CHALMERS UNIVERSITY OF TECHNOLOGY

Copula Dependence Structure on Real Stock Markets

by

Jan Lennartsson and Min Shu

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Approved, Thesis Committee:

____

Gothenburg, Sweden

January, 2005

ABSTRACT

Copula Dependence Structure on Real Stock Markets

by

Jan Lennartsson and Min Shu

Linear correlation is only an adequate means of describing the dependence between two random variables when they are jointly elliptically distributed. When the joint distribution is not elliptical the linear correlation coefficient becomes just one of many possible ways of summarizing the dependence structure between the variables. In this thesis project, based on both long term data and short term tick data, the stochastic dependencies among several stocks and risk-free bonds are investigated. One of the objectives of the thesis, besides improving the general understanding of dependence structures between different assets, is to investigate what is the significance of correlation analysis within those dependence structures. The motivation for focusing in part on this problem, is that on the one hand correlation analysis is used as an important tool in portfolio analysis, while on the other hand it is known that correlation in general might give a very poor picture of the the true dependence structure. One method in this thesis is to fit various kind of copulas found in the literature as well as a new one constructed by us to be suitable for the dependencies observed.

KEYWORDS: Copula; Time series; Dependence; Correlation.

Acknowledgments

We are greatly indebted to Patrik Albin for being the fantastic inspirator that he is, as well as for his comments and suggestions.

Further, we are very thankful to Martin Norling for helping us catch data.

We also thank Mathematical Sciences at Chalmers for offering a master's thesis students Lab room.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

Consider a portfolio of N stocks and bonds. Managers of such a portfolio are typically interested in the portfolio value at a certain time T in the future. From a statistical point of view, a problem, dependence among default events, is faced when modelling losses on portfolios.

Dependence across financial markets have been widely studied in the past decades. Three general alternative methods are available in multivariate analysis for studying dependence:

- One approach is to use a joint distribution based on the most commonly used joint distribution in theoretical and empirical finance being the multivariate normal distribution.

- The second approach that has been used in empirical study is to compute conditional correlation.

- The third alternative approach in multivariate analysis is to use a copula model for directly modelling dependence.

The third approach is adopted in this thesis. The study of copulas and their application in financial markets is a rather modern phenomenon. Compared to the joint distribution approach or correlation-based approach, a copula model is a more convenient tool in studying the dependence structure. In statistics, a copula is a function that connects marginal distributions to restore the joint distribution and various copula functions represent various dependence structures between variables. In a copula model, the primary task is to choose an appropriate copula function and a corresponding estimation procedure. Marginal distributions are treated as nuisance functions. This reorientation has desirable advantages in empirical finance. One of the primary goals is to investigate the dependence in order to better understand portfolio allocation. The marginal distributions of asset returns in individual markets may be very complicated and may not easily fit within existing parametric models.

In this project, the dependence structure of several stocks are estimated by using a mixture copula approach. The purpose is to find a simple yet flexible model to summarize the dependence structure. The mixture is composed of the Joe survival copula, the Gumbel copula and the AMH copula.

The mixed method facilitate the separation of the concepts of degree of dependence and structure of dependence, and these concepts are embodied in two different groups of parameters-association parameters and weight parameters.

The datasets of the projects come from typical investment funds in the market. There are not any extremely positive dependence structure and negative dependence structure between the different financial assets involed. Our basic problem is to find an appropriate copula model for the dependencies between these assets.

The thesis report is organized as follow: Section 2 reviews some basic concepts about copulas, and introduces mixture model. Section 3 describes some basic copula models. Section 4 discusses correlation. Section 5 shows how to generate random numbers from copula models. In Section 6 our main statistical investigation is carried out. Section 7 describes our resulting mixture copula model, to model the observed depnedencies. In Section 8 we make conclusions.

Chapter 2
Basic Features of Copulas

In this chapter, we summarize the basic definitions that are necessary to understand the concept of copulas. We then illustrate the most important properties of copulas that are needed to understand the usage of copulas in finance.

We follow the notation used in Nelsen (1999). Furthermore, we will restrict ourselves to copulas in two dimensions. The generalization to n dimensions is not difficult.

2.1 Definition of the Copula

In the statistics literature, the idea of a copula arose as early as the 19th century in the context of discussions of non-normality in multivariate cases. Modern theories about copulas can be dated to about forty years ago when Sklar (1959) defined copulas and showed some of their fundamental properties: By Sklar's theorem, for a copula C,

$$F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = C(F_{X_1}(x_1), F_{X_2}(x_2), ..., F_{X_n}(x_n))^*. \tag{2.1}$$

It is clear that a copula is a mapping from $\mathbf{I^n}$† to $\mathbf{I}$, i.e. a multivariate distribution with uniform marginals on $\mathbf{I}$. From (2.1), it is evident 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* $C : \mathbf{I^n} \to \mathbf{I}$ *is called a* copula *if the following conditons hold:*

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

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

2. *for every* $u_k \in \mathbf{I}$

$$C(1, 1, ...u_k, ..., 1) = u_k;$$

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

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

---

* $X_1, ..., X_n$ are random variables, $F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$ denotes the joint distribution function and $F_{X_i}(x_i)$ denotes the marginal distribution function of X_i.

† I denots the interval [0,1].

Now we will restrict ourselves to the bivariate copula.

Definition 2.1.2 *A* bivariate copula *is a function* $C : \mathbf{I} \times \mathbf{I} \to \mathbf{I}$ *with the following properties:*

1. For every $u, v \in \mathbf{I}$

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

2. for every $u, v \in \mathbf{I}$

$$C(u, 1) = u$$

and

$$C(1, v) = v;$$

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

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

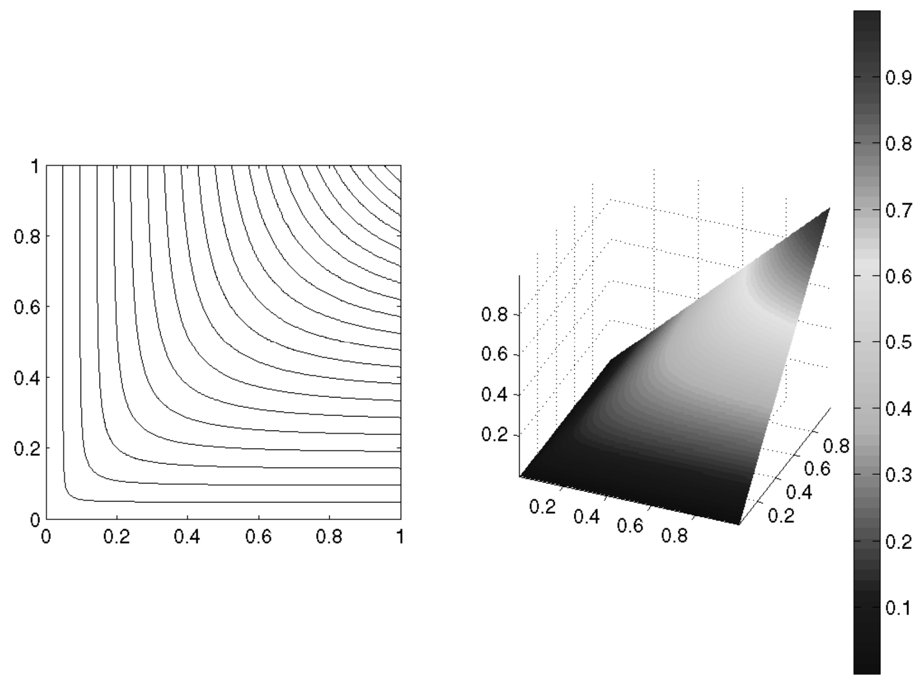

Figure 2.1 Example of a copula function.

A function that fulfils Property 1 is said to be *grounded*, Property 3 is the two-dimensional analogue of a nondecreasing one-dimensional function. A function with this feature is therefore called *2-increasing*, see Figure 2.1.

2.2 The Probability Density Function of Copulas

Due to virtual similarity of all copula functions, it is hard to visualize differences between these distribution functions. So rather it is convenient to study density functions of copulas.

The density of a copula C is given by

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

if C is a continuously differentiable function of u and v.

