CHALMERS | GÖTEBORG UNIVERSITY

MASTER THESIS

Copula Dependence Structure on Stock Market with Application to Risk

Shaoxuan Guan

Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg, Sweden 2011

Thesis for the Degree of Master of Science

Copula Dependence Structure on Stock Market with Application to Risk

Shaoxuan Guan

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE – 412 96 Göteborg, Sweden

Göteborg, June 2011

Abstract

Traditionally, dependency between stock returns has been expressed by linear correlation and been measured whether they follow the same trend of movement under the assumption of joint elliptical distribution. However, this method cannot fully describe the dependencies between two stock returns, such as lower tail, upper tail or center dependency. In this thesis, I make use of the copula analysis method to model these dependencies. One objective of this thesis is to analyze the stock returns in the real stock market, then apply the copula technique to compile the real data and investigate the intrinsic link and relevant structure between them. Further, I will study rare events under extreme value theories. Together with the use of copula technique that models default correlation, I will construct peaks-over-threshold model to analyze the dependence structure between different stock returns and simulate the real world data. Firstly, the copula model is estimated based on the data in two dimension. Then, the application is extended to three dimension, and the three-dimensional copula model is constructed by the similar method.

KEYWORDS: Copula, Dependence, Correlation, Extreme Value Theory

Acknowledgements

I would like to express my gratitude to my supervisor, Patrik Albin, not only for his inspiration, help and advice, but also for giving me the opportunity to have an in-depth look at this subject.

Meanwhile, I am very grateful of Peter Hegarty for agreeing to be my thesis work examiner.

Contents

Chapter 1

Introduction

In finance, diversification is one of the most important techniques for reducing investment risk, which means reducing risk by investing in a variety of different assets. If the stock prices do not move up or down in perfect synchrony, a diversified portfolio will have less risk than the weighted average risk of its constituent stocks, and often less risk than the least risky of its constituents. Diversification relies on the lack of a tight positive relationship among the stock returns.

Nowadays, linear correlation is the most used way to model dependence structure because of its simplicity and convenience. However, linear correlation is only adequate when the random variables are jointly elliptically distributed, since it makes no distinction between large and small realizations and it does not distinguish between positive and negative returns.

In statistics, statisticians have found another way, the copula technique, to represent dependence. It is based on the idea that a simple transformation can be made of each marginal variable in such a way that each transformed marginal variable has a uniform distribution. The dependence structure can then be expressed as a multivariate distribution on the obtained uniforms, and a copula is precisely a multivariate distribution on marginally uniform random variables. Finance mathematicians suggest that the copula tecnique might be a better way to describe the dependence structure of stock returns. Therefore, in this thesis both correlation and copula will be used to find out the relationship between two or among three different stocks.

Further, recent studies in finance have highlighted the importance of rare events in stock pricing and portfolio choice. These rare events might be in the form of large changes in investment returns, a stock market crash, major defaults, or the collapse of risky asset prices. Extreme value theory is a practical and useful tool for modeling and quantifying risk. Thus, in many cases the most efficient way of studying these rare events is through extreme value theory. Together with the use of copula functions in modeling default correlation, we can get better analysis of the dependence structure among different stocks, and thus valuate stock prices and construct desired portfolio.

All in all, this thesis is organized as follows: Chapter 2 reviews some basic concept about the copula technique and introduces some copula examples. Chapter 3 focus on correlation and the comparison between correlation and copula. Chapter 4 discusses risk issues under Black-Scholes model and peaks-over-threshold model. Chapter 5 shows how to generate random numbers from copula models and test the fitness of copula model. In Chapter 6, I will investigate the dependence structure of different stocks. The risk application of the copula technique is covered in Chapter 7. Conclusions are drawn in Chapter 8.

Chapter 2

The Copula Function

In this Chapter, first I will introduce the notion of the copula function together with its properties and probabilistic interpretation. Then I will move on to discuss some other notions relevant to Copula tecnique, such as copula density function and empirical copula. In the end, I will enumerate five Archimedean copulas that are applied in diverse areas.

2.1 Definition and Properties

In the statistics literature, the idea of the copula technique arose in the 19th century. The technique is based on the multivariate cases of non-normality. Modern Copula theories can be dated to about fifty years ago when Sklar (1959) defined copulas and showed some of their fundamental properties: By Sklars theorem, let $X = (X_1, X_2, ..., X_n)$ be a random vector with joint distribution function $F(x_1, x_2, ..., x_n)$, for a copula C,

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

where $F_i(x_i)$ is the marginal distribution function of X_i.

Here I denotes the interval $[0, 1]$, and I^n defines the space $[0, 1]^n$. It is clear that a copula is a mapping from I^n to I, i.e. a multivariate distribution with uniform marginal on I . From (2.1), it is easy to see that the marginal dependence can be separated from the dependence structure between the variates, and that it makes sense to interpret C as the dependence structure of the multivariate random vector X.

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

1. For all $u = (u_1, u_2, ..., u_n) \in I^n$

$$C(u) \geq 0;$$

2. for every $u_k \in I$

$$C(1, 1, \ldots, u_k, \ldots, 1) = u_k;$$

3. for every U_{i_2} , U_{i_1} with $U_{i_2}U_{i_1} \geq 0, \forall i$

$$C(u_{1_2}, u_{2_2}, ..., u_{n_2}) - \textstyle\sum_{i,j,...,q \backslash i=j=...=q} C(u_{1_i}, u_{2_j}, \ldots, u_{n_q}) + C(u_{1_1}, u_{2_1}, \ldots, u_{n_1}) \geq 0.$$

Confining to the bivariate copula and the three-variate copula. The multivariate copula can be derived in a straight forward way.

Definition 2.1.2. *A map C: $I \times I \to I$ is called a copula if the following conditions hold:*

1. For every $u, v \in I$

$$C(u, 0) = C(0, v) = 0;$$

2. for every $u, v \in I$

$$C(u, 1) = u \text{ and } C(1, v) = v;$$

3. for every $u_1, u_2, v_1, v_2 \in I$ with $u_1 \leq u_2$ and $v_1 \leq v_2$

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Definition 2.1.3. *A map C: $I^3 \to I$ is called a copula if the following conditions hold:*

1. For every $u, v, w \in I$

$$C(u, 0, 0) = C(0, v, 0) = C(0, 0, w) = 0;$$

2. for every $u, v, w \in I$

$$C(u, 1, 1) = u, C(1, v, 1) = v \text{ and } C(1, 1, w) = w;$$

3. for every $u_1, u_2, v_1, v_2, w_1, w_2 \in I$ with $u_1 \leq u_2$, $v_1 \leq v_2$ and $w_1 \leq w_2$

$$C(u_2, v_2, w_2) - \sum_{i,j,k \setminus i=j=k} C(u_i, v_j, w_k) + C(u_1, v_1, w_2) \geq 0.$$

A function that fulfills Property 1 is said to be grounded, and a function with the feature of property 3 is called n-increasing.

2.2 The Probability Density Function of Copulas

Since there is not much virtual differences between copula functions, it is convenient to study density functions of copulas.

The density of a copula C is given by

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v), \tag{2.2}$$

where C is continuously differentiable function of u and v.

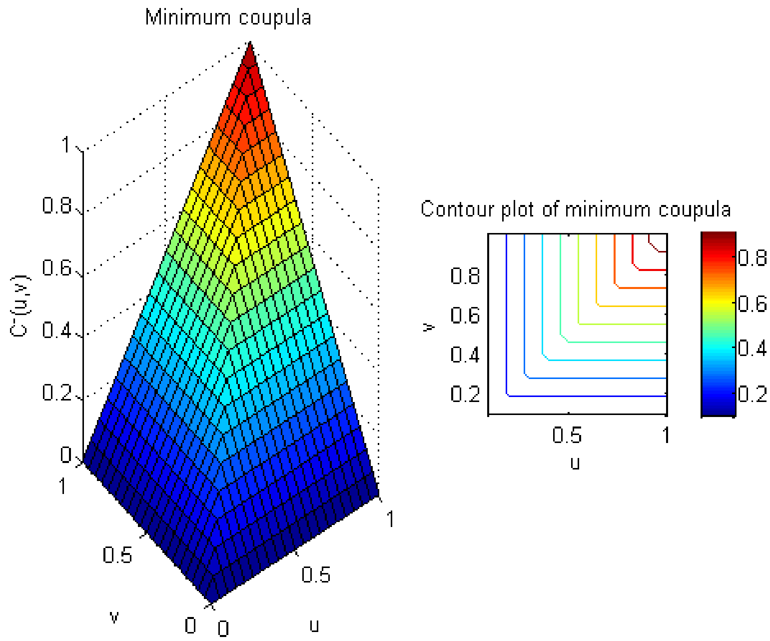

Figure 2.1: C^{-} copula

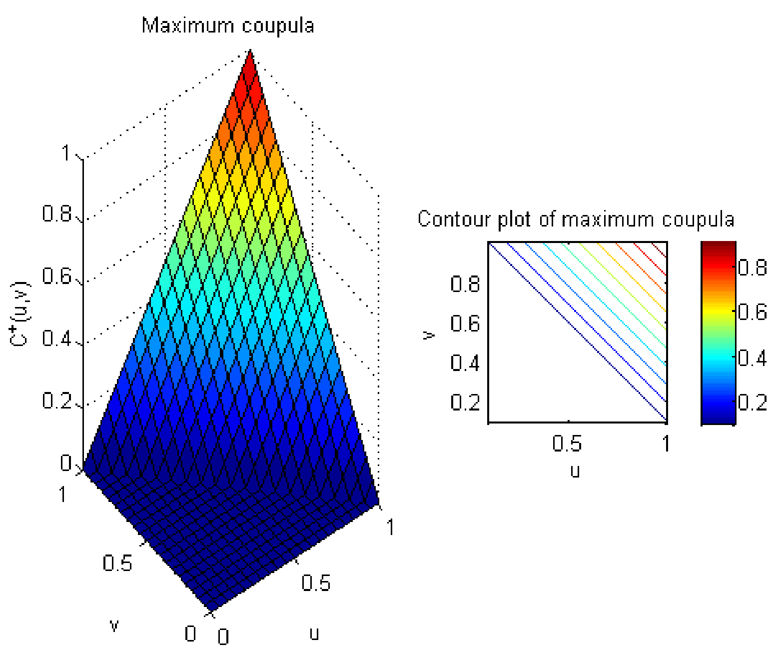

Figure 2.2: C^{+} copula

2.3 Fréchet Bounds

Definition 2.3.1. *The minimum copula $C^-: I^2 \to I$ and the maximum copula $C^+: I^2 \to I$ are given by:*

$$C^-(u,v) = \min(u,v) \text{ and } C^+(u,v) = \max(u+v-1,0).$$

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 Fréchet-Hoeffding inequality, which refer to C^- as the Fréchet-Hoeffding lower bound and C^+ as the Fréchet-Hoeffding upper bound.

2.4 Dependence Structure

PQD and NQD In copula dependence structure, positive dependence is termed to show that large value (small value) of the random variables occur together, while negative dependence tends to express that large value of one variable occurs with small value of the other. They are defined as follows:

Definition 2.4.1. *Two random variables X and Y are called positively quadrant dependent (PQD) 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

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

The notion of negative quadrant dependence(NQD) is analogical reversing the inequalities in (2.4) and (2.5).

Π **Copula** First I give the definition of Π copula.

Definition 2.4.2. *The copula Π :$I^2 \to I$ is given by*

$$\Pi(u,v) = uv. \tag{2.6}$$

If X and Y have joint distribution function F, with continuous marginal distributions F_1 and F_2 respectively, copula C and (2.4) holds for all (x,y) i.e.

$$F(x,y) \geq F_1(x)F_2(y), \tag{2.7}$$

then for all $(u, v) = (F_1(x), F_2(y))$

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)) \geq F_1(F_1^{-1}(u))F_2(F_2^{-1}(v)) = uv = \Pi(u, v). \tag{2.8}$$

In short,

$$C(u, v) \geq \Pi(u, v). \tag{2.9}$$

Therefore, if (2.9) holds for all $u, v \in I$, the copula is called a PQD copula, and an NQD copula is defined analogously. Π copula is the separator of PQD and NQD.

Theorem 2.4.1. *Let X and Y be continuous random variables. The X and Y are independent if and only if $C_{xy} = \Pi$.*

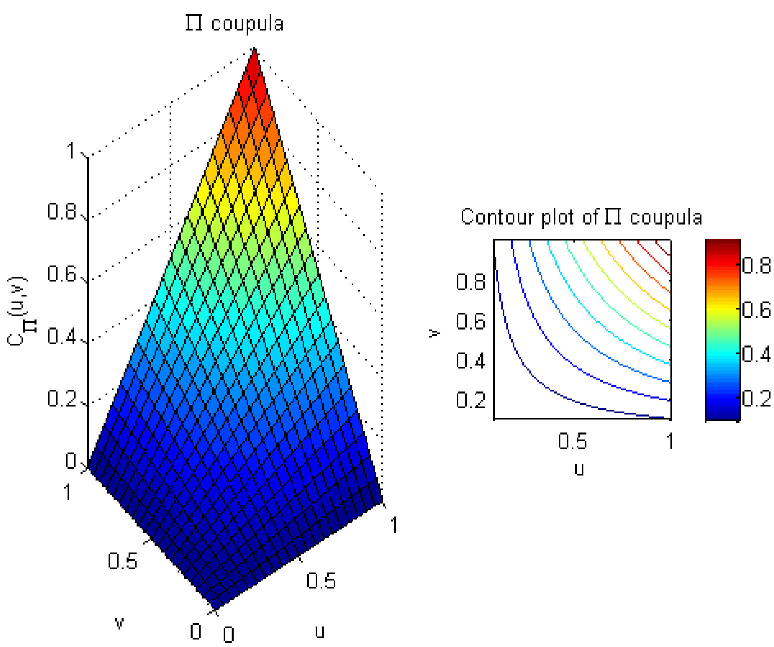

Figure 2.3: Π copula

Tail Dependence Tail dependence is a way to describe the amount of extremal value dependence. Copula functions can be used to investigate tail dependence according to simultaneous booms and crashes on different markets.

