
Note that the second term on the right hand side $f_{2|1}(x_2 \mid x_1)$ can be written in terms of a pair-copula and a marginal distribution using (5.8). Since the last term $f_{3|1,2}(x_3 \mid x_1, x_2)$ we can pick one of the conditional variables, say x2, and use a form similar to (5.9) to arrive at

$$f_{3|12}(x_3 \mid x_1, x_2) = c_{13|2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)) * f_{3|2}(x_3 \mid x_2). \tag{5.10}$$

This decomposition involves a pair-copula and the last term can then be decomposed into another pair-copula, using (5.8) again, and a marginal distribution. This yields, for a three dimensional density, the full decomposition

$$f(x_1, x_2, x_3) = f_1(x_1) * c_{12}(F_1(x_1), F_2(x_2)) * f_2(x_2) * c_{31|2}(F_{3|2}(x_3 \mid x_2), F_{1|2}(x_1 \mid x_2)) * c_{23}(F_2(x_2), F_3(x_3)) * f_3(x_3). \tag{5.11}$$

The follow step we will use above formula to fit our data sets, which is divide into two groups: **group 1** for electricity price (F_1), coal (F_2) and CO_2 (F_3); **group 2** for electricity price (F_1) , gas (F_4) and oil(F_5).

5.3.2 Fit copulas for three dimension data sets

It is similar numerical method with two dimension model, which is also to calculate the minimum Kolmogorov-Smirnov distance between fitted copulas and empirical copula, which of the distance value is plotted in Figure 5.6 , the five devolatilized logreturn series of electricity future price of 2011 (ENYOR), Coal, CO_2(emission price), Gas and Oil, divided by two groups.

For the various three dimension copulas and each triple variates of logreturn series, by calculating the minimum KS distance between the empirical copulas and theoretical models, the copula's parameters are searched. See Table 5.3

The p-value should be calculated to see whether there is a copula suitable. During the calculation of p-values for these two groups, we found that there is no such a good model , since all the p-value are far smaller than 5%
Because triple variates have more complex dependency structure, there is no single copula model is fitted, then let us find its mixture one.

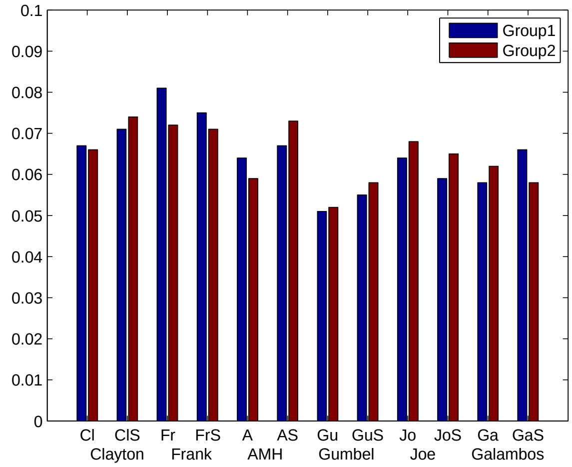

Figure 5.6: KS distance of copulas compared with the empirical copulas

Copula	Group 1	Group 2
Clayton	$\alpha = 2.971, D = 0.067$	$\alpha = 3.315, D = 0.066$
Clayton Survival	$\alpha = 1.842, D = 0.071$	$\alpha = 2.437, D = 0.074$
Frank	$\alpha = 8.421, D = 0.081$	$\alpha = 9.052, D = 0.072$
Frank Survival	$\alpha = 7.945, D = 0.075$	$\alpha = 8.859, D = 0.071$
AMH	$\alpha = 0.997, D = 0.064$	$\alpha = 0.893, D = 0.059$
AMH Survival	$\alpha = 0.972, D = 0.067$	$\alpha = 0.912, D = 0.073$
Gumbel	$\alpha = 2.654, D = 0.051$	$\alpha = 2.720, D = 0.052$
Gumbel Survival	$\alpha = 2.793, D = 0.055$	$\alpha = 2.809, D = 0.058$
Joe	$\alpha = 2.251, D = 0.064$	$\alpha = 3.105, D = 0.068$
Joe Survival	$\alpha = 2.753, D = 0.059$	$\alpha = 2.801, D = 0.065$
Galambos	$\alpha = 0.881, D = 0.058$	$\alpha = 0.901, D = 0.062$
Galabos Survival	$\alpha = 0.926, D = 0.066$	$\alpha = 0.912, D = 0.058$

Table 5.3: Minimal KS distances and copula parameter value for triple variates data

Chapter 6

Mixture Copulas for bivariate copulas and multivariate copulas

6.1 Mixture copulas for bivariate copulas

Simulation Method: First, we estimate the association parameters by finding the minimum Kolmogorov-Smirnov distance of between the empirical copula and every theoretical copula (Table 5.2). Second, we use the minimum Kolmogorov-Smirnov distance to estimate the shape parameters of the mixture copula, which is the coefficient of every single copula.