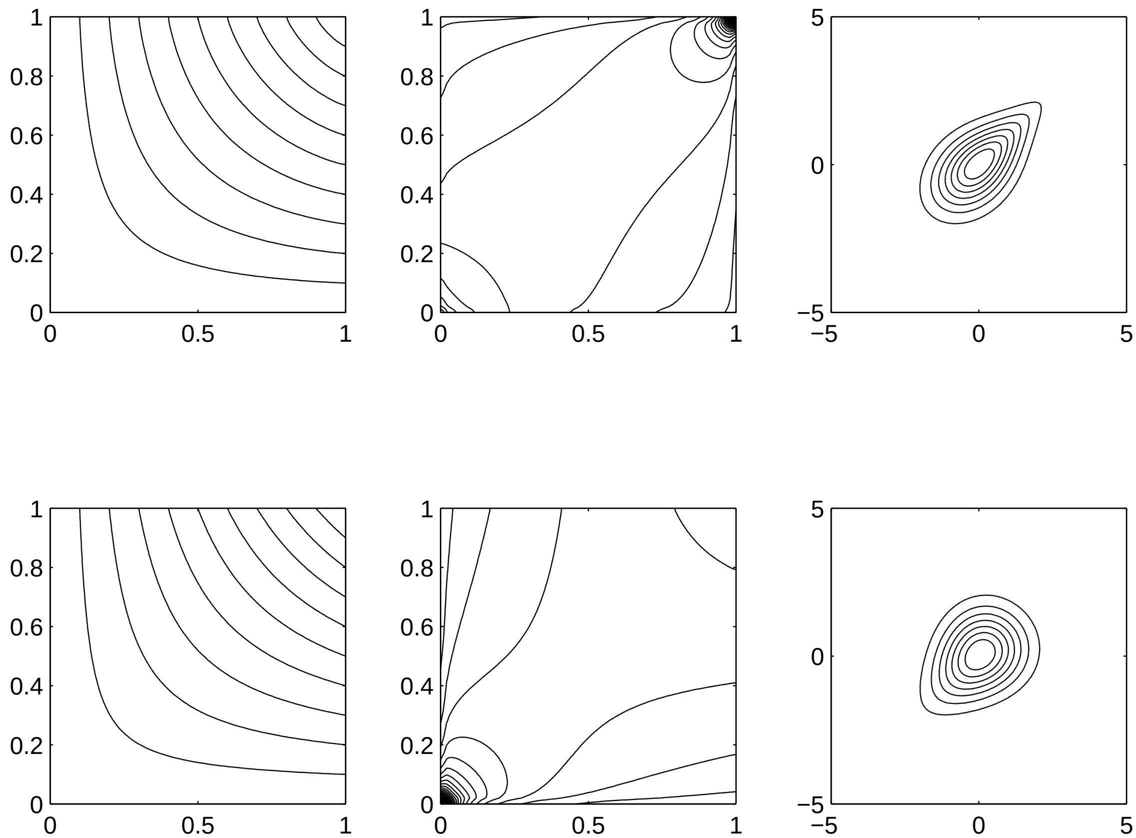

Figure 2.2 First column displays copula functions, second column probability density functions of these copulas, and third column density functions of distributions with these copulas and standard normal marginal distributions.

From Figure 2.2, it seems that the the first column displays two almost equal copula function, but from the two columns to the right, it is clear that their dependence structures are in fact quite different.

Theorem 2.2.1 *Let C be a copula. For every $u \in \mathbf{I}$, the partial derivative $\partial C/\partial v$ exists for almost all $v \in \mathbf{I}$. For such u and v one has*

$$0 \leq \frac{\partial}{\partial v} C(u,v) \leq 1. \tag{2.2}$$

The analogous statement is true for the partial derivative $\partial C/\partial u$. In addition, the functions $u \to C_v(u) \equiv \partial C(u,v)/\partial v$ and $v \to C_u(v) \equiv \partial C(u,v)/\partial u$ are well-defined and nondecreasing almost everywhere on $\mathbf{I}$.

2.3 Frechet Bounds

Definition 2.3.1 *The copulas $W : \mathbf{I^2} \to \mathbf{I}$ and $M : \mathbf{I^2} \to \mathbf{I}$ are given by*

$$W(u,v) = \min(u,v)$$

and

$$M(u,v) = max(u+v-1,0).$$

Both W and M denote perfect dependence but in two completely different ways. For every copula C and every $(u,v) \in \mathbf{I^2}$

$$W(u,v) \geq C(u,v) \geq M(u,v). \tag{2.3}$$

Inequality (2.3) is the copula version of the Frechet-Hoeffding inequality, which refer to M as the Frechet-Hoeffding upper bound and W as the Frechet-Hoeffding lower bound, see Figure 2.3.

2.4 Dependence Structure

We begin with some ”positive” and ”negative” dependence properties: positive dependence properties expressing the notion that ”large” (or ”small”) values of the random variables tend to occur together, and negative dependence properties expressing the notion that ”large” values of one variables tend to occur with ”small” values of the other.

Definition 2.4.1 *The copula $\Pi : \mathbf{I^2} \to \mathbf{I}$ is given by*

$$\Pi(u,v) = u\,v.$$

Definition 2.4.2 *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 equivalently

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

Negative quadrant dependence (NQD) *is defined analogously by reversing the inequalities in (2.4) and (2.5).*

If X and Y have joint distribution function H, with continuous marginal distributions F and G, respectively, and copula C, and (2.4) holds i.e.

$$H(x, y) \geq F(x)G(y)$$

for all (x, y), then

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)) \geq F(F^{-1}(u))G(G^{-1}(v)) = uv = \Pi(u, v)$$

for all (u,v)=(F(x),G(y)), i.e.

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

This proves that the Π copula is the separator of PQD and NQD.

For copulas, if (2.6) holds for all $u, v \in \mathbf{I}$, the copula is called a *PQD copula*, and an *NQD copula* is defined analogously. Figure 2.3 shows a relation between PQD and NQD copulas.

The concept of tail dependence is a way to describe the amount of extremal value dependence. It is method to measure strength of positive tail dependence. Here, copula functions may be used to compute and investigate tail dependence assessing the evidence of simultaneous booms and crashes on different markets.

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

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

and the upper tail dependence *by*

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

It can be verified that $0 \leq \lambda_U, \lambda_L \leq 1$. When $\lambda_U \approx 1$ or $\lambda_L \approx 1$, there is a strong tail dependence.

2.5 Survival Copulas

In many application, the random variables of interest represent the lifetimes of individuals or objects in some population. The probability of an individual living or surviving beyond time x is given by the survival function (or survivor function, or reliability function).

For a pair (X, Y) of random variables with a joint distribution function H, the joint *survival function* is given by $\overline{H}(x, y) = P[X > x, Y > y]$. The margins of $\overline{H}$ are the univariate survival function $\overline{F}$ and $\overline{G}$, respectively. Then we have

$$\overline{H}(x, y) = 1 - F(x) - G(y) + H(x, y) = \overline{F}(x) + \overline{G}(y) - 1 + C(1 - \overline{F}(x), 1 - \overline{G}(y)),$$

so we define:

Definition 2.5.1 *A bivariate copula* $C_{\text{survival}} : \mathbf{I^2} \to \mathbf{I}$ *is called the* survival copula *of a copula* C *if*

$$C_{\text{survival}}(u, v) = u + v - 1 + C(1 - u, 1 - v). \tag{2.9}$$

It can easily be verified that C_{survival} is a copula if C is a copula. The survival copula switches upper and lower tail dependence, see figure 2.4.

The density of the survival copula c_{survival} and the density of the original copula c are related by

$$c_{\text{survival}}(u, v) = c(1 - u, 1 - v).$$

Hence they are mirror images about $(u, v) = (1/2, 1/2)$

For example, if a copula features positive upper tail dependence means, then the probability of both variables being in the upper tail is relatively high. And then its survival copula, its mirror image, has positive lower tail dependence, so that the probability of both variables being in lower tails is high.

2.6 Mixture Copula

Discrete mixture models, see Hu (2004), arise in the theory of reliability when individuals belong to one of n distinct distributions with certain proportions.

Definition 2.6.1 *Let* $C_1^{\alpha_1}, \ldots, C_n^{\alpha_1}$ *be copulas with parameters* $\alpha_1, \ldots, \alpha_n$*, and* $\beta_1, \ldots, \beta_n \geq 0$ *numbers such that* $\beta_1 + \beta_2 + ... + \beta_n = 1$*. A* mixture copula *is given by*

$$C_{\text{mixture}}(u, v) = \beta_1 C_1^{\alpha_1}(u, v) + ... + \beta_n C_n^{\alpha_n}(u, v). \tag{2.10}$$

Mixture models may be used to obtain more versatile copula models, for example, allowing 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 embodied in two different groups of parameters- the *association parameters* α and the *weight parameters* β. The association parameters are parameters in each copula that control the degree of dependence, while the weight parameters reflects the shape of the dependence.