Definition 2.4.3. *For a copula C the lower tail dependence is given by*

$$\lambda_l = \lim_{u \to 0} \frac{C(u, u)}{u}, \tag{2.10}$$

and the upper tail dependence is given by

$$\lambda_u = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u}. \tag{2.11}$$

It can be proved that $0 \le \lambda_u, \lambda_l \le 1$. Whenever $\lambda_u \approx 1$ or $\lambda_l \approx 1$, there is a strong tail dependence.

The maximum copula C^+ is denoted comonotonic copula since it describes prefect positive dependence, while the minimum copula C is denoted countermonotonic since it describes perfect negative dependence. The figures shows that C^+ has maximum upper and lower tail dependence, on the contemporary C has zero upper and lower tail dependence.

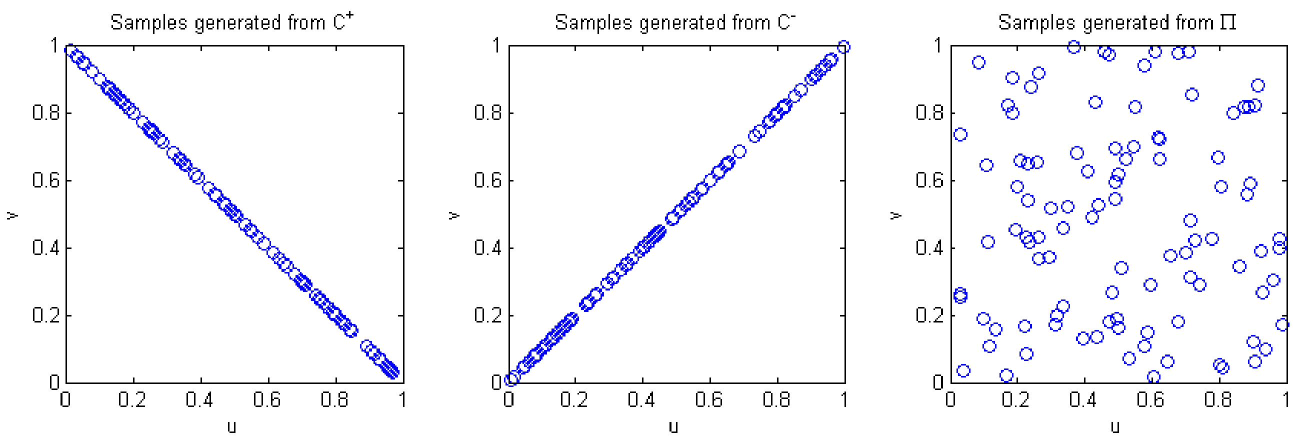

Figure 2.4: Samples generated from C^+, C^- and Π copula

2.5 Empirical Copulas

The empirical copula is obtained through empirical cumulative density transform of the original data.

Definition 2.5.1. *Let $(x_i, y_i)_{i=1}^N$ represent a sample of size N from a continuous bivariate distribution. The empirical copula is given by*

$$C_e(u, v) = \frac{\#\{(x_i, y_i) : F_X(x_i) \le u, F_Y(y_i) \le v\}}{N}, \tag{2.12}$$

and its probability density function is given by

$$c_e(u, v) = \frac{1}{N} \sum_{k=1}^{n} \delta(u - F_X(x_i), v - F_Y(y_i)). \tag{2.13}$$

the function δ can be approximated by normal-kernel smoothing.

2.6 Examples of Archimedean Copulas

Archimedean copulas is an important class of copulas. It allows to construct a copula from a real valued function φ which is called the generator of the copula. In order that $C(u_1, \ldots, u_n)$ satisfies the rectangular condition, φ need to be convex.

$$C(u_1, \ldots, u_n) = \varphi(\sum_{i=1}^{n} \varphi^{-1}(u_i)) \tag{2.14}$$

Archimedean copulas allow for a great variety of dependence structures. They have closed form expressions and are not concluded from multivariate distribution using Sklars Theorem.

I will now discuss five examples of bivariate Archimedean copulas.

Gaussian Copula The bivariate Gaussian copula is defined as

$$C^{\rho}_{Gaussian}(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left\{-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right\} \frac{dudv}{2\pi\sqrt{1-\rho^2}}, \tag{2.15}$$

where Φ is the joint distribution function of a standard bivariate Gaussian distribution, Φ^{-1} is the inverse probability distribution function of a standard bivariate Gaussian distribution, and ρ is the correlation coefficient $-1 < \rho < 1$.

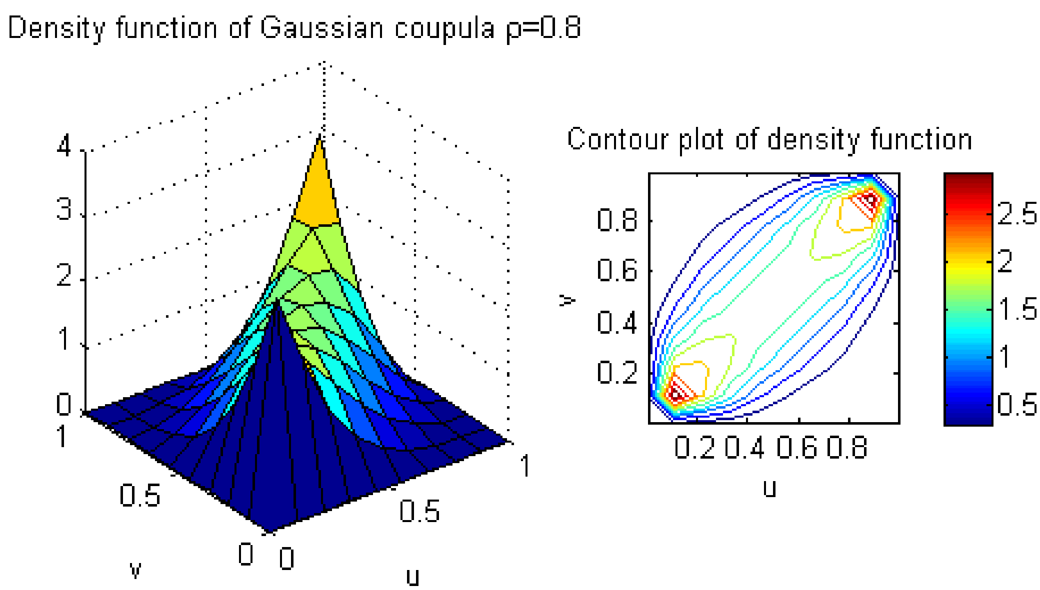

Figure 2.5: Example of Gaussian copula

Gaussian copula is one of the most usually used copula, in particularly for finance modeling. The gaussian dependence structure is symmetric, If two variates of the stock market follow the Gaussian copula, they would have the same probability of simultaneous crashing than simultaneous booming. Moreover, its tail dependence coefficient $\lambda_u = \lambda_l = 0$, so the tail of

Gaussian copula is normal tail. What is more, a gaussian dependence structure with $\rho > 0$ means that the variates are positive quadrant dependent, analogously, if $\rho < 0$ the variates are negative quadrant dependent. For example, if we take $\rho < 0$ for the two market returns, it displays that there is a larger probability for them to move in the opposite way, which means that if one variate boom, the other variate has higher probability to crash.

Student's t Copula The bivariate Student's t copula is defined as

$$C_t^{\rho,\nu}(u,v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} (1 + \frac{u^2 - 2\rho uv + v^2}{\nu(1-\rho^2)}) \frac{dudv}{2\pi\sqrt{1-\rho^2}}, \tag{2.16}$$

where $-1 < \rho < 1$. t_ν^{-1} is the inverse probability distribution function of the Student's t distribution, and ν is the degree of freedom.

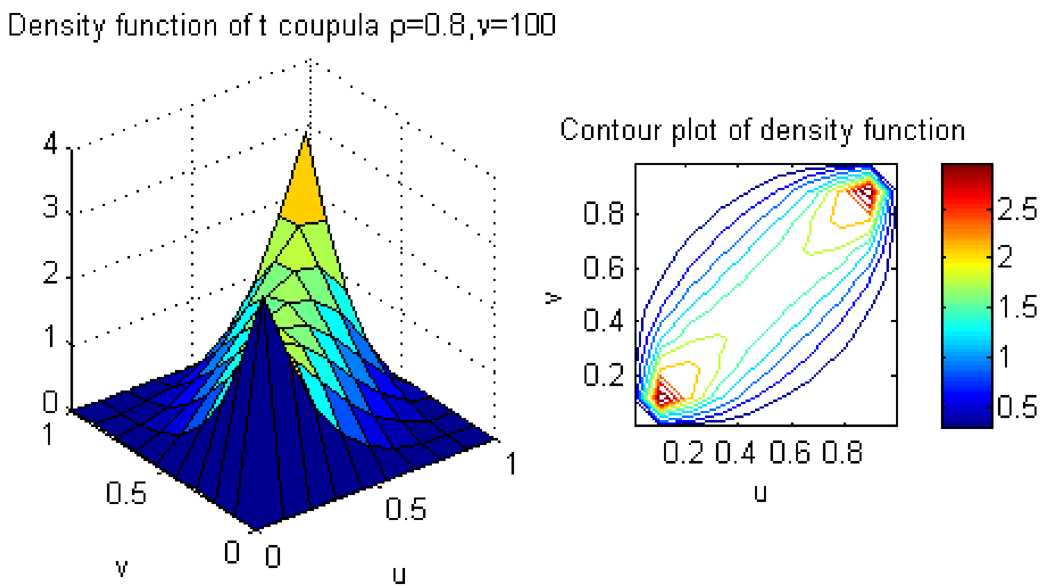

Figure 2.6: Example of Student's t copula

Student's t copula's dependence structure is symmetric. Its tail dependence coefficient $\lambda_u = \lambda_l = 0$, and it suggests that the tail of Student's t copula is nearly normal tail, but have more observations in the tails than Gaussian copula.

Frank Copula The bivariate Frank copula is defined as

$$C_{Frank}^{\alpha}(u,v) = \log_\alpha(1 + \frac{(u^{-\alpha}-1)(v^{-\alpha}-1)}{\alpha - 1}), \tag{2.17}$$

where $0 \leq \alpha \leq 1$ or $\alpha \geq 1$.

The Frank copulas dependence structure is symmetric and its tail dependence coefficients are $\lambda_u = \lambda_l = 0$. Compared with Gaussian Copula, the frank copula have a large probability in the middle region, therefore its tails must be lighter.

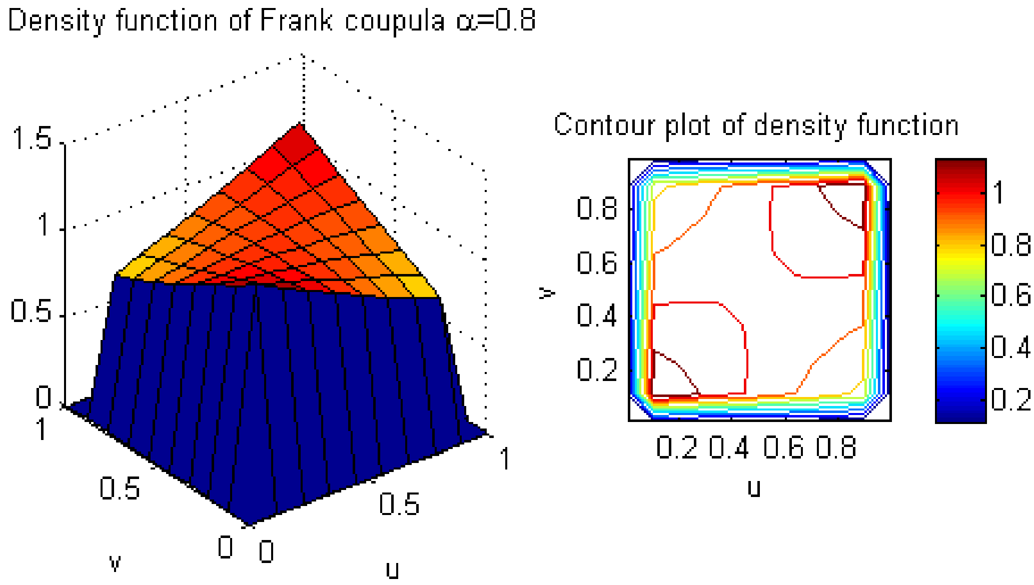

Figure 2.7: Example of Frank copula

Gumbel Copula The bivariate Gumble copula is defined as

$$C^{\alpha}_{Gumbel}(u,v) = e^{-((-\log(u))^{\alpha}+(-\log(v))^{\alpha})^{1/\alpha}}, \tag{2.18}$$

where $\alpha \geq 1$.