From Nelsen (1999): copula after mixing is still a copula. After simulating the copulas that estimate lower tail, upper tail and central dependence and mixing them together in mixing copula,we finally found our model copula. We mixed of three copulas, Frank copula , Gumbel Survival copula and Clayton copula. The results are shown in Table 6.1

$$C_{mix}(u,v) = \beta_1 C_{frank}(u,v,\alpha_1) + \beta_2 C_{gumbelsurvival}(u,v,\alpha_2) + \beta_3 C_{clayton}(u,v,\alpha_3) \tag{6.1}$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ are association parameters in the mixture copula which effect the degree of dependence, and $\beta = (\beta_1, \beta_2)$ are weight parameters in the mixture copula which reflect the shape dependence structures. Follows that choosing β_1,β_2,$\beta_3 \in [0,1]$, β_3 =1-($\beta_1 + \beta_2$) with $\beta_1 + \beta_2 \leq 1$. See Figure 6.1, plots some mixture copula density function and its contour with different values of vectorα and β.

Finally we need to test goodness-of- fit of the model I found, may be specified in two ways,

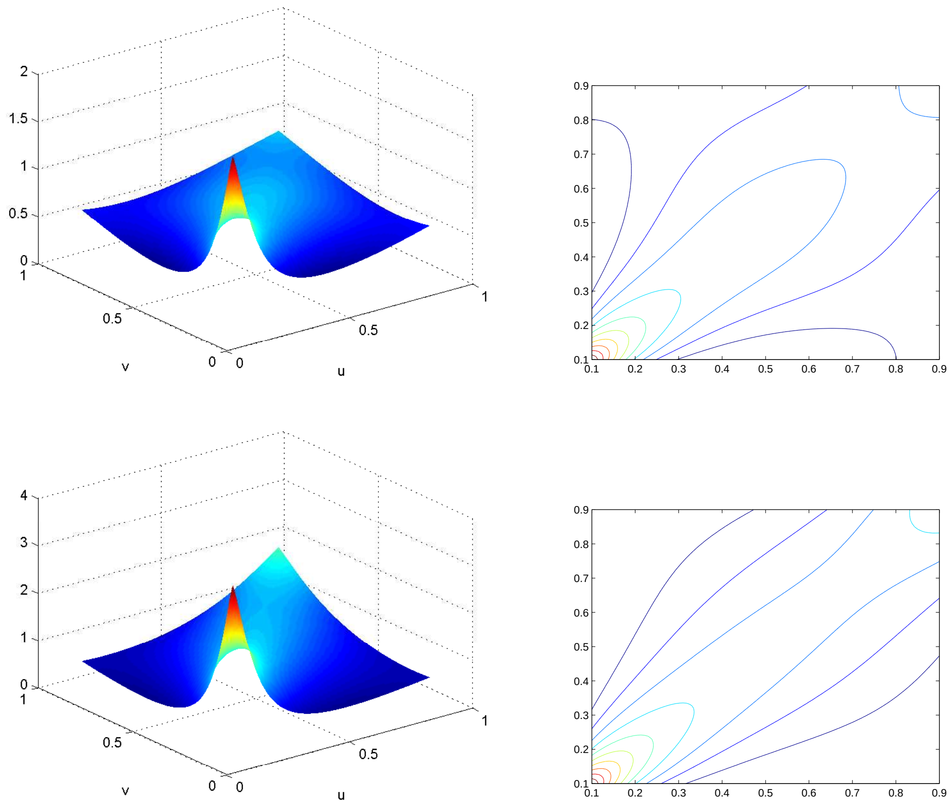

Figure 6.1: At top displays the mixture copula density function with $\beta_1 = 0$, $\beta_2 = 1/3$, $\beta_3 = 1/2$ and $\alpha_1 = 1.94$, $\alpha_2 = 8.15$, $\alpha_3 = 2.48$, at the bottom displays the mixture copula density function with $\beta_1 = 1/3$,$\beta_2 = 1/3$,$\beta_3 = 1/3$ and $\alpha_1 = 1.94$, $\alpha_2 = 8.15$, $\alpha_3 = 2.48$

	ENYOR&COAL	ENYOR&CO_2	ENYOR&GAS	ENYOR&OIL
α_1	2.457	1.753	1.524	2.390
α_2	2.645	2.60	1.581	2.306
α_3	1.947	1.545	0.973	0.856
β_1	0.013	0.526	0.273	0.115
β_2	0.399	0.415	0.439	0.356
β_3	0.588	0.069	0.302	0.529