2.7 Empirical Copulas

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

Definition 2.7.1 *Let* $(x_k, y_k)_{k=1}^n$ *denote a sample of size* n *from a continuous bivariate distribution. The* empirical copula *is the function* $\hat{C}_{\text{emp}}$ *given by**

$$\hat{C}_{\text{emp}}(u,v) = \frac{\sharp\{(x_k,y_k) : F_X(x_k) \le u, F_Y(y_k) \le v\}}{n},$$

and the empirical copula density *function* $\hat{c}_{\text{emp}}$ *is given by*

$$\hat{c}_{\text{emp}}(u,v) = \frac{1}{n}\sum_{k=1}^{n} \delta(u - F_X(x_k), v - F_Y(y_k)).$$

*$\sharp$ denotes the number of elements of a set.

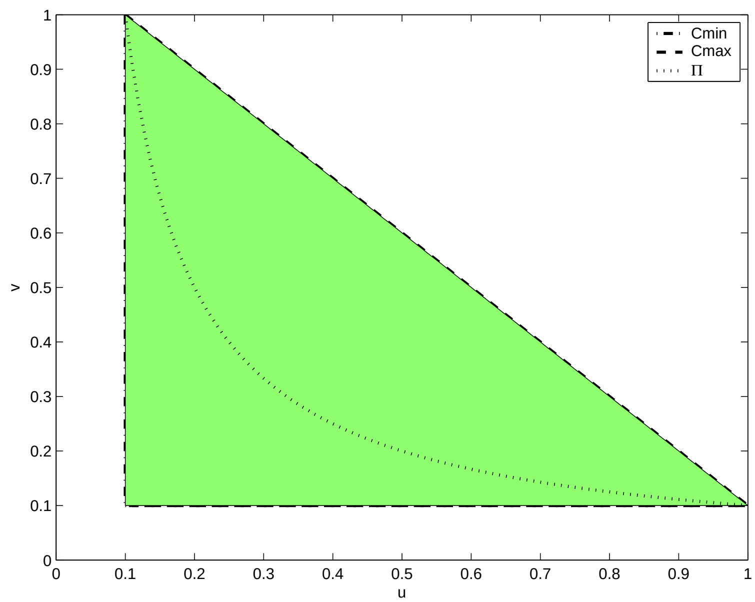

Figure 2.3 Frechet Hoeffding bounds

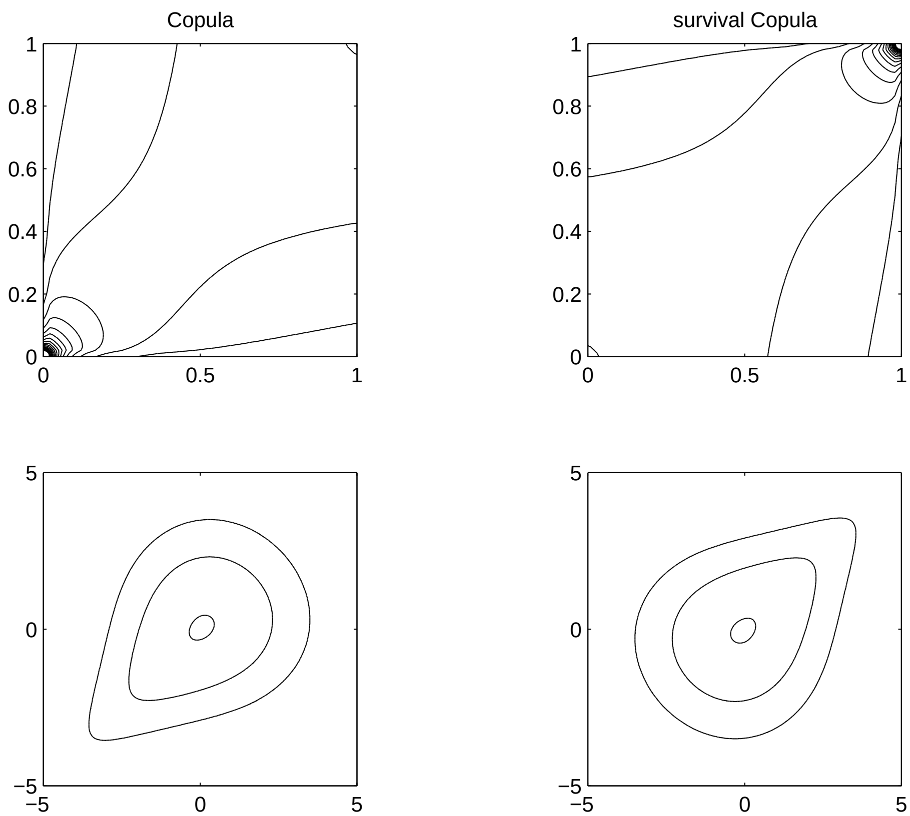

Figure 2.4 Upper figures display the density of a copula and corresponding survival copula density, lower figures display density of the corresponding distribution that has this copula and standard normal marginals.

Chapter 3
Examples of Copulas

The copula W is called *comonotonic* copula since it describes perfect positive dependence, and M is called *countermonotonic* since it describes perfect negative dependence, see Figure 3.1.

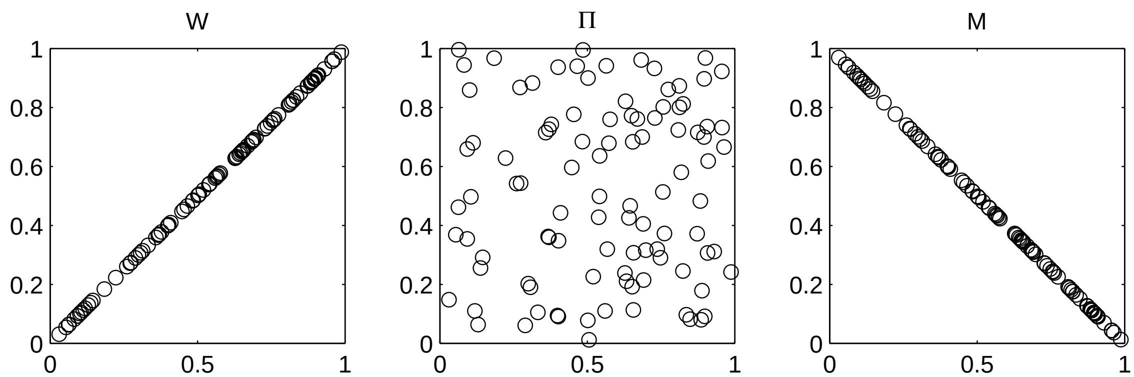

Figure 3.1 100 samples generated from W, Π and M copulas.

Table 3.1 shows that W has maximum upper and lower tail dependence, whereas Π and M both have zero upper and lower tail dependence.

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

3.1 Archimedean Copulas

In this section, we concentrate on an important class of copulas called *Archimedean copulas.* If $C(u_1, u_2, ..., u_n) = \phi(\sum_{i=1}^{n} \phi^{-1}(u_i))$ with generator ϕ then the copula is called archimedean. These copulas allow for a great variety of dependence struc-

	W	Π	M
λ_U	1	0	0
λ_L	1	0	0

Table 3.1 Tail dependence of copulas W, Π and M.

tures. They have closed form expressions and they are not derived from multivariate distribution using Sklar's Theorem. They find an application for a number of reasons:

1. The ease with which they can be constructed;
2. the great variety of facilities of copulas which belong to this class;
3. the many nice properties possessed by the members of this class.

For an account of this history, see Schweizer (1991) and the references cited therein.

Gaussian Copula One of the most frequently used copulas, especially for financial modelling, is the bivariate *Gaussian copula* C_{Gauss}. It is defined by

$$C^{\rho}_{\text{Gauss}}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \exp\left\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right\}\frac{dx\,dy}{2\pi\sqrt{1-\rho^2}} = \Phi_{\rho}(\Phi^{-1}(u),\Phi^{-1}(v)). \tag{3.1}$$

Here Φ^{-1} is the inverse probability distribution function of the standard univariate Gaussian distribution, while Φ_{ρ} is the joint distribution function of a standard bivariate Gaussian with the correlation coefficient ρ, which is the only parameter of the Gaussian copula ($-1 < \rho < 1$).

Figure 3.2 clearly displays that a Gaussian dependence structure is symmetric. Two variates of a stock market with a Gaussian copula dependence structure implies that the variates are equally likely to boom together as to crash together.

A Gaussian dependence structure with $\rho > 0$ means that the variates are positive quadrant dependent, analogous if $\rho < 0$ the variates are negative quadrant dependent.

If $\rho < 0$ in the example of market returns, it implies that it is a higher probability for the variates to move opposite ways, which means that if one variate boom, the other variate exhibits higher probability to crash.