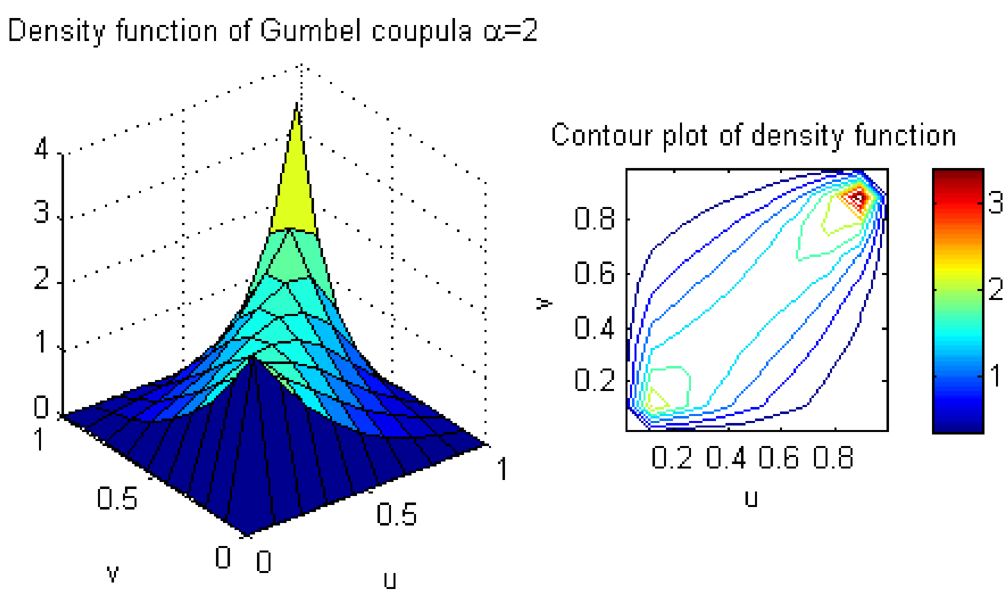

Figure 2.8: Example of Gumbel copula

The Gumble copulas dependence structure is asymmetric and its tail dependence coefficients are $\lambda_u = 2 - 2^{1/\alpha}$ and $\lambda_l = 0$. Since its upper tail is heavier than the lower tail, if two variates of the stock market follow the Gumbel copula, they would have a larger probability of simultaneous booming than simultaneous crashing.

When $\alpha = 1$, the Formula (2.18) reduces to $C^1_{Gumbel}(u, v) = \Pi(u, v)$, the independent copula.

Clayton Copula The bivariate Clayton copula is defined as

$$C^{\alpha}_{Clayton}(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}, \tag{2.19}$$

where $\alpha > 0$.

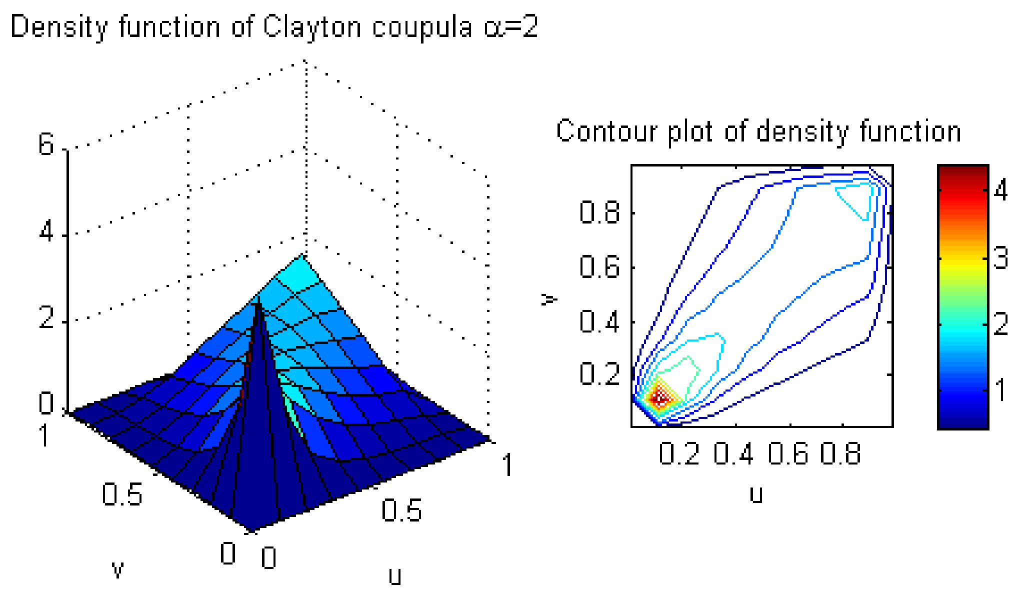

Figure 2.9: Example of Clayton copula

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

Chapter 3

Correlation

In this chapter, first I will review the linear correlation formula. Then I will briefly discuss the dependence structure of linear correlation. In the end, I will make comparison between correlation and copula.

3.1 Definition

Linear correlation is an important statistical technique to investigate how strongly pairs of variables are related. Initially, correlation is founded on the assumption of multivariate normally distributed variables, and it is aimed to describe their dependencies. Still, correlation analysis feature as an important tool to measure dependencies on for example, returns in stock markets.

The linear correlation coefficient between two random variables X and Y is

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}. \tag{3.1}$$

3.2 Dependence Structure

Correlation is a measure of linear dependence.

If $P[Y = aX + b] = 1$,where $a \neq 0$, X and Y are perfectly linearly dependent, 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 a stock market, implies that there is a larger probability for the variates to move in the same direction, which means that if one variate booms, it is much likely that another variate also goes up. While a negative correlation suggests the opposite. If one variate booms, it is much likely that another variate drops down. If the correlation between two variates equals to zero, the two variate might be independent.

3.3 Correlation and Copula

Even through the correlation is considered as one of the most widely used concepts in modern finance, it is also one of the most misunderstood concepts.

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

$$Cov(X,Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)(c(F_X(x), F_Y(y)) - 1)dxdy. \tag{3.2}$$

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

$$Cov(X,Y) = \int_0^1 \int_0^1 F_X^{-1}(u)F_Y^{-1}(v)(c(u,v) - 1)dudv. \tag{3.3}$$

It is obvious that all the information of copula is not included in the above covariance formula. We can conclude that copulas illustrate the complete dependence structure while covariance is only a measure of linear dependence. Therefore we cant derive copula from given covariance.

In addition, we have $\partial^2\Pi(u,v)/\partial u\partial v = 1$, then by the (3.3) the covariance is 0, as well as all other relations between the two random variables involved.

Chapter 4

Discussion about Risks

In finance, the two key factors that most investors pay close attention to are risk and return. Usually, log returns is the measure of return, while the variance of log return is widely used for presenting risk. In this section, I will investigate two models, the Black-Scholes model and the peaks-over-threshold model, and analyze their implication on risk measures.

4.1 Black-Scholes Model

The Black-Scholes model is a mathematical description of financial markets and derivative investment instruments. Under specific assumptions, I assume that the price of a stock follows a geometric Brownian motion. That is,

$$dS(t) = (\mu + \frac{\sigma}{2})S(t)dt + \sigma S(t)dB_t, \tag{4.1}$$

where B_t is a Brownian motion.

So, I get the solution to equation (4.1)

$$S(t) = S(0)e^{\mu t + \sigma B_t}, \tag{4.2}$$

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

Consider the logreturn of the stock price $X(t)$:

$$X(t) = \log \frac{S(t)}{S(t-\Delta)} = \log(e^{\mu(t)+\sigma B_t - \mu(t-\Delta) - \sigma B_{t-\Delta}}) = \mu\Delta + \sigma(B_t - B_{t-\Delta}), \tag{4.3}$$

where Δ is time interval between sample point. Hence, the logreturns of stock prices are the increments of a Brownian motion, i.e. they are stationary and independent.

However, is that true in the real stock markets? The answer is more likely to be no. By examining the data of stock price, it can easily be found that the volatility is not constant. Therefore, Black-Scholes Model may not be appropriate in this case.

4.2 Peaks-Over-Threshold Model

In real world, the stock returns have statistical distribution with heavy-tails. Rapid changes and complex interdependencies force us to go beyond standard statistical models and simplifying assumptions of normality to develop more sophisticated methodologies which capture downside risk.

Extreme value theory is a branch of statistics dealing with the extreme deviations from the median of probability distributions. It is a useful tool for estimating the tail of stock return distribution. Since only the data in the tail are needed to calibrate the parameters of this distribution, extreme value theory allows us to obtain an effective tail estimation without making any particular assumption about the shape of the tail.

In the following, I will investigate an example of extreme value theory application, the peaks-over-threshold model, which is based on an asymptotic approximation of the excess return distribution by generalized Pareto distribution.

In peaks-over-threshold model, the historical dataset is sorted, and the observations exceeding a specified threshold are fitted to a generalized Pareto distribution. The expression for the generalized Pareto distribution is as follows:

The cumulative distribution function is

$$F_{(\xi,\mu,\sigma)}(x) = \begin{cases} 1-(1+\frac{\xi(x-\mu)}{\sigma})^{-1/\xi} & \text{for } \xi \neq 0, \\ 1-\exp(-\frac{x-\mu}{\xi}) & \text{for } \xi = 0. \end{cases} \tag{4.4}$$

And the probability density function is:

$$f_{(\xi,\mu,\sigma)}(x) = \frac{1}{\sigma}(1+\frac{\xi(x-\mu)}{\sigma})^{-(1/\xi+1)} = \frac{\sigma^{1/\xi}}{(\sigma+\xi(x-\mu)^{1/\xi+1}} \tag{4.5}$$

for $x \geq \mu$ when $\xi \geq 0$, and $x \leq \mu - \xi/\sigma$ when $\xi < 0$, where $\mu \in R$ is the location parameter, $\sigma > 0$ the scale parameter and $\xi \in R$ the shape parameter.

The peaks-over-threshold model offers a unifying way to model the tail of a severity distribution, since for many different underlying distributions the generalized Pareto distribution can be fitted above a sufficiently high threshold. Thus, the peaks-over-threshold method allows us to use the generalized Pareto distribution as the principal modeling tool for the probabilistic risks of large losses.

Chapter 5

Construction of Copula Models

After acquiring the necessary knowledge about copula, I am ready to construct copula models. First, I will present the way how to generate random variables whose dependence structure is define by a copula. Second, I will demonstrate the method to test the fitness of a specific copula.

5.1 The Generation Method

Method 1 Assume that all the parameters of the joint distribution is already known. The joint density function is bounded, i.e. $f(x, y) \leq N$, and random variables come from a black box. Then a random variate (X, Y) from this dependence structure can be generated in the following way:

1. Generate two uniform distributed random variables ξ, η from the certain box;
2. Generate a uniform variable, γ from $[0, N]$;
3. If $f(\xi, \eta) > \gamma$, accept (ξ, η) to the data sets, otherwise we need to go back to the first step.

Since $c(u, v)$ may not be bounded, the uniform random variables may have to be transformed:

$$F(x, y) = C(F_X(x), F_Y(y)). \tag{5.1}$$

Further, its joint density function is given by :

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^2}{\partial x \partial y} C(F_X(x), F_Y(y)) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y), \tag{5.2}$$

$c(F_X(x), F_Y(y)) f_X(x) f_Y(y) \leq N$ is bounded.

Method 2 Another method to generate random variables from a chosen copula is formulated by using conditional sampling.

Let c_v denote the conditional distribution function for the random variables U at the given value v of V,

$$c_v(u) = P[U < u | V = v]. \tag{5.3}$$

Since the density function of a uniform distribution constantly equal to one, $f_X(u) = f_Y(v) = 1$. We have

$$c_u(v) = P[V < v|U = u] = \int_{-\infty}^{v} \frac{f(u,y)}{f_X(u)} dy = \int_{0}^{v} c(u,y)dy = \frac{\partial}{\partial x} C(x,v)|_{x=u} = C_u(u,v), \quad (5.4)$$

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 $v \in I$.

In accordance with Formula (5.4), we can use the following conditional distribution method to generate the data:

1. Generate a uniformly distribution random variables $\xi \in I$;
2. Generate $(V|U = \xi)$ from the conditional copula;
3. Now $(\xi, (V|U = \xi))$ will be a random variate with desired distribution.

Moreover, we can generalize the bivariate generation method showed above to three dimension. Since

$$\begin{aligned} f(x,y,z) &= \frac{\partial^3}{\partial x \partial y \partial z} F(x,y,z) \\ &= \frac{\partial^3}{\partial x \partial y \partial z} C(F_X(x), F_Y(y), F_Z(z)) \\ &= c(F_X(x), F_Y(y), F_Z(z)) f_X(x) f_Y(y) f_Z(z), \end{aligned} \quad (5.5)$$

then

$$\begin{aligned} c_{uv}(w) &= P[W < w|U = u, V = v] \\ &= \int_{-\infty}^{w} \frac{f(u,v,z)}{f_X(u) f_Y(v)} dz \\ &= \int_{0}^{w} c(u,v,z)dz \\ &= \frac{\partial^2}{\partial x \partial y} C(x,y,w)|_{x=u,y=v} \\ &= C_{uv}(u,v,w). \end{aligned} \quad (5.6)$$

The generation procedure for a three-dimensional variate with desired distribution is as follows:

1. Generate a uniformly distribution random variables $\xi \in I$;
2. Generate $(V|U = \xi)$ from the conditional copula (5.4);
3. Generate $(W|U = \xi, V = (V|U = \xi))$ from the conditional copula (5.6);
4. $(\xi, (V|U = \xi), (W|U = \xi, V = (V|U = \xi)))$ is the random variate in desire.