Table 6.1: parameter value of mixture copula for different paired data

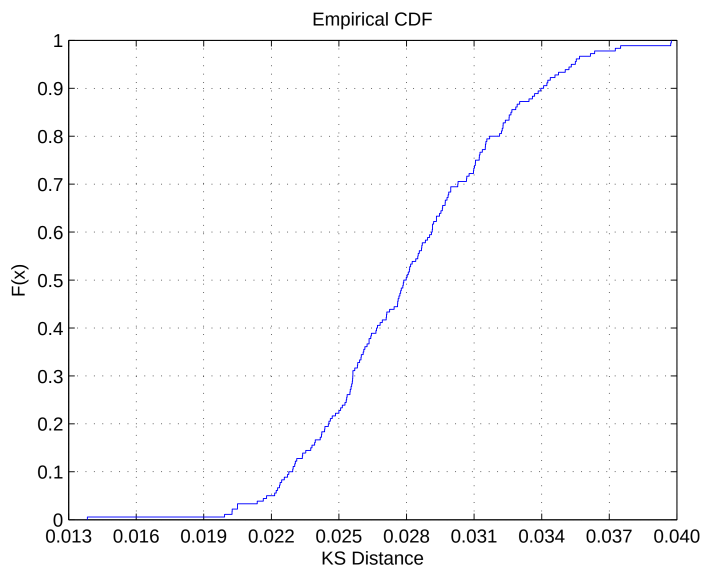

Figure 6.2: KS distance of copulas compared with the empirical copulas

one is a Kolmogorov-Smirnov test , the other is chi-square test.

For the first method: the calculated Kolmogorov-Smirnov distance (greatest distance) between fitted mixture copula and empirical distribution for every pair of data sets, that is $D_{12} = 0.043$, $D_{13} = 0.038$, $D_{14} = 0.036$. Then using the bootstrap method was used o find the approximate p-value of the generated data sets from the copula model, the results plots in Figure 6.2 and which shows that all the distance $D_{12} = 0.043$, $D_{13} = 0.038$, $D_{14} = 0.036$ are too large to reject the fitted mixture model.

For the second method is a chi-square test, the *p-value* of paired future price of ENYOR and Coal, ENYOR and Gas, ENYOR and Oil is larger than 5%, so accept the fitted model, which means my mixture model fits real data sets as well.

Compared with the single Clayton Copula, my mixture copulas is more precise. See the comparison of results in Table 6.2.

	ENYOR&COAL	ENYOR&CO_2	ENYOR&GAS	ENYOR&OIL
$C_{mixture}$	0.156	0.029	0.573	0.108
$C_{gumbelsurvial}$	8.43e-04	2.759e05	1.306e05	7.672e04

Table 6.2: comparison of p-value between mixture copula and clayton copula

6.2 Mixture copulas for multivariate copulas

Using the same method to find mixture copulas for triple variates of real world (u, v, z) finally we mixed of three copulas, Clayton copula , Jeo Survival copula and AMH copula. The results are shown in Table 6.3

$$C_{mix}(u, v, z) = \beta_1 C_{clayton}(u, v, z, \alpha_1) + \beta_2 C_{Joesurvival}(u, v, z, \alpha_2) + \beta_3 C_{AMH}(u, v, z, \alpha_3) \tag{6.2}$$

	Group 1	Group 2
α_1	2.971	3.315
α_2	2.753	2.801
α_3	0.997	0.893
β_1	0.335	0.416
β_2	0.399	0.573
β_3	0.366	0.011

Table 6.3: parameter value of mixture copula for different paired data

This time chi-square test is used, the *p-value* of triple future price of $Group1\&Group2$, are both larger than 5%, which is $p = 0.059$ and $p = 0.106$ so accept the fitted model, which means this mixture model fits real data sets as well.

Chapter 7

Conclusion

The assumption of dependence structures of variates is not always true. From the empirical copulas, we know that the dependence structure of real world data is more complex. Usually, tail dependence is stronger than in central regions of data. Moreover, unlike the multivariate Gaussian distributions, the real market has asymmetric dependence, which always cline to crash dependency. Therefore, the linear correlation is suitable for the fitting Gaussian Processes, but not suitable in the real market, as the data come from more complex distributions. In one word, correlation cannot reflect the dependence in the electricity market, and the previous chapters have shown that copula models describe the dependence much better, especially the mixture copulas.