The Gaussian copula's variates are only multivariate normal distributed if its marginal distributions are normal.

We can check that the tail dependence coefficients are $\lambda_U = 0$ and $\lambda_L = 0$. Seebined with Figure 3.2, we defined the tail of Gaussian copula is normal tail.

AMH Copula The bivariate *AMH copula* C_{AMH} is defined as

$$C^{\alpha}_{\text{AMH}}(u,v) = \frac{u\,v}{1-\alpha(1-u)(1-v)}. \tag{3.2}$$

The only parameter of the AMH copula is the parameter α, $0 < \alpha < 1$.

Figure 3.3 clearly displays that a AMH dependence structure is asymmetric and the lower tail is heavier than upper tail. Moreover, its tail dependence coefficients $\lambda_U = \lambda_L = 0$, so we conclude the its upper tail is light tail and its lower tail is light tail or normal tail.

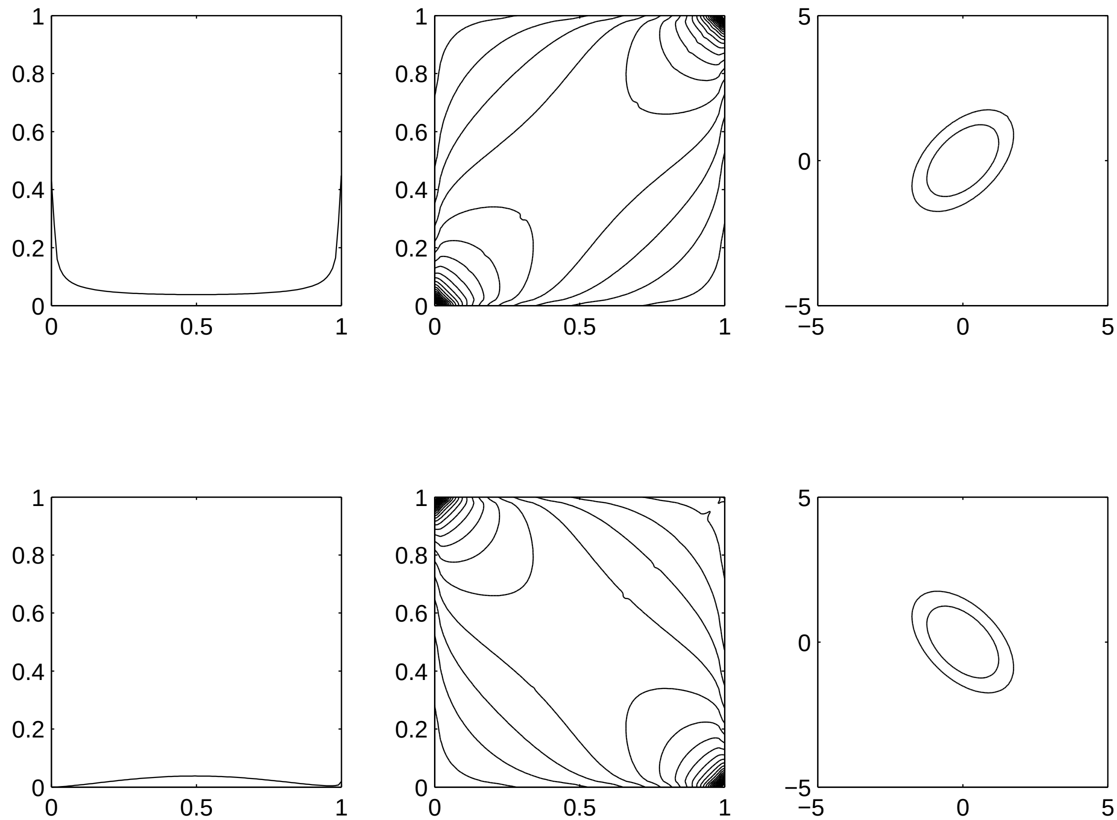

Figure 3.2 First row is the density of Gaussian copula with $\rho = 0.5$ and the second row is the density of Gaussian copula with $\rho = -0.5$. First column displays the cross section of the density on diagonal $u = v$, second column displays copula density function and third column displays joint distribution if marginal distributions are normal.

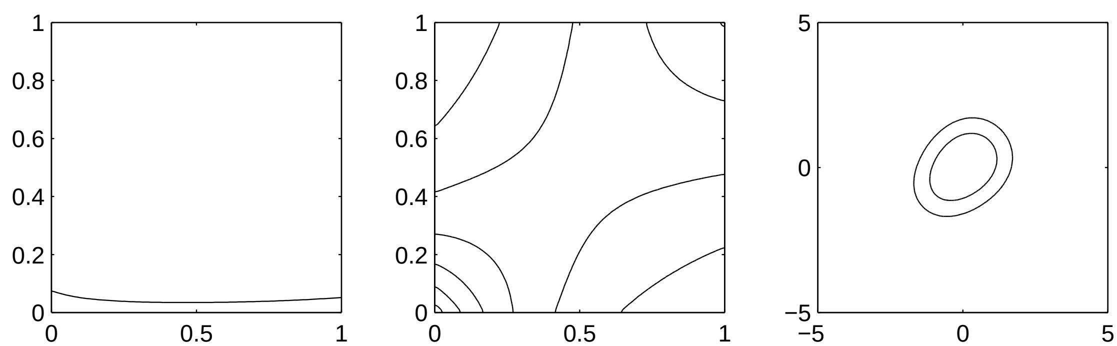

Figure 3.3 First column displays the cross section of the density of $C^{\alpha}_{\mathrm{AMH}}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution when marginal distributions are standard normal.

Frank Copula The bivariate *Frank copula* C_{Frank} is defined as

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

The only parameter of the Frank copula is the parameter α, where $0 < \alpha < 1$ or $\alpha > 1$.

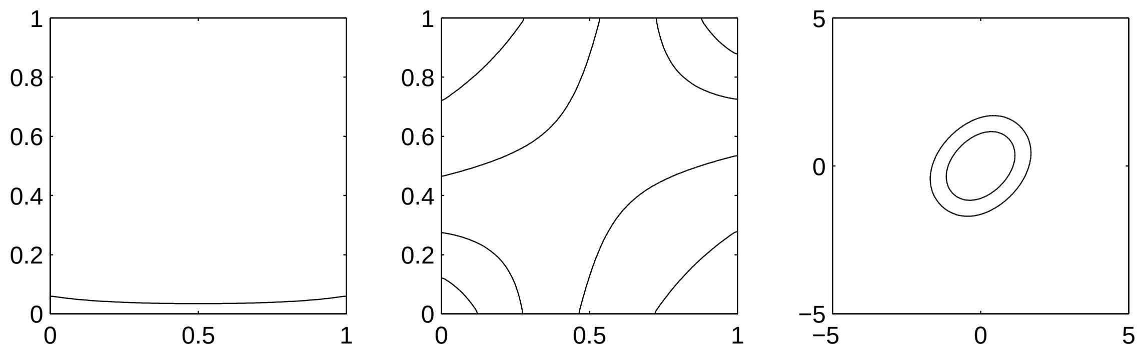

Figure 3.4 First column displays the cross section of the density of $C_{\text{Frank}}^{\alpha}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution when marginal distributions are standard normal.

Figure 3.4 displays that a Frank copula's dependence structure is symmetric and by Appendix I, its tail dependence coefficients are $\lambda_U = \lambda_L = 0$. Compared with a Gaussian copula, the Frank copula has more probability in the middle region, so that the tails must be lighter.

Gumbel Copula The bivariate *Gumbel copula* C_{Gumbel} is defined as

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

The only parameter of the Gumbel copula is the parameter α, where $\alpha \geq 1$. For $\alpha = 1$, expression (3.4) reduces to $C_{\text{Gumbel}}^{1}(u,v) = \Pi(u,v)$, the independent copula.

From Figure 3.5, a Gumbel dependence structure is asymmetric and the upper tail is heavier than the lower tail. A Gumbel copula implies that two markets are more likely to boom together than to crash together. And its tail dependence coefficients are $\lambda_U = 2 - 2^{1/\alpha}$, $\lambda_L = 0$, so that its upper tail is heavy, and its lower tail light or normal. The expression for λ_U also shows that the larger is α, the heavier is the upper tail.

Joe Copula The bivariate *Joe copula* C_{Joe} is defined as

$$C^{\alpha}_{\text{Joe}}(u,v) = 1 - ((1-u)^{\alpha} + (1-v)^{\alpha} - (1-u)^{\alpha}(1-v)^{\alpha})^{1/\alpha}. \tag{3.5}$$

The only parameter of the Joe copula is the parameter α, $\alpha > 1$.