It is easy to generate multidimensional variates with a desirable distribution in the same way.

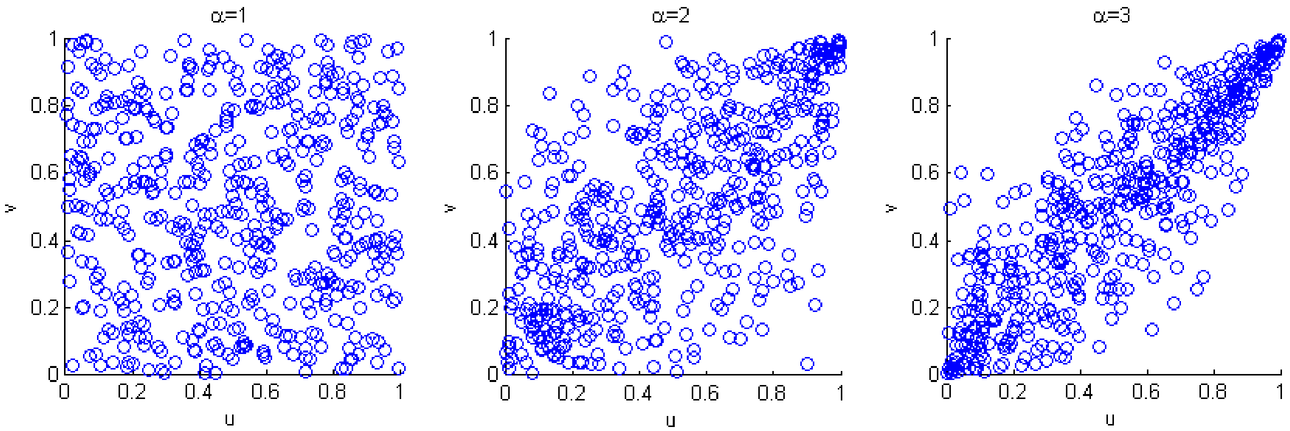

Figure 5.1: Random pairs generated from Gumble copula

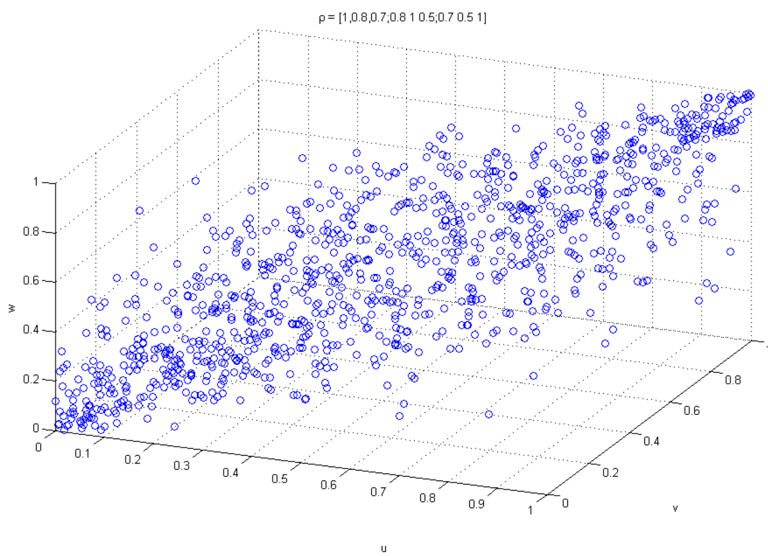

Figure 5.2: Random numbers generated from Gaussian copula

5.2 Test the Goodness of Fit for a Specific Copula

The following method will be used to test the fitness of the copula models:

1. Find parameters of the hypothetical theoretical copula function by using maximum likelihood method.

2. Calculate hypothetical theoretical copula from the oringinal data;

3. Calculate empirical copula from the oringinal data;

4. Find Kolmogorov -Smirnov(KS) distance of the empirical copula to the hypothetical

theoretical copula by formula

$$D_{KS}^{2D} = \max_{u,v\in[0,1]} |C_{emp}(u,v) - C_{the}(u,v)|, \tag{5.7}$$

or

$$D_{KS}^{3D} = \max_{u,v,w\in[0,1]} |C_{emp}(u,v,w) - C_{the}(u,v,w)|. \tag{5.8}$$

Chapter 6

Data Exploration and Investigation

In this section, first I will study the statistical characteristics of the stock returns, and use the traditional way, linear correlation, to investigate the dependency between the stock pairs. Then I will construct bivariate copula models and apply the new method to investigate the dependency between different stock returns. Finally, I will extend the bivariate copula models to three-variate copula models.

As we know, Swedbank Robur is the largest Swedish fund in terms of assets under management. Ericsson, H&M and Nordea Bank are the three stocks that take the biggest portion of this fund. The daily log returns of these three stocks in 2010 are collected to perform my analysis.

6.1 Data Exploration

The adjusted closing price is a useful tool when examining historical returns. It gives analysts an accurate representation of the firm's equity value beyond the simple market price, since it accounts for all corporate actions such as stock splits, dividends and rights offerings. Thus, we start with taking a look at the adjusted closing prices of the three stocks.

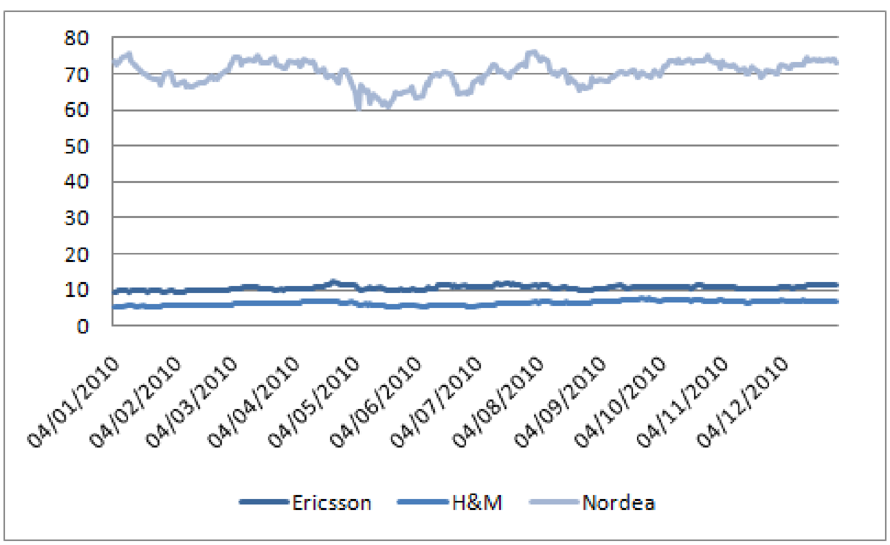

Figure 6.1: Adjusted closing prices

From Figure (6.1)we can see that the adjusted closing prices of the three stocks shake over time, and have slightly upward (Ericsson and H&M) or downward (Nordea) tendency.

Secondly, I calculate the daily returns (log returns) of each stock, and present them in Figure (6.2).

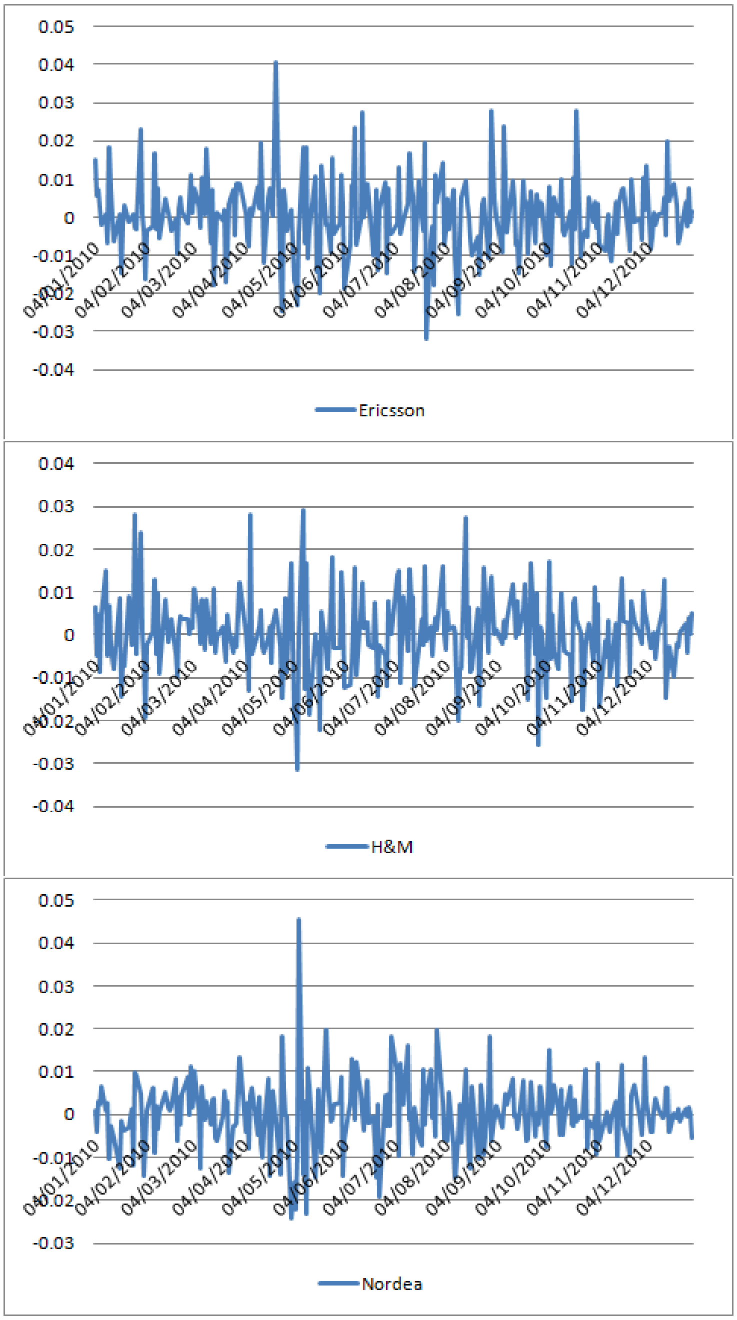

Figure 6.2: Daily returns

The distributions of the daily returns as well as the number of data, mean and standard deviation are summarized in the histograms, see Figure (6.3).

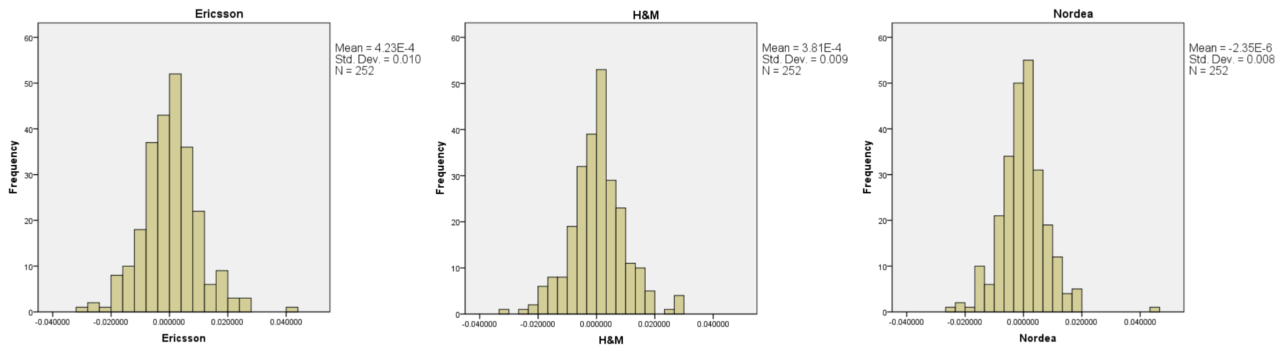

Figure 6.3: Histogram

From the histograms, it is not difficult to see that the daily returns of each stock approximately follow normal distribution, but display with heavy tails. It is also suggested by the normal quantile-quantile plot of data in Figure (6.4). The tails clearly diverge from the normal distribution. However, this is not compliant with the assumption of Black-Scholes model. Therefore, in our case, the Black-Scholes model may not be a precise way to valuate daily return or stock price.

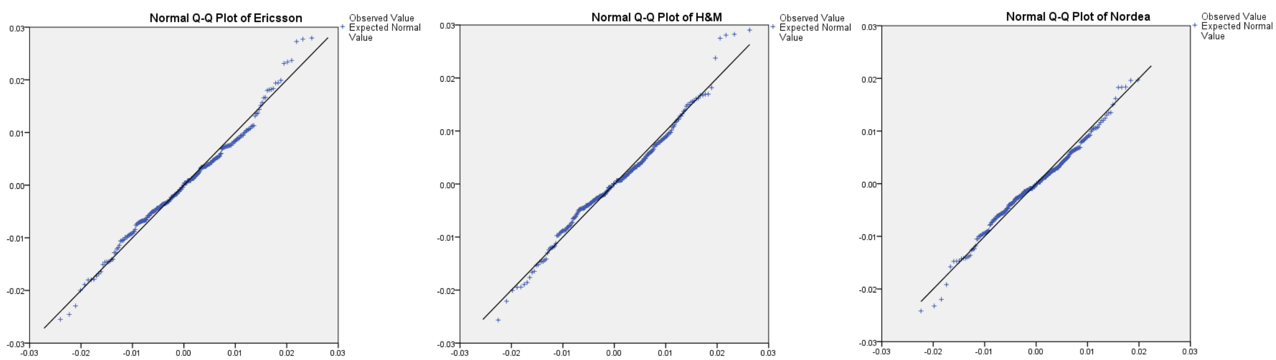

Figure 6.4: Normal quantile-quantile plot

6.2 Linear Correlation Investigation

Correlations between stocks and their significance level is calculated and shown in Table (6.1).