We need to mention that under 2D conditions, we could assume that data sets come from a Gumbel Survival copula, but we cannot assume that they are jointly Gaussian distributed, otherwise, it may lead to critical mistake, since we must consider the true data distribution is heavy lower tail, that is crash dependence. Mixture copula can fit the real data quite well, since its weight parameter effect shape of dependence, besides association parameters reflect its dependence degree. For the 3D case, we can also use the copula generator to derive the copula function of triple variates, however, the pair copula method applied in this thesis is more comprehensive and easy to apply.

Appendices

.1

The Kolmogorov -Smirnov distance is the gretest distance between the empirical distribution and a theoretical distribution for the data sets,i.e. in the term of copulas:

$$D_{KS} = Max \mid C_{emp}(u, v) - C_{theory}(u, v) \mid$$

for $u, v \in [0, 1]$, where C_{emp} is the empirical copula and C_{theory} is the theoretical copula.

.2

A p-value of 0.05, for instance, means that you would have only 5% chance of eliciting the sampling being tested if the null hypothesis was actually true. A p-value is close to zero signals that the null hypothesis is false, otherwise, accept the null hypothesis,where 0.05 is a typical threshold used in the real world test.

In this thesis I calculate a p-value by the chi-square statistics test to check whether copula models fit the loagreturn data sets.

The chi-square statistic test of m boxes is given that,

$$\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is observed frequency,and E_i is expected frequency .

$$E_i = n[C(u_i, v_i) - C(u_{i-1}, v_i) - C(u_i, v_{i-1}) + C(u_{i-1}, v_{i-1})]$$

Here $u_i > u_{i-1}$, $v_i > v_{i-1}$,$u_0 = v_0 = 0$ and $u_k = v_k = 1$ must hold.

Bibliography

[1] Friedrich Schmid, *: Measuring Conditional Dependence*. http://www.case.hu-berlin.de/events/Archive/Workshops/Copulae/talks/Schmid.pdf, 2007.

[2] Etienne Cuvelier, Monique Noirhomme-Fraiture, *: Clayton copula and mixture decomposition*. www.fundp.ac.be/pdf/publications/55680.pdf , 2005.

[3] Leslie Lamport, Yuri Goegebeur*: Modelling dependence through copulas*. www.econ.kuleuven.be/insurance/pdfs/LectureNotesUbatuba.pdf, 2003.

[4] Claudia Kluppelberg,Gabriel Kuhn, *: Copula Structure Analysis for High-dimensional Data*.
www-m4.ma.tum.de/Papers/Klueppelberg/copstruc070128.pdf, 2009.

[5] Xianling Ren , Shiying Zhang, *: Copula function selection criterion based on nonparameteric kernel density estimation*.
http://www.cqvip.com/qk/96188x/201001/1001258282.html, 2010.

[6] Florence Wu, Emiliano A. Valdez, *: Simulating Exchangeable Multivariate Archimedean Copilas and its Applications*.

http://wwwdocs.fce.unsw.edu.au/actuarial/research/papers/2006/SimulatingArchimedean Copulas.pdf, 2006.

[7] Jan Lennartsson, Miu Shu *: Copula Dependence Structure on Real Stock Markets*. Master Thesis, Chalmers University of Technology,
http://www.math.chalmers.se/ palbin/janmin.pdf 2005

[8] Alain Angeralides,Ottmar Croniez *: Modelling electridity prices with Ornstein-Uhlenbeck processes*. Master Thesis, Chalmers University of Technology,
http://www.math.chalmers.se/ palbin/CorrectedThesis30Aug.pdf

[9] Joe, H. *: Families of m-variate distributions with given margins and* $m(m-1)/2$ *bivariate dependence parameters.*. Schweizer, B., and Taylor, M. D.,editors, Distributions with fixed marginals and related topics, pages 120-141, 1996.

[10] Nadaraya,E.A. *: On estimating regression*. Theory of Probability and its Applications 9 141-142. 1964

[11] Nelssen,R.B. *: A Introduction to Copulas*. Springer, New York. 1999

[12] Umberto Cherubini, *: Copula methods in finance*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2004.

[13] Aas, Czado, *: Pair-copula constructions of multiple dependence*. Sonderforschungsbereich 386, Paper 487, 2006.
http://epub.ub.uni-muenchen.de/

[14] John H., J.Einmatl, *: Estimating a Multidimensional Extreme-Value Distribution*. Eindhoven University of Technology, The Netherland, 1993
http://arno.uvt.nl/show.cgi?$fid = 14233$

[15] Watson,G.S. *: Smooth regression analysis*. Sankhya, Series A 26, 359-372.

[16] Chatfield, C. *:The Analysis of Time series an introduction*. Chapman & Ha .