Clearly a Joe copula's dependence structure, see Figure 3.6, is asymmetric and its tail dependence coefficients are $\lambda_U = 2 - 2^{1/\alpha}$ and $\lambda_L = 0$.

Even though the tail dependence of C^{α}_{Joe} is equal to $C^{\alpha}_{\text{Gumbel}}$, it clearly does not mean equal dependence structures apart from the tails. For example, the Joe copula does feature a much lighter lower tail.

Cook-Johnson Copula The bivariate *Cook-Johnson (CJ)* copula C_{CJ} is defined as

$$C^{\alpha}_{\text{CJ}}(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}. \tag{3.6}$$

The only parameter of the CJ copula is the parameter α, $\alpha > 0$.

The Cook-Johnson copula's dependence structure is also asymmetric and its tail dependence coefficients are $\lambda_U = 0$ and $\lambda_L = 2^{-1/\alpha}$, see Figure 3.7. If two variates of the stock market follow the CJ copula, then they feature a larger probability of simultaneous crashing than simultaneous booming.

BB1 Copula The bivariate *BB1 copula* C_{BB1} is defined as

$$C^{\alpha_1,\alpha_2}_{\text{BB1}}(u,v) = (1 + ((u^{-\alpha_1} - 1)^{\alpha_2} + (v^{-\alpha_1} - 1)^{\alpha_2})^{1/\alpha_2})^{-1/\alpha_1}. \tag{3.7}$$

The parameters of the BB1 copula are $\alpha_1 > 0$ and $\alpha_2 \geq 1$.

The BB1 copula's dependence structure is asymmetric, and the copula emphasizes both tails. The density function of the copula is concentrated closly to the line $u = v$, see Figure 3.8. The tail dependence coefficients are $\lambda_U = 2 - 2^{\frac{1}{\alpha_2}}$ and $\lambda_L = 2^{-1/(\alpha_1\alpha_2)}$.

BB6 Copula The bivariate *BB6 copula* C_{BB6} is defined as

$$C^{\alpha_1,\alpha_2}_{\text{BB6}}(u,v) = 1 - (1 - e^{-((-\log(1-(1-u)^{\alpha_1}))^{\alpha_2} + (-\log(1-(1-v)^{\alpha_1}))^{\alpha_2})^{1/\alpha_2}})^{\alpha_1}. \tag{3.8}$$

The parameters of the BB6 copula are $\alpha_1 > 0$ and $\alpha_2 \geq 1$.

The BB6 copula's dependence structure is asymmetric and looks like that of BB1, with the density concentrated close to the line $u = v$, see Figure 3.9. However, its upper tail is heavier than its lower tail, and both these tails are heavy.

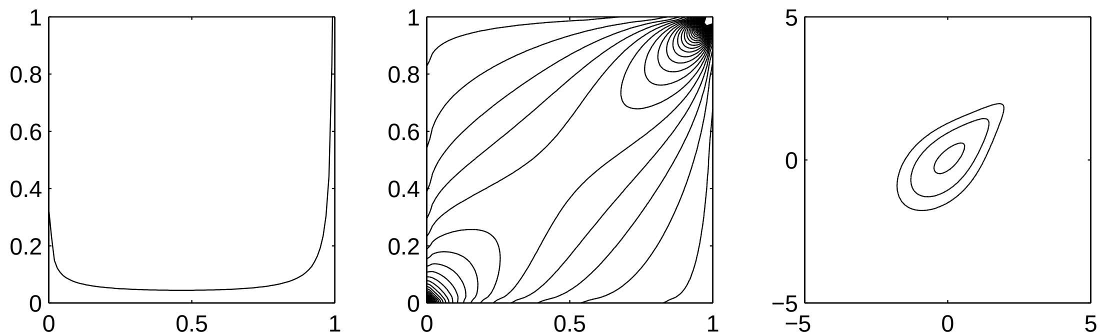

Figure 3.5 First column displays the cross section of the density of $C^{\alpha}_{\mathrm{Gumbel}}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution when marginal distributions are standard normal.

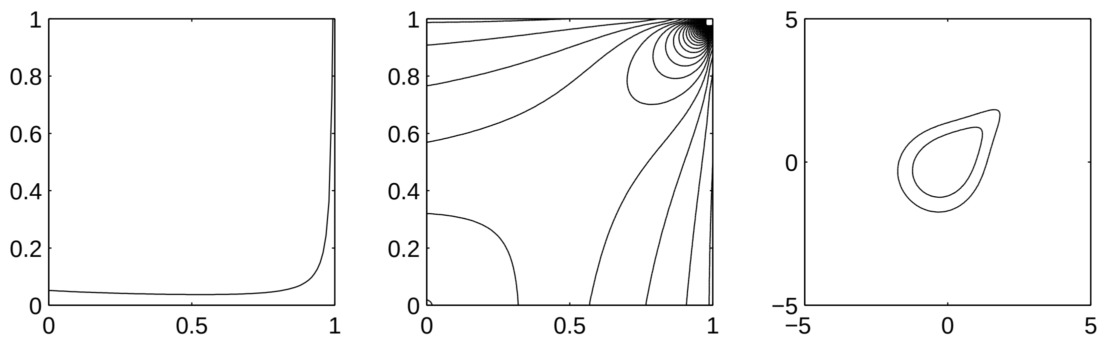

Figure 3.6 First column displays the cross section of the density of $C^{\alpha}_{\mathrm{Joe}}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution if marginal distributions are standard normal.

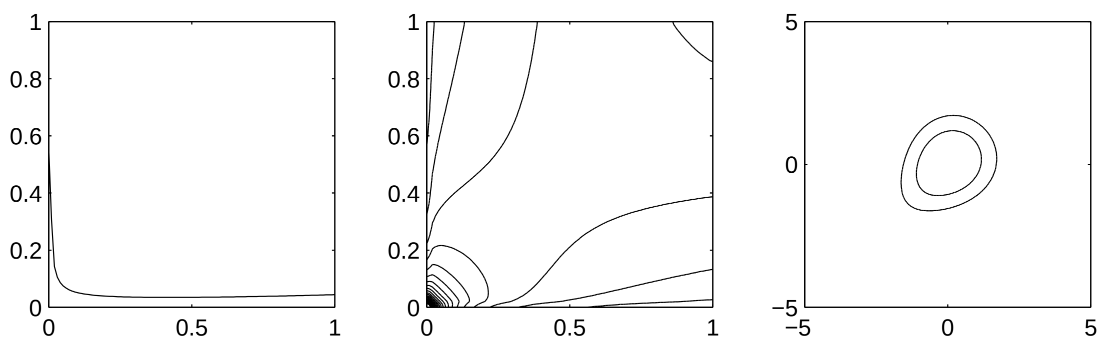

Figure 3.7 First column displays the cross section of the density of C^{α}_{CJ} on diagonal $u = v$, second column displays copula density function and third column displays joint distribution if marginal distributions are standard normal.

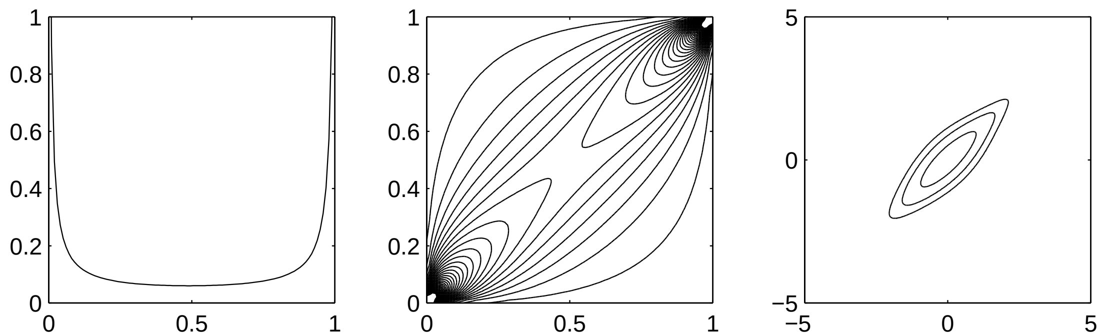

Figure 3.8 First column displays the cross section of the density of $C^{\alpha_1,\alpha_2}_{\mathrm{BB1}}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution if marginal distributions are standard normal.