Correlation (Significant Level)	Ericssion	H&M	Nordea
Ericsson	1	0.589(0.000)	0.029(0.643)
H&M	0.589(0.000)	1	0.099(0.117)
Nordea	0.029(0.643)	0.099(0.117)	1

Table 6.1: Correlation and significant level

We can easily find out that the stock returns are all positively correlated. The correlation between Ericsson and H&M is significantly different than zero, while the correlation between Ericsson and Nordea, as well as the correlation between H&M and Nordea are not.

Then I display the stock return pairs in the scatter plots.

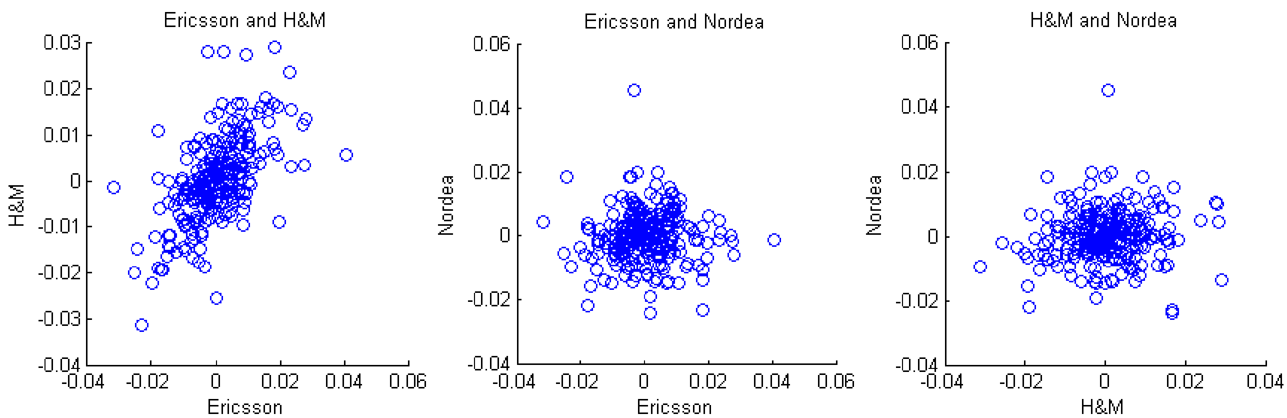

Figure 6.5: Stock returns displayed in pairs

Figure (6.5) further illustrates that linear correlation may not be a reliable measure to describe the dependence structure, because it is obvious that the stock pairs do not have joint elliptical distribution, which is assumed by correlation. What is more, it appears to be lots of outliers (extreme values) in the scatter plots, and they might or cannot provide information about the stocks. Since correlation can be very sensitive to excluding these outliers, we have to make our own judgement on whether keep or remove them, which could render bias in the analysis.

Therefore, we can conclude that linear correlation fail to be a precise way to describe dependence structure in this case.

6.3 Finding Bivariate Copula for Data Sets

In this section, I will find suitable bivariate copulas for stock return pairs showed in Figure (6.5) and then test the fitness of each copula.

First, I transform the stock returns to the copula scale (unit square) using a kernel estimator of the cumulative distribution function. The results are shown in the scatter plots, Figure (6.6).

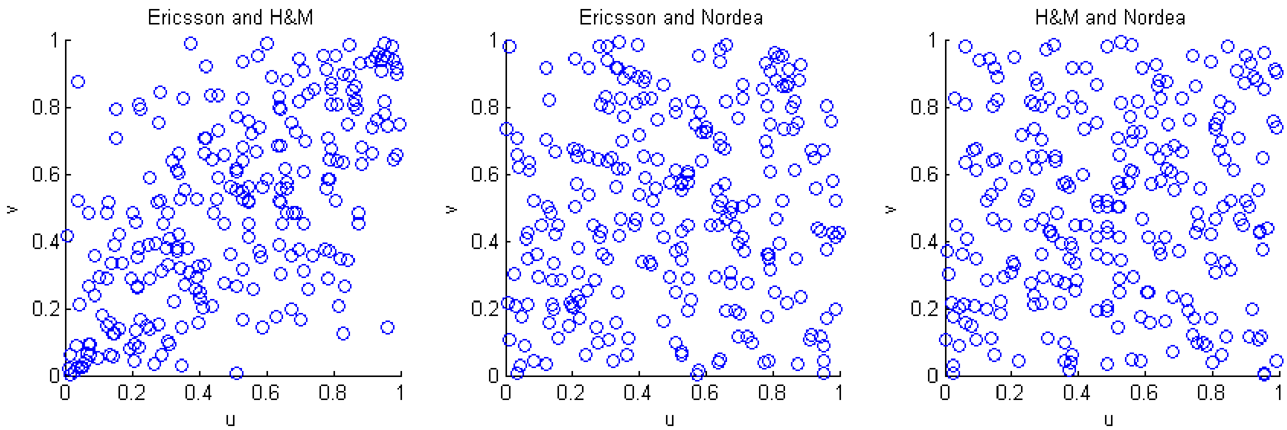

Figure 6.6: Marginal distribution of stock returns displayed in pairs

Compared with Π copula, the input data are not distributed randomly. The three stocks are more likely to boom or crash at the same time.

Secondly, I calculate the empirical copula between each two stock returns. The empirical copulas and their contour plots are displayed in the following three figures.

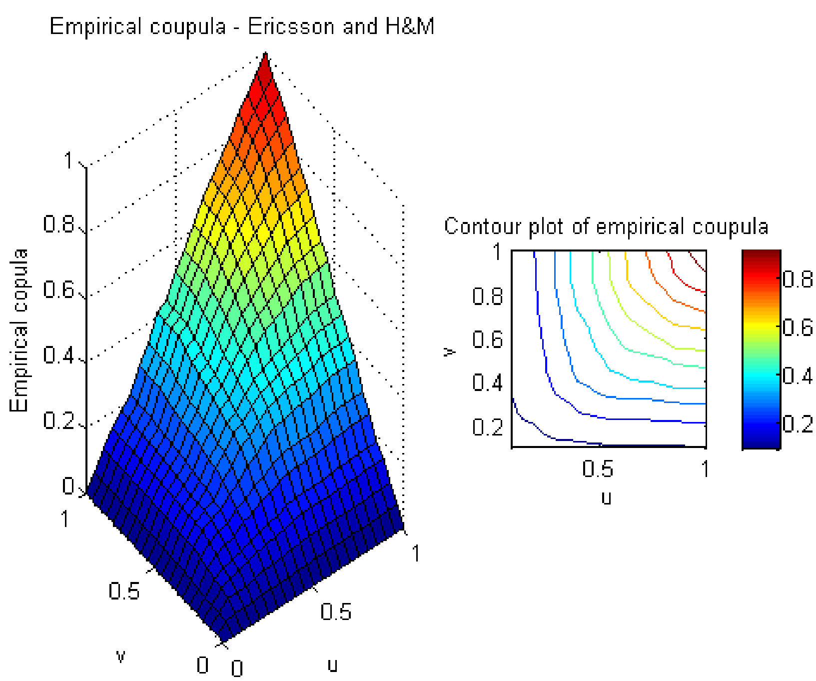

Figure 6.7: Empirical copula and corresponding contour plot - Ericsson and H&M

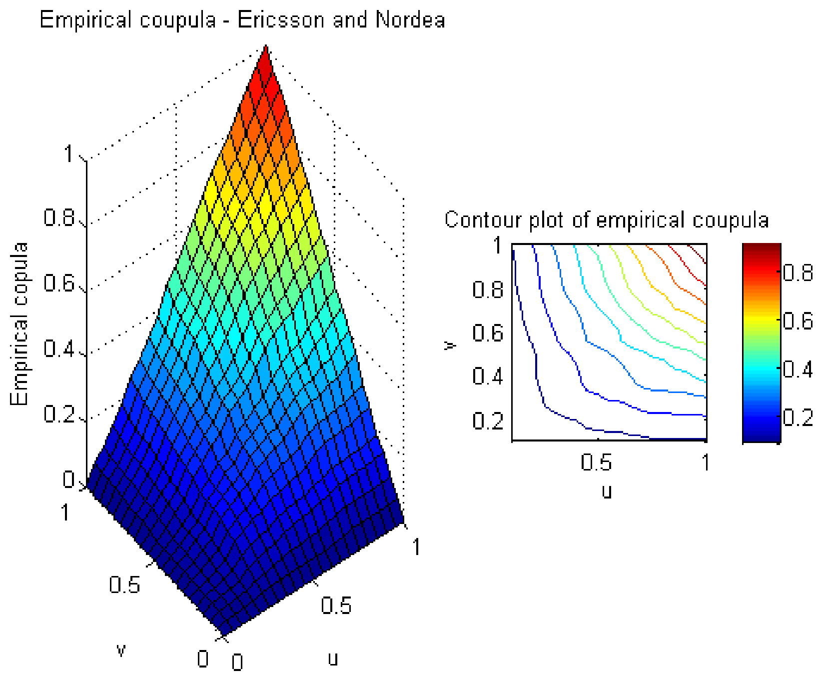

Figure 6.8: Empirical copula and corresponding contour plot - Ericsson and Nordea

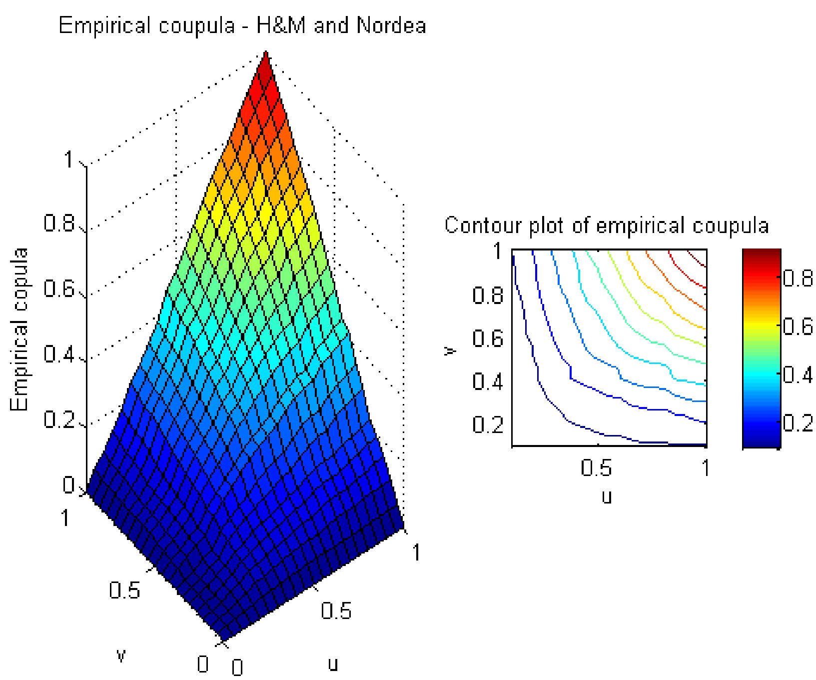

Figure 6.9: Empirical copula and corresponding contour plot - H&M and Nordea

Thirdly, I employ maximum likelihood method to fit a copula to a bivarate data set, and find a suitable copula parameter for each copula function.

Then I calculate the KS distance between empirical copula and theoretical copulas in order to test the fitness of each Archiemedean copula. The smaller KS distance suggests the better fit, since here I take the empirical copula as a benchmark. The parameters and KS distances for each copula got from different data sets are summarized in Table (6.3), Table (6.3) and Table (6.3).

	Parameter	KS Distance
Gaussian	$\rho = 0.6053$	$KS = 0.0407$
Student's t	$\rho = 0.6171, \nu = 5.4990$	$KS = 0.0379$
Frank	$\alpha = 4.8740$	$KS = 0.0415$
Gumbel	$\alpha = 1.6917$	$KS = 0.0389$
Clayton	$\alpha = 1.1983$	$KS = 0.0631$

Table 6.2: Copula parameter and KS distance for Ericssion and H&M

	Parameter	KS Distance
Gaussian	$\rho = 0.0572$	$KS = 0.0375$
Student's t	$\rho = 0.0676, \nu = 24.0313$	$KS = 0.0356$
Frank	$\alpha = 0.5835$	$KS = 0.0332$
Gumbel	$\alpha = 1.0034$	$KS = 0.0472$
Clayton	$\alpha = 0.1258$	$KS = 0.0377$

Table 6.3: Copula parameter and KS distance for Ericssion and Nordea

	Parameter	KS Distance
Gaussian	$\rho = 0.1080$	$KS = 0.0372$
Student's t	$\rho = 0.1122, \nu = 5.2562$	$KS = 0.0357$
Frank	$\alpha = 0.7691$	$KS = 0.0373$
Gumbel	$\alpha = 1.0854$	$KS = 0.0369$
Clayton	$\alpha = 0.1528$	$KS = 0.0405$

Table 6.4: Copula parameter and KS distance for H&M and Nordea

For an easy comparison, the following histogram, Figure (6.10), is adopted to give us a visual display of KS distance .