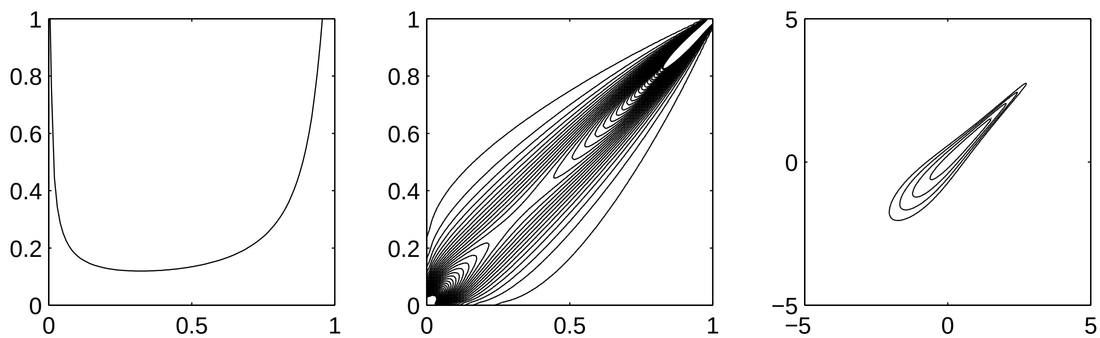

Figure 3.9 First column displays the cross section of the density of $C^{\alpha_1,\alpha_2}_{\mathrm{BB6}}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution if marginal distributions are standard normal.

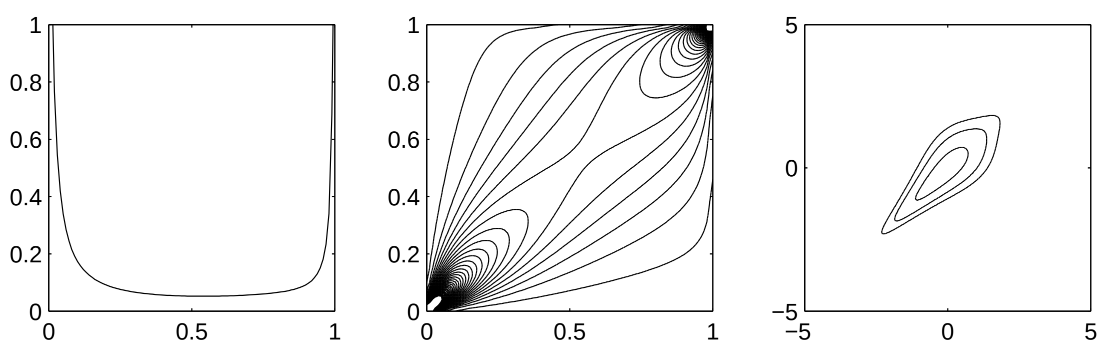

Figure 3.10 First column displays the cross section of the density of $C^{\alpha_1,\alpha_2}_{\mathrm{BB7}}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution if marginal distributions are standard normal.

BB7 Copula The bivariate *BB7 copula* C_{BB7} is defined as

$$C_{\text{BB4}}^{\alpha_1,\alpha 2}(u,v) = 1 - (1 - ((1-(1-u)^{\alpha_1})^{-\alpha_2} + (1-(1-v)^{\alpha_1})^{-\alpha_2} - 1)^{-1/\alpha_2})^{1/\alpha_1}. \quad (3.9)$$

The parameters of the BB7 copula are $\alpha_1 > 0$ and $\alpha_2 \geq 1$.

The BB7 copula's dependence structure is almost symmetric and its density concentrated at the center, but the tails are heavier than for the Gaussian copula, see Figure 3.10.

3.2 Extreme Value Copulas

Another important class of copulas is the extreme value class: A copula is said to be an *extreme value copula (EV)* if for all $t > 0$ the scaling property. $C(u^t, v^t) = (C(u,v))^t$ holds $\forall\, u, v \in \mathbf{I}$.

EV copulas are max-stable, meaning that, if (X_1, Y_1), (X_2, Y_2), ...(X_n, Y_n) are independent identically distributed (i.i.d.) random pairs from an EV copula C and $M_n = \max\{X_1, X_2, ...X_n\}$ and $N_n = \max\{Y_1, Y_2, ...Y_n\}$, then the copula for (M_n, N_n) is also C. The EV copulas can be represented in the form:

$$C_{\text{EV}}(u,v) = e^{\log(uv) A(\log(u)\log(v)/\log(uv))}, \quad (3.10)$$

where the function A is called the *dependence function.*

Galambos Copula The *Galambos copula* has the following form:

$$C_{\text{Galambos}}^{\alpha}(u,v) = uv e^{((-log(u))^{-\alpha} + (-log(v))^{-\alpha})^{-1/\alpha}}. \quad (3.11)$$

The only parameter of the Galambos copula is the parameter $\alpha \geq 0$.

Figure 3.11 shows that a Galambos copula's dependence structure is asymmetric. The tail dependence coefficients are $\lambda_U = 2^{-1/\alpha}$ and $\lambda_L = 0$.

BB5 Copula The *BB5 copula* is a two-parameter extension of the Gumbel copula and has the following form:

$$C_{\text{BB5}}^{\alpha_1,\alpha_2}(u,v) = e^{-(-\log(u))^{-\alpha_1} - (-\log(v))^{-\alpha_1} + ((-\log(u))^{-\alpha_1\alpha_2} + (-\log(v))^{-\alpha_1\alpha_2}))^{-1/\alpha_2}}. \quad (3.12)$$

The parameters of the BB5 copula are $\alpha_1 \geq 0$ and $\alpha_2 > 1$.

Clearly a BB5 copula's dependence structure is also asymmetric. As illustrated in Figure 3.12, a BB5 copula implies that two markets are more likely to boom together than to crash together and its tails are heavy.

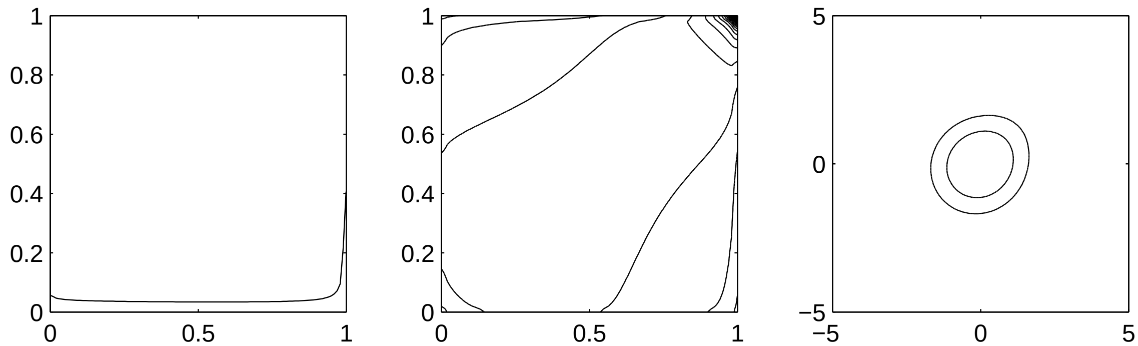

Figure 3.11 First column displays the cross section of the density of $C^{\alpha}_{\text{Galambos}}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution if marginal distributions are standard normal.

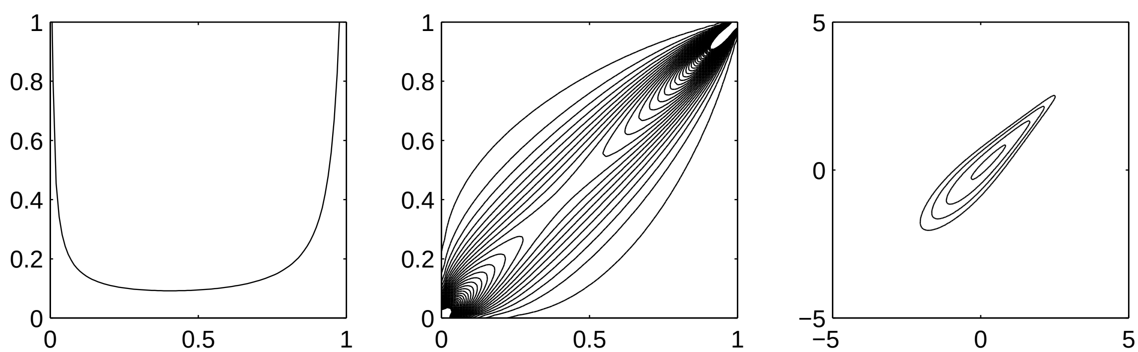

Figure 3.12 First column displays the cross section of the density of $C^{\alpha_1,\alpha_2}_{\text{BB5}}$ on diagonal $u = v$, second column displays copula density function and third column displays joint distribution if marginal distributions are standard normal.