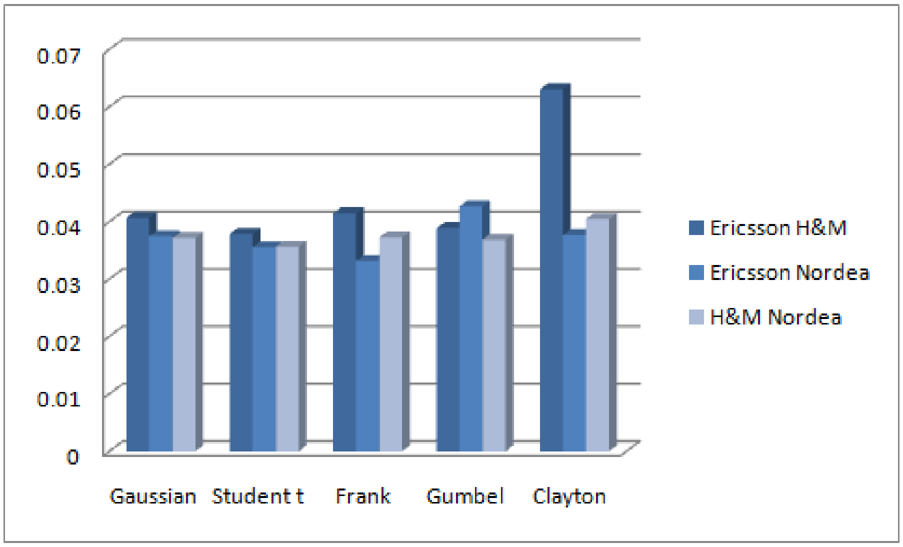

Figure 6.10: Histogram of KS distance

As is discussed before, Gaussian copula, Student's t copula and Frank copula have symmetric dependence structure, while Cumbel copula and Clayton copula have asymmetric dependence structure. Following Gumbel copula dependence structure, stocks would have a larger probability of simultaneous booming than simultaneous crashing, while following Clayton copula dependence structure, stocks behave conversely.

Therefore, Ericsson and H&M as well as H&M and Nordea have the least KS distance with Student's t copula. They are more likely to follow Student's t copula dependence structure. They would have almost the same probabilty of simultaneous booming or crashing. Ericsson and Nordea have the least KS distance with Frank copula. They may follow Frank copula dependence structure. They would also have almost the same probabilty of simultaneous booming or crashing, but have a larger probability in the middle region which means theirs tails must be lighter.

6.4 Finding Three-Variate Copula for Data Sets

A similar numerical method is adopted for the three-dimensional model. Take Guassian copula and Student's t copula as examples.

First, we get the three-variate data set of stock returns, Figure (6.11).

Secondly, I transform the stock returns to an unit cube, Figure (6.12).

After transforming the stock returns to the copula scale, I calculate the empirical copula of the three stock returns. Sadly, it is difficult to show high-dimension empirical copula.

Then I use maximum likelihood method to fit a copula to the three-variate data set, and find a suitable copula parameter for Gaussian and Student's t copula function.

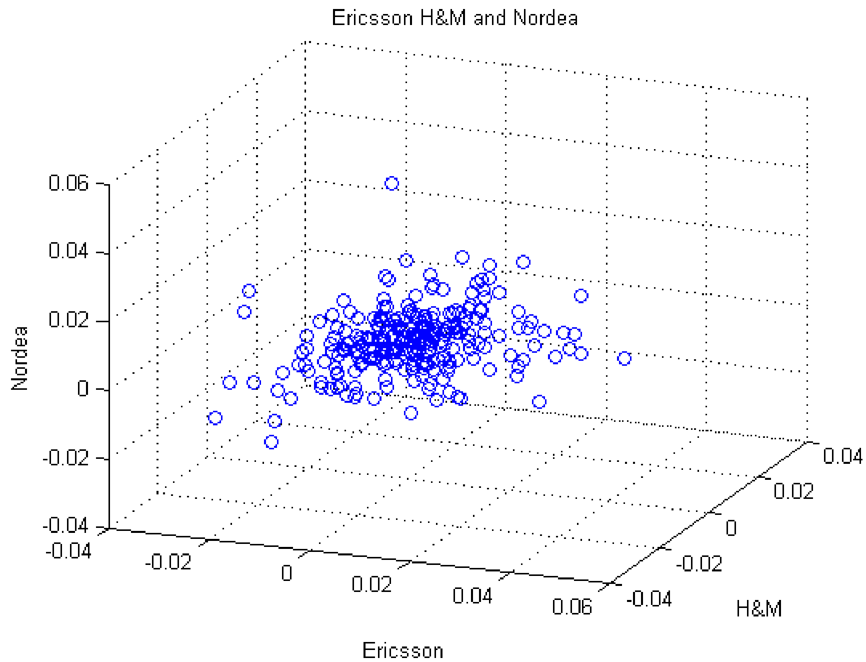

Figure 6.11: Stock returns

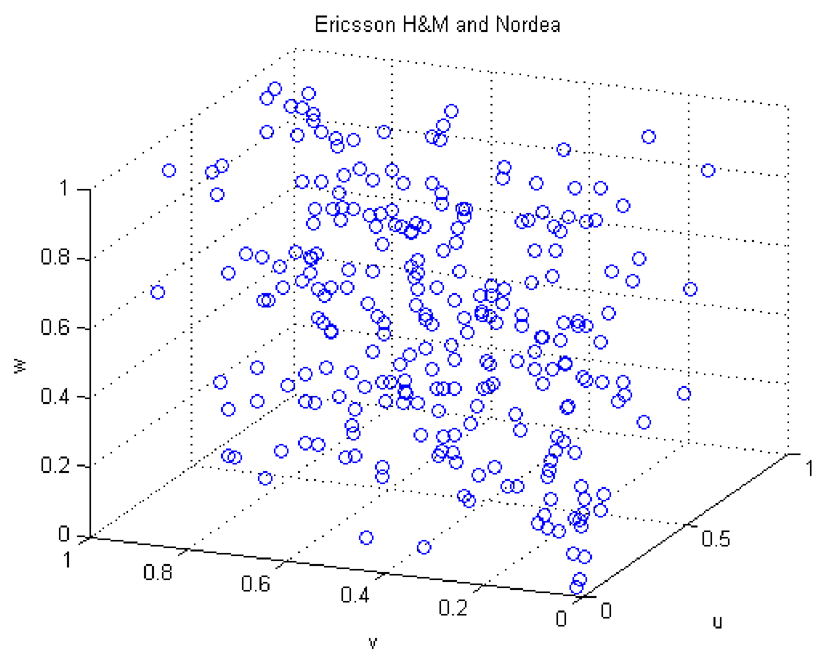

Figure 6.12: Marginal distribution of stock returns

For Gaussian copula:

$$\rho = \begin{bmatrix} 1.0000 & 0.6053 & 0.0572 \\ 0.6053 & 1.0000 & 0.1080 \\ 0.0572 & 0.1080 & 1.0000 \end{bmatrix}. \tag{6.1}$$

For Student's t copula:

$$\rho = \begin{bmatrix} 1.0000 & 0.6375 & 0.0881 \\ 0.6375 & 1.0000 & 0.1314 \\ 0.0881 & 0.1314 & 1.0000 \end{bmatrix}, \tag{6.2}$$

and the degree of freedom $\nu = 8.3633$.

In the end, I calculate the KS distance between the fitted Gaussian copula and empirical copula, which equals to 0.2473, and the KS distance between the fitted Student's t copula and empirical copula, which equals to 0.2412.

The KS distances for the three-variate copula models are much larger compared to all of the KS distances for the bivariate copula models.

Chapter 7

Risk Applications

In this section, I will model the risk of a portfolio, which consists of the three stocks I discuss above, with a simulation technique using copula method as well as extreme value theory. The process firstly constructs the sample marginal cumulative distribution function of each stock return using a Gaussian kernel estimate for the interior and a generalized Pareto distribution estimate for the upper and lower tails. A bivariate or a three-variate copula is then fit to the data. In the end, I will simulate the stock returns in two as well as three dimensions by using copula technique and extreme value theory. For reference, I will also simulate and plot centered returns using Gaussian distribution.

7.1 Estimating the Piecewise Cumulative Density Functions

The selection of an appropriate threshold is one of the main concerns about the peaks-over-threshold model, and there are many ways to find the desirable threshold. Here, I just adopt a simple way to find upper and lower thresholds, such that 10% of the residuals is reserved for each tail.

Then I will fit the amount by which those extreme residuals in each tail fall beyond the associated threshold to a parametric generalized Pareto distribution by maximum likelihood method.

The piecewise distribution function allows not only interpolation within the interior of the Cumulative Density Function, but also extrapolation in each tail. Extrapolation allows estimation of quantiles outside the historical record, which is invaluable for risk management applications.

For comparision, I also plot the cumulative density function of the Gaussian distribution with the same risk. This is adopted by Black-Scholes model, which assumes that the stock returns follow Gaussian (normal) distribution. The mean and the standard deviation as the only required two parameters of the Gaussian distribution can be easily derived from the past data.

From Figure (7.1), Figure (7.2) and Figure (7.3), we can see that piecewise cumulative density function is relatively more flat. It allows the stock returns to have larger distribution in both the upper and lower tails than corresponding Gaussian cumulative density function, which complies with the behaviour of stock returns in the real stock market.

In the following two sections, I will take stock returns of Ericsson and H&M as an example to perform my analysis with bivariate copulas.

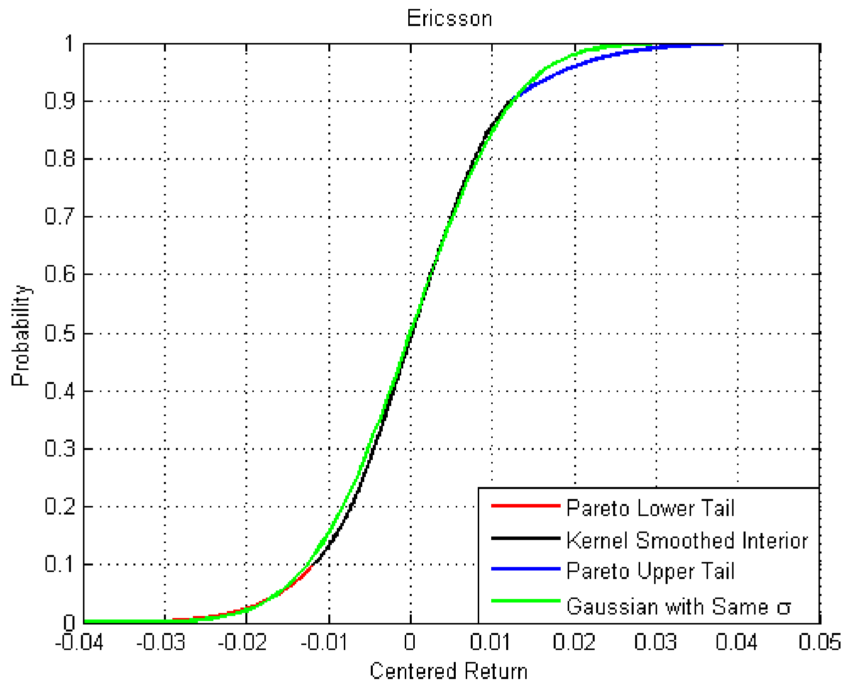

Figure 7.1: Semi-parametric/piecewise cumulative density function: Ericsson

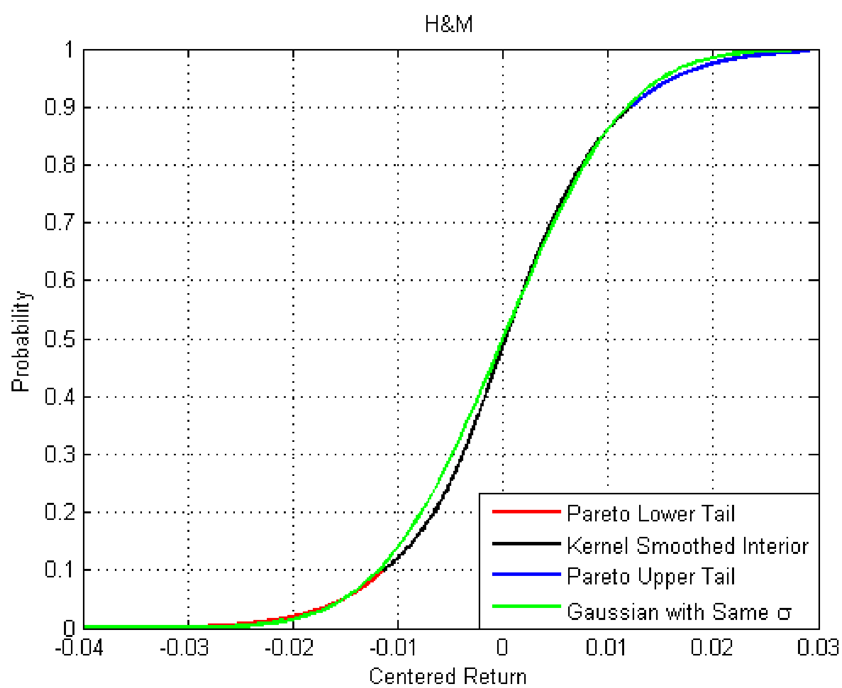

Figure 7.2: Semi-parametric/piecewise cumulative density function: H&M

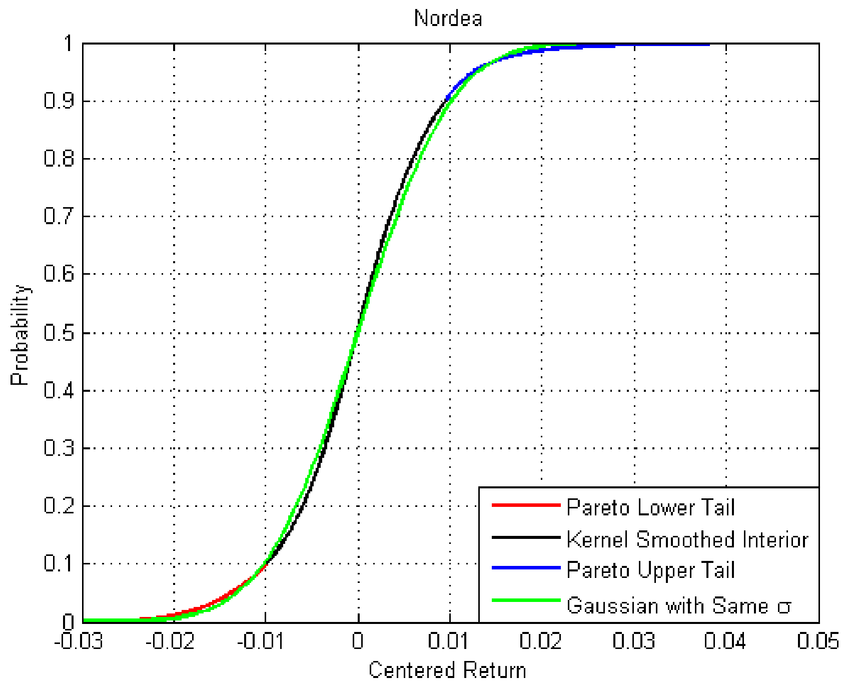

Figure 7.3: Semi-parametric/piecewise cumulative density function: Nordea

7.2 Fitting Bivariate Copulas for Data Sets

Firstly, I use the daily returns to estimate the parameters of different bivariate copulas by maximum likelihood method. The parameters of each copula are summarized in the following Table (7.2).

	Parameter
Gaussian	$\rho = 0.6013$
Student's t	$\rho = 0.6141, \nu = 6.0231$
Frank	$\alpha = 4.8787$
Gumbel	$\alpha = 1.6780$
Clayton	$\alpha = 1.1517$

Table 7.1: Copula parameter for Ericssion and H&M

We can see that the parameters are totally different from the parameters we get in the copula model without application of extreme value theory.

7.3 Simulating Stock Returns with Bivariate Copula

Now that the copula parameters have been estimated, I will first simulate jointly-dependent uniform variates. Then, by extrapolating the Pareto tails and interpolating the smoothed interior, I will transform the uniform variates derived from copularnd to daily centered returns via the inverse cumulative density function of each stock return. In the end, I will display the simulated stock returns in the following five figures.

The following five figures suggest that the stock returns generated from Gausian copula, Student's t copula and Frank copula are symmetric, while the stock returns generated from Gumbel copula and Clayton copula are asymmetric. The stock returns generated from Student's t copula have heavier tails than that from Gaussian copula. Conversely, the stock returns generated from Frank copula have lighter tails than that from Gaussian copula. What is more, the upper tail of the stock returns generated from Gumbel copula is heavier than the lower tail, which means two stock returns have a larger probability of simultaneous booming than simultaneous crashing. However, the lower tail of the stock returns generated from Gumbel copula is heavier than the upper tail, and they would have a larger probability of simultaneous crashing than simultaneous booming.

It is indicated that the choices of copula functions may have different effects on the simulated stock returns. Therefore, it is necessary for analysts to make investigation of the stock return distribution in advance, so that they can choose a suitable copula for their data sets.

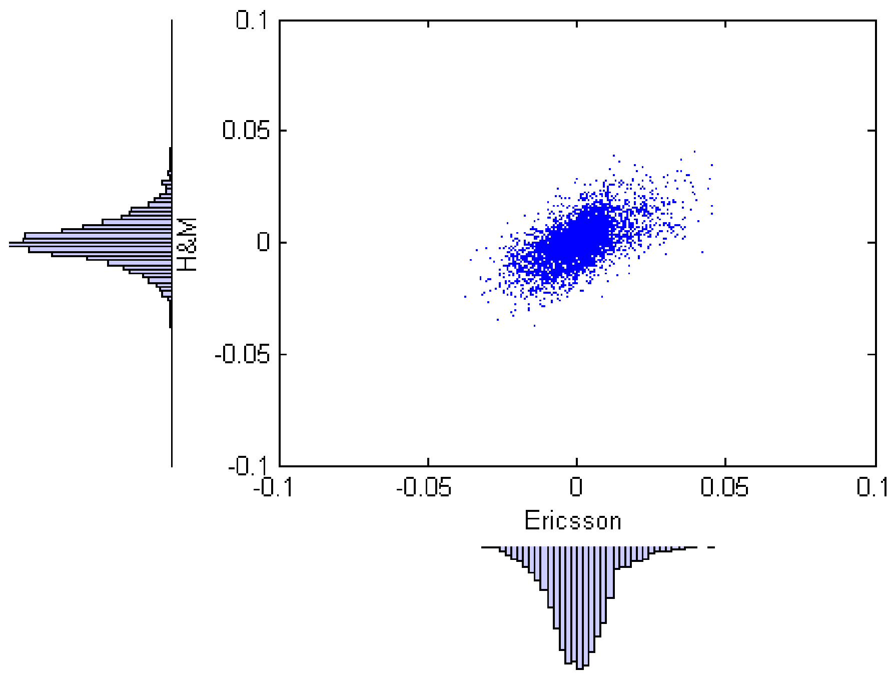

Figure 7.4: Simulation of stock returns by bivariate Gaussian copula

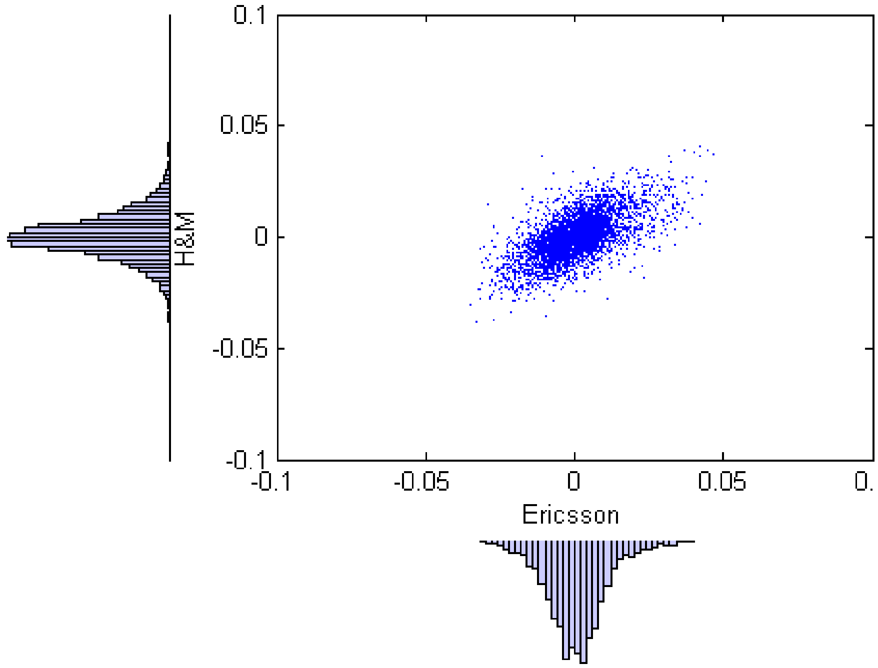

Figure 7.5: Simulation of stock returns by bivariate Student's t copula

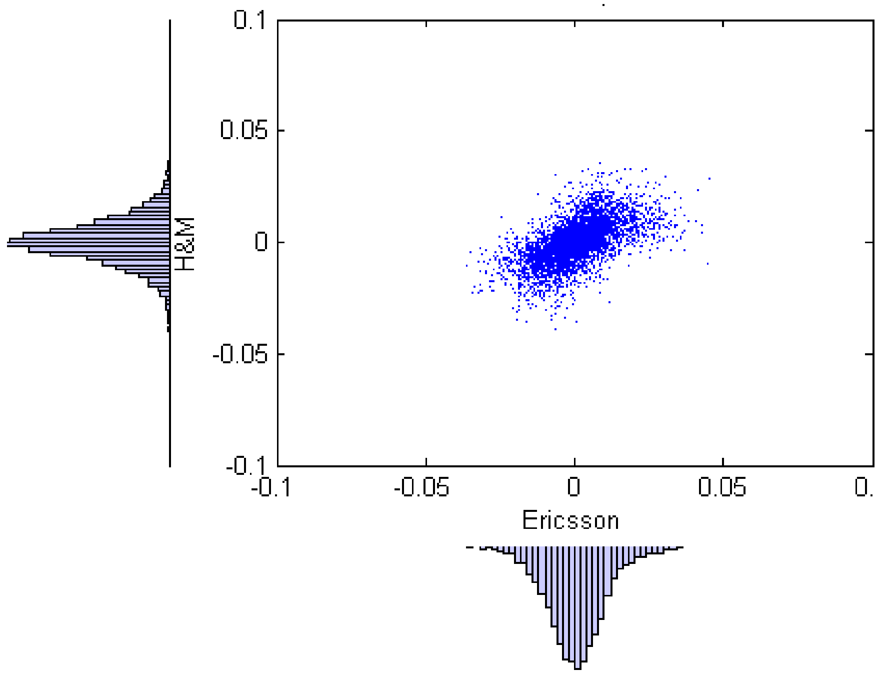

Figure 7.6: Simulation of stock returns by bivariate Frank copula

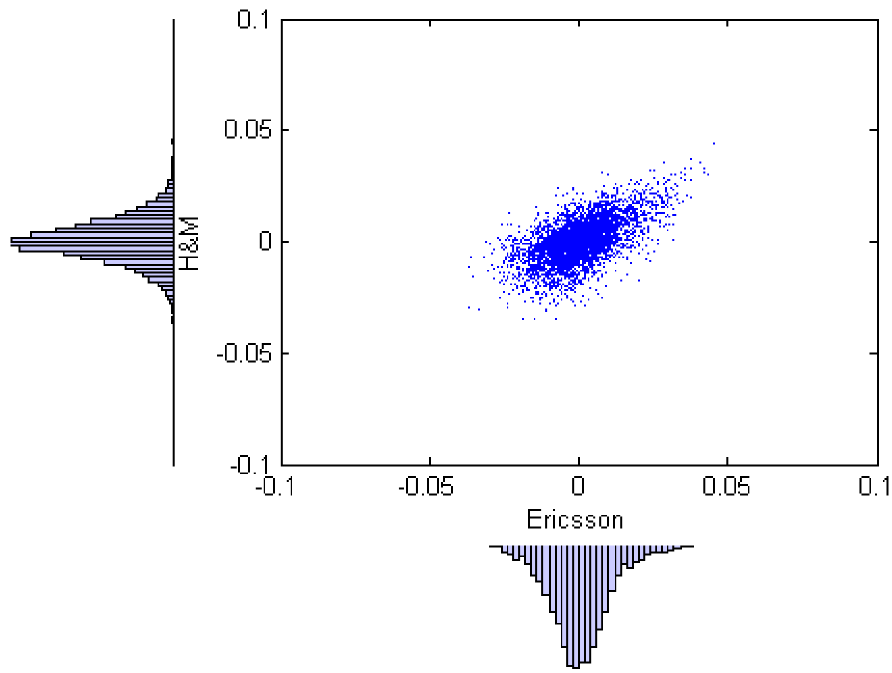

Figure 7.7: Simulation of stock returns by bivariate Gumbel copula

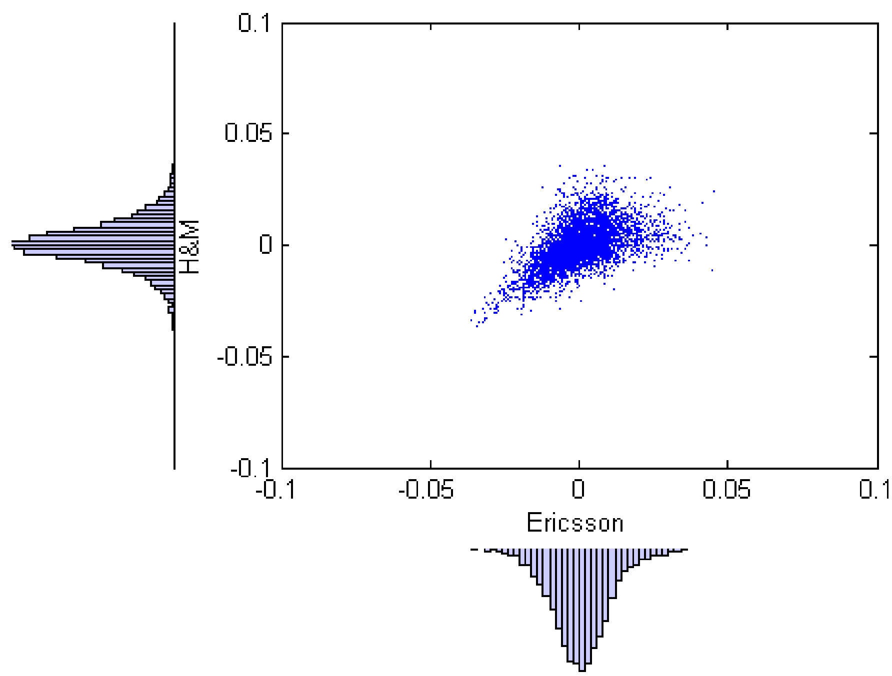

Figure 7.8: Simulation of stock returns by bivariate Clayton copula

For reference, I will simulate and plot centered returns with Gaussian distribution, see Figure (7.9).