3.3 The Frechet Family

Definition 3.3.1 *Let* $\alpha, \beta \in \mathbf{I}$ *with* $\alpha + \beta \leq 1$*, and set*

$$C^{\alpha,\beta}(u,v) = \beta M(u,v) + (1-\alpha-\beta)\Pi(u,v) + \alpha W(u,v): \tag{3.13}$$

This comprehensive two-parameter copula is called a Frechet copulas.

The parameters α, β are linked to non-parametric dependence measures by particularly simple analytical formulas. For example, the correlation of a Frechet copula is

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

A Frechet copula with $\beta = 0$ is a PQD copula, and when $\alpha = 0$, it is a NQD copula.

Chapter 4
Correlation

Correlation is a statistical technique which can show whether and how strongly pairs of stochastic variables are related. Correlation is essentially founded on the assumption of multivariate normally distributed returns, in order to adequatly describe dependencies. Still, correlation analysis feature as an important tool to measure dependencies on for example, returns in stock markets.

We begin with considering pairs of real valued random variables X and Y with finite variances.

The linear correlation coefficient between X and Y is

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}},$$

where $\mathrm{Cov}(X,Y)$ is the covariance between X and Y, $\mathrm{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])]$, and $\mathrm{Var}(X)$ denotes the variance of X.

Correlation is a measure of linear dependence. In the case of perfect linear dependence, i.e.

$$Y = aX + b \; a.s.,$$

or $P[Y = aX + b] = 1$, where $a \neq 0$, then $\rho(X,Y) = \mathrm{sign}(a)$ is -1 or 1.

The correlation coefficient may take on any value between positive and negative one,

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

The sign of the correlation coefficient defines the direction of the relationship, either positive or negative. A positive correlation coefficient, in the example of market returns, implies that it is a higher probability for the variates to move same ways, which means that if one variate booms, it is a high probability that the other variate would also boom.

If two random variables X and Y are jointly normal distributed with covariance $\mathrm{Cov}(X,Y)$, then all dependencies between these two variates are captured in the covariance. This means that X is independent from $Y - \mathrm{Cov}(X,Y)/\mathrm{Var}(Y)X$, because

$$\mathrm{Cov}(X, Y - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}X) = \mathrm{Cov}(X,Y) - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}\mathrm{Cov}(X,X) = 0.$$

If two real world random variables $\hat{X}$ and $\hat{Y}$ have a dependence structure that can be described by correlation analysis, then $\hat{X}$ and $\hat{Y} - \mathrm{Cov}(\hat{X},\hat{Y})/\mathrm{Var}(\hat{Y})\hat{X}$ should be independent, by the above argument, and the Π copula should fit its empirical copula perfectly.

However, the correlation, as well as being one of the most ubiquitous concepts in modern finance, is also one the of the most misunderstood concepts.

Assume that the copula function of a pair of random variables X and Y is known. Then their covariance is given by

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

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

$$\mathrm{Cov}(X,Y) = \int_0^1\int_0^1 F_X^{-1}(u)F_Y^{-1}(v)\left(c(u,v) - 1\right) dudv. \tag{4.1}$$

It is now clear that all information of the copula is not captured by the covariance. Copulas define the complete dependence structure while covariance only is a measure of linear dependence: One cannot compute c given $\mathrm{Cov}(X,Y)$ in (4.1)!

Of course, by definition of the Π copula, we have $\partial^2\,\Pi(u,v)/\partial u\partial v = 1$. Then by (4.1), the covariance is zero, as well as all other relations between the two random variables involved.

Chapter 5
Generation of Random Number Using Copula Models

In Chapter 2, we presented the definition of copulas, their most important properties and several classes of copulas. So now we are ready to generate random numbers. The strategy is to give a general guide on how to generate pairs of random variables whose dependence structure is defined by a copula. Helpful for understanding, we exemplify with a Gumbel copula which is discussed in more detail.

5.1 The General Method

Assume that all parameters of the joint distribution is known. The joint density function is bounded, $f(x, y) \leq M$, and random variables come from a closed box. Then a random variate (X, Y) from this dependence structure can be created as follows:

1. Generate two uniform distributed random variables ξ, η from the box;
2. generate a uniform variable, γ from 0 to M;
3. if $f(\xi, \eta) > \gamma$, then accept (ξ, η) to the data set, otherwise go back to 1.

Since $c(u, v)$ may not be bounded, the uniform random variables may have to be transformed. Consider two random variables with normal marginal distributions, then

$$F(x, y) = C(\Phi(x), \Phi(y)).$$

Further, the joint density function is given by

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^2}{\partial x \partial y} C(\Phi(x), \Phi(y)) = c(\Phi(x), \Phi(y))\, \phi(x)\, \phi(y),$$

where $d\Phi(x)/dx = \phi(x)$ and $c(\Phi(x), \Phi(y))\, \phi(x)\, \phi(y) \leq c(0.5, 0.5) \leq M$ is bounded. Now the box is not closed since normal distributed variables can take any value. But assuming that no values lay outside some large perimiter, then all properties of the above generating data model is fulfilled.

The second method, using conditional distributions, works for all copula functions:

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

$$c_v(u) = P[U < u | V = v].$$

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

$$c_v(u) = P[U \leq u \,|V = v] = \int_{-\infty}^{u} \frac{f(x,v)}{f_Y(v)}\, dx = \int_{0}^{u} c(x,v)\, dx = \frac{\partial}{\partial y} C(u,y)|_{y=v} = C_v(u,v), \tag{5.1}$$

where $C_v(u,v)$ is the partial derivative of the copula C. From (2.2), we know that $c_v(u)$ is nondecreasing and exists for all $u \in \mathbf{I}$.

With the result (5.1) at hand, we have the following second method to generate the data, as follows:

1. Generate a uniformly distributed random variable ξ over $\mathbf{I}$;
2. generate $(U|V = \xi)$ from the conditional copula;
3. now $(\xi, (V|U = \xi))$ will be a random variate with the distribution desired.

5.2 Generation Random Numbers According to the Gumbel Copula

In this section, a detailed example is presented. For the sake of simplicity, we assume the parameter α of Gumbel copula equals 1.5, and use the conditonal copula method to generate 500 paired random numbers.

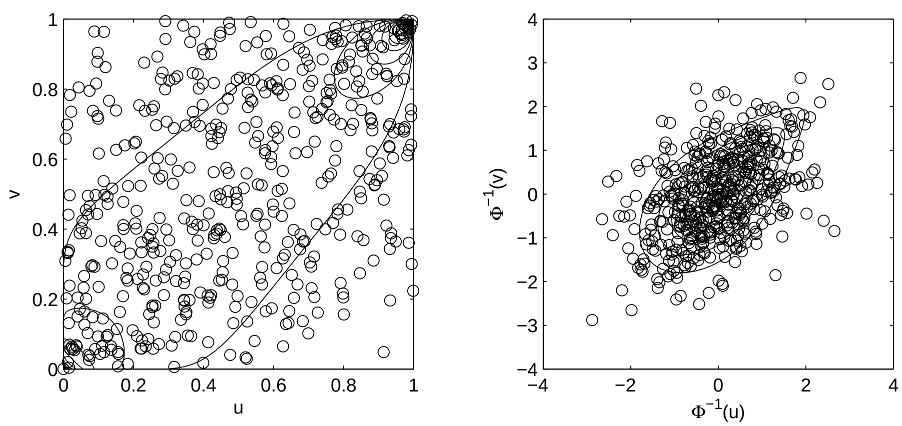

Figure 5.1 First column displays variates with uniform marginal distributions from a Gumbel copula and the copula density, second column displays same variates with standard normal marginal distribution and joint density.

In Figure 5.1, the paired random numbers are displayed. And it is clear that the data sets well fit the Gumbel copula model by comparing to its contour lines.

5.3 Robust Estimation of Significance

The bootstrap method will be used to measure the goodness of fit for the models. The detail method is:

1. Generate pairs (ξ, η) from model of data sets sample size;
2. find empirical copula function from the generated data;
3. calculate the Kuiper distance of the empirical copula to the model copula (see Appendix D);
4. repeat the above procedure 1000 times;
5. plot the empirical distribution function of the observed Kuiper distances.

By checking the empirical distribution function of the Kuiper distance, approximate p-values for the Kuiper distance statistic can be found.

Chapter 6
Investigation

The investigation is diversified into two parts. Firstly the dependence structure with time of a single asset, and secondly the dependence structure of two different assets, which is the major part.

When modelling the price process of an asset on a true market, normally the logreturns of the asset are viewed as i.i.d. data. But how well does this assumption hold? Are the logreturn really independent in time?