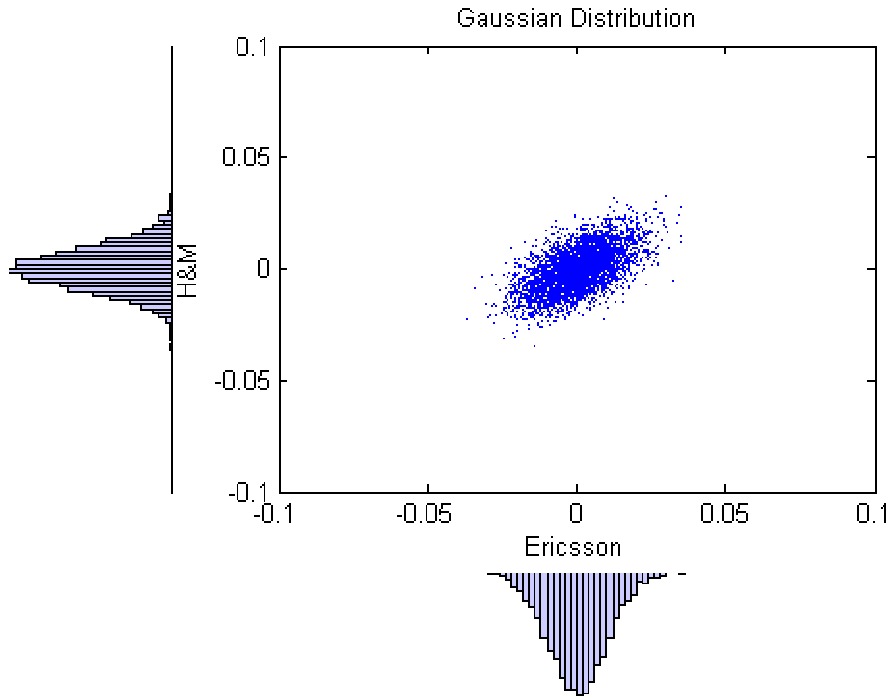

Figure 7.9: Simulation of stock returns by Gaussian distribution

It can be easily find out that the stock returns simulated by the models that I construct are relatively decentralized and display with heavier tails compared to the stock returns purely following Gaussian distribution.

7.4 Fitting Three-Variate Copulas for Data Sets

From this section, I will extend the model into three dimension.

First, I will use the daily returns to estimate the parameters of the three-variate Gaussian copula and Student's t copula.

For Gaussian copula, I get:

$$\rho = \begin{bmatrix} 1.0000 & 0.6013 & 0.0566 \\ 0.6013 & 1.0000 & 0.1094 \\ 0.0566 & 0.1094 & 1.0000 \end{bmatrix}. \tag{7.1}$$

For Student's t copula, I get:

$$\rho = \begin{bmatrix} 1.0000 & 0.6119 & 0.0776 \\ 0.6119 & 1.0000 & 0.1024 \\ 0.0776 & 0.1204 & 1.0000 \end{bmatrix}, \tag{7.2}$$

and the degree of freedom $\nu = 9.1868$.

7.5 Simulating Stock Returns with Gaussian Copula and Student's T Copula

Now that the copula parameters have been estimated, the same method will be used to generate stock returns. Firstly, I will simulate jointly-dependent uniform variates. Secondly, by extrapolating the Pareto tails and interpolating the smoothed interior, I will transform the uniform variates derived from copularnd to daily centered returns via the inverse cumulative density function of each stock return. Finally, I will plot the simulated stock returns in the following figures.

For comparison, I will also simulate and plot centered returns that follow Gaussian distribution.

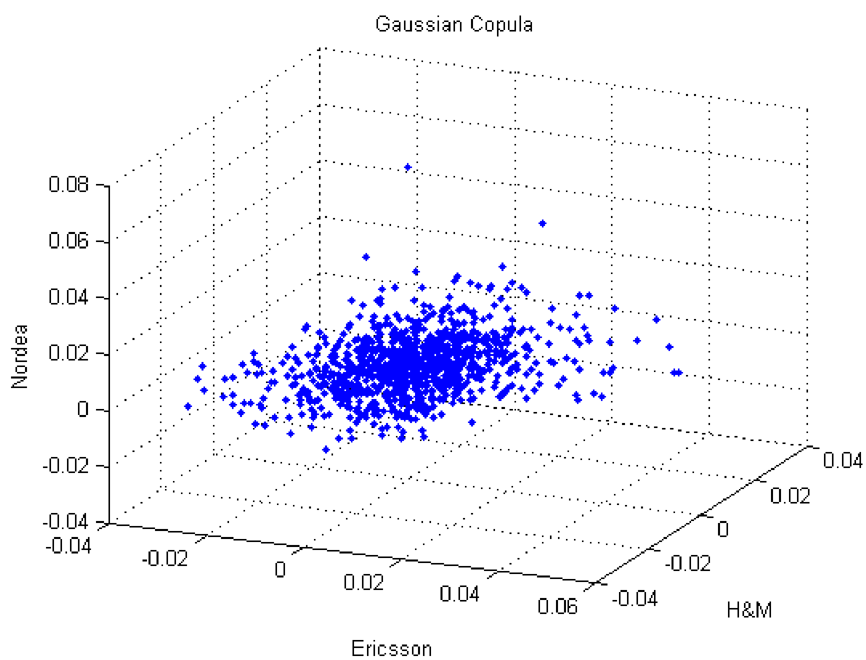

Figure 7.10: Simulation of stock returns by Gaussian copula

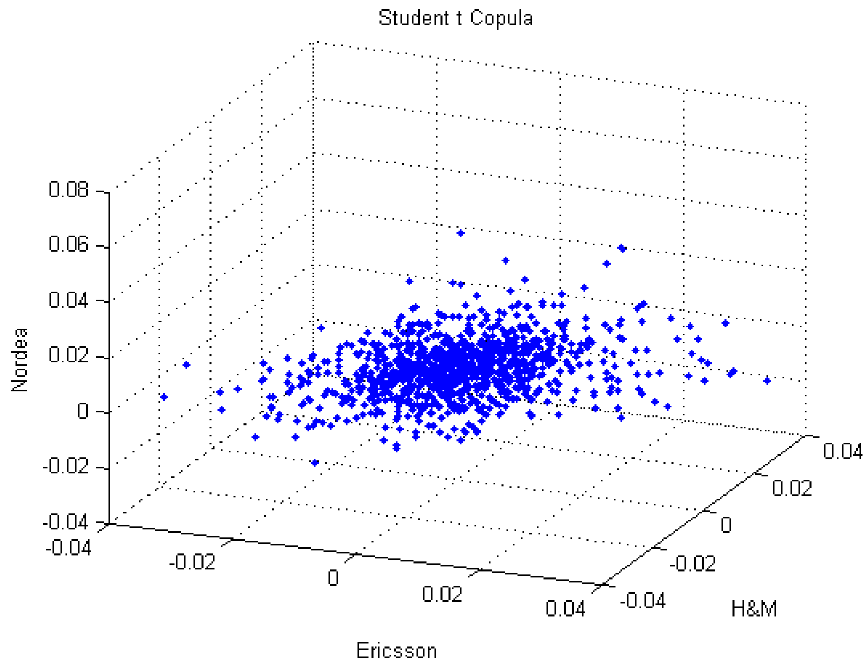

Figure 7.11: Simulation of stock returns by Student's t copula

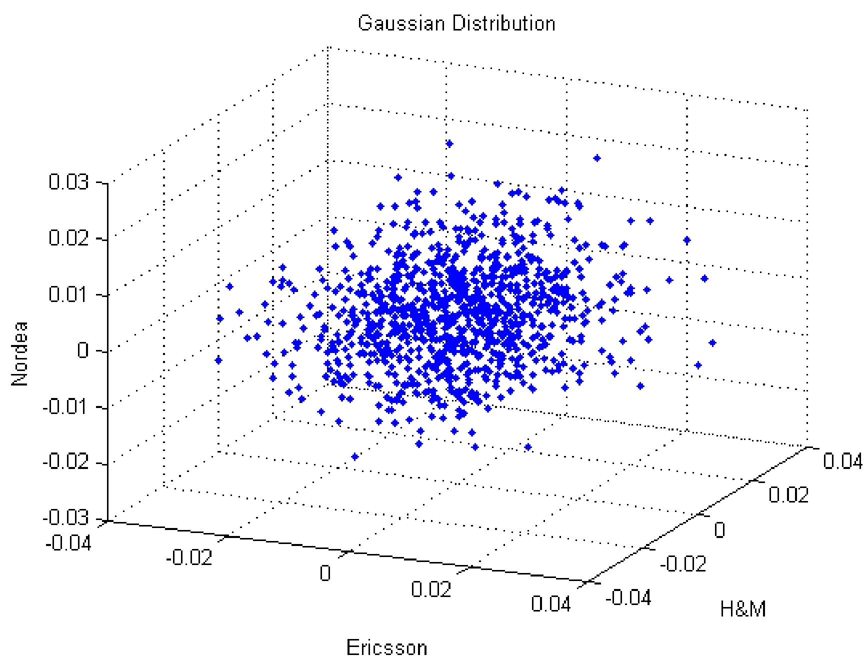

Figure 7.12: Simulation of stock returns by Gaussian distribution

The first two figures above also suggest that the stock returns generated from Gausian copula and Student's t copula are symmetric. Moreover, the stock returns generated from Student's t copula have heavier tails than that from Gaussian copula.

Compared to the stock returns following Gaussian distribution, it is obvious that the stock returns generated by peaks-over-threshold model with application of copula technique are relatively decentralized and display with heavier tails, which can be closer to the real world data.

Chapter 8

Conclusion

In this thesis, I have used both correlation and copula to investigate the dependence structure of the stock returns. And I have got an up-close look at five different Archimedean copulas and their special characteristics. Then, I have cooperated copula technique with peaks-over-threshold model to better assess the dependency of the extreme values. In the end, I have a few points to lay stress on:

Linear Correlation versus Copula

As shown in this thesis, there are several drawbacks of linear correlation. However, copula overcomes these drawbacks. On one hand, linear correlation assume the two variates follow joint elliptical distribution, and only linear dependence can be explained by the correlation. While this is often not in accordance with the reality as we discussed in Section 6.2. On the other hand, linear correlation is inferior in adequately capturing the outliers' information, which can be invaluable to risk managers (also discussed in Section 6.2). However, one of the key strengths of the copula method is its ability to model extreme events, in a financial application, this is a highly valuble property, since the extreme events are what, in many cases, is the most important ones to model.

Moreover, theoretically copula contains more information about dependence structure than correlation, which is derived in Section 3.3.

Therefore, despite of the simplicity and convenience of linear correlation, copula can be a better way to describe dependence structure between different stock returns.

Choice of Copula Functions and Their Parameters

Different copula functions have different characteristics (discussed mainly in Section 2.6), which might imply different dependence structure (discussed in Section 7.3 and 7.5). Hence, when applying the model to real world data it would therefore be advisable to use different copula functions and compare their goodness of fit with KS-distance (the method is shown in Section 5.2 and illustrated in Section 6.3 and 6.4).

What is more, the value of the copula parameter can also be critical to the dependence structure (see Figure (5.1)). In this thesis, I mostly use maximum likelihood method to get optimal values of the parameters. Since the copula parameter is critical for estimating the dependence structure, this topic could be the subject of further investigation.

Copula Model with Risk Application

Because of the huge increase of volatility and erratic behavior of financial markets, analysts has raised their attention to risk management. At the same time, copula has recently become the most significant new tool, since it can handle the co-movement between markets, risk factors and other relevant variables which are studied in finance in a flexible and relatively precise way.

The traditional models, such as Black-Scholes model, assume that the stock return distribution is multivariate conditional normal, which unfortunately diverges from the empirical evidence (discussed in Section 4.1 and shown in Section 6.1). However, peak-over-threshold model can fit the real world data much better.

All in all, the copula plus peaks-over-threshold model outperforms the traditional model. Because it not only can precisely model the co-movement of stock returns, but also takes consideration of extreme booming or crashing.

References

[1] Castillo, E.: Extreme Value Theory in Engineering. San Diego, Academic Press,1988.

[2] Claudia Kluppelberg, Gabriel Kuhn: Copula Structure Analysis for High-dimensional Data.

http://www-m4.ma.tum.de/Papers/Klueppelberg/copstruc070128.pdf, 2009.

[3] Embrechts, P., A. J. McNeil and D. Straumann: Correlation and Dependence in Risk Management: Properties and Pitfalls, To appear in Risk Management: Value at Risk and Beyond, ed. By M. Dempster and H. K. Moffatt, Cambridge University Press, 1999.

[4] Friedrich Schmid: Measuring Conditional Dependence.

http://www.case.huberlin.de/events/Archive/Workshops/Copulae/talks.pdf, 2007.

[5] 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

[6] Naess, A. and P. H. Clausen. Statistical Extrapolation and the Peaks Over Threshold Method. OMAE996422. In Proceedings 18th International Conference on Offshore Mechanics and Arctic Engineering. New York: ASME. 1999.

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

[8] Paul Embrechts, Filip Lindskog, Alexander McNeil:Modelling Dependence with Copulas and Applications to Risk Management

http://www.math.ethz.ch/finance