According to the distribution of portfolio enterprisers, the investment policy of the fund is reflected in the structure of its investment portfolio. What dependence structure between the real assets are preferable? The value of a fund is the sum of all assets within the fund. So to avoid crashes, the assets should display negative dependence. Meanwhile, to allow maximum profit, the assets should display positive dependence.

6.1 Data

In order to investigate the different dependence structures of assets on real-world markets, long term data sets and short term data sets were gathered.

Long term data were catched from the finance homepage of Yahoo of the variates in the ProFunds Ultra Telecommunications Inv. fund (TCPIX).

Five minutes short term tick data were catched from the OMX homepage of the variates in the SEB Sverigefond 1.

Statistical Presentation of Long Term Data Sets We briefly present the basic statistics of the logreturn* series of the holding of ProFunds Ultra Telecommunications Inv. before investigating the distribution of their dependence. The series are CenturyTel Inc. (CTL), SBC Communications Inc. (SBC) and Alltel Corp. (AT). (Of these, the fund does no longer hold CTL.)

From inspection of Figure 6.2, we see that all three data sets have a non-Gaussian distribution and display heavy tails.

In Table 6.1, we summarize the computations of the first two empirical moments of the logreturn series.

From Figure 6.3, it is clear that the logreturns are time dependent. The data is crash dependent, and further investigation verifies that the variates also displays negative dependence. This means that the data set is not i.i.d. The same property

*We use the notation; $S(t)$ stock price, $X(t)$ logreturn and $A(t)$ devolatilized logreturn.

	Mean	Variance
CTL	$3.032e-04$	$4.318e-04$
SBC	$4.579e-04$	$4.157e-04$
AT	$2.168e-04$	$4.795e-04$

Table 6.1 Statistics for three logreturn series on the period from 03-Oct-96 to 24-Sep-04 with a total of $N = 2000$ data.

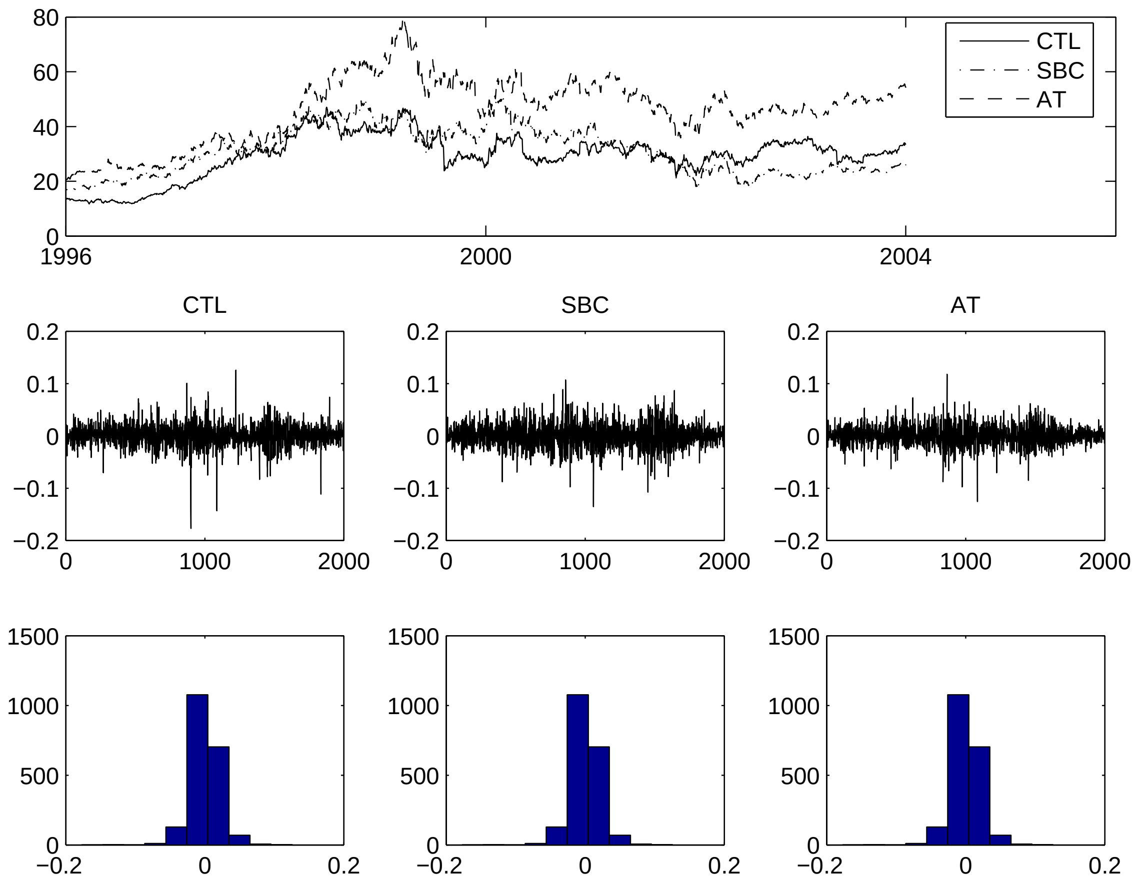

Figure 6.1 At top stocks prices from 03-Oct-96 to 24-Sep-04, in middle logreturn values of same time period and at bottom histogram of the logreturn data.

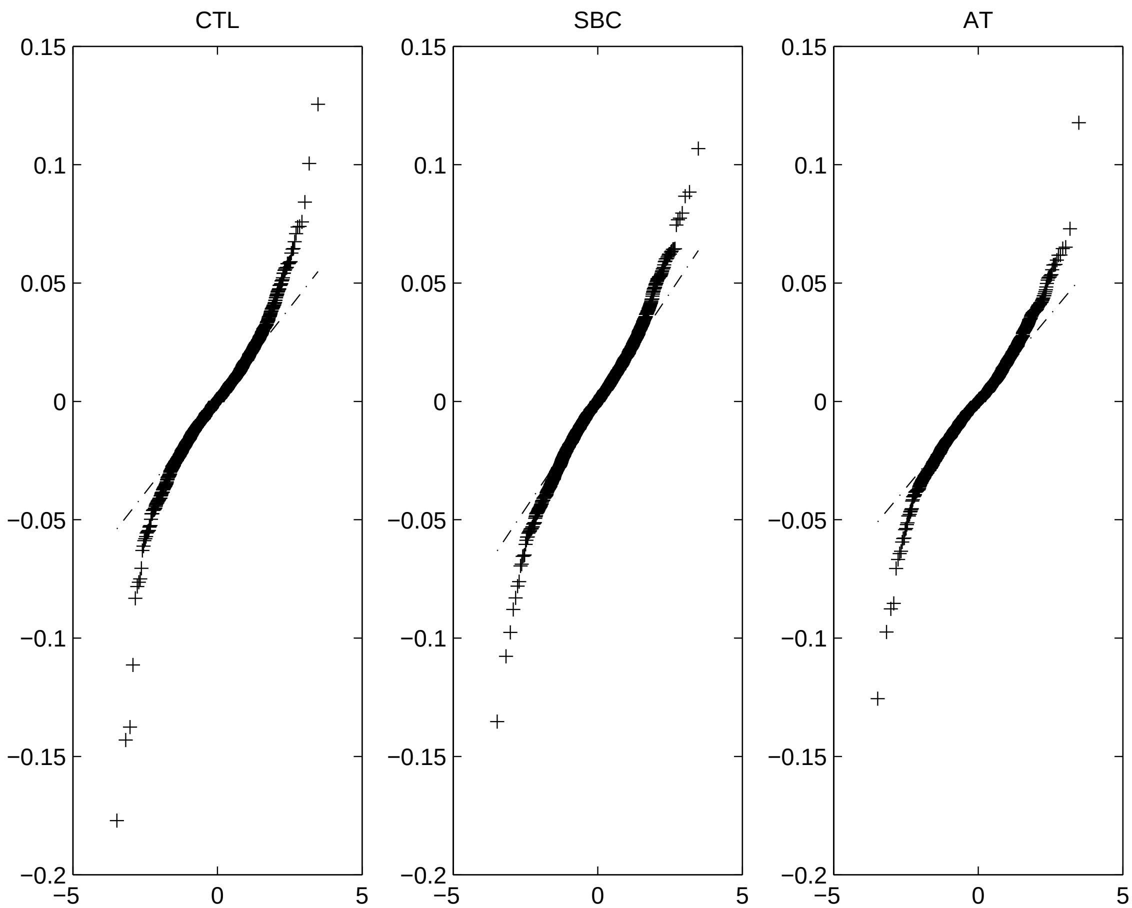

Figure 6.2 Normal quantile-quantile plot of logreturns of data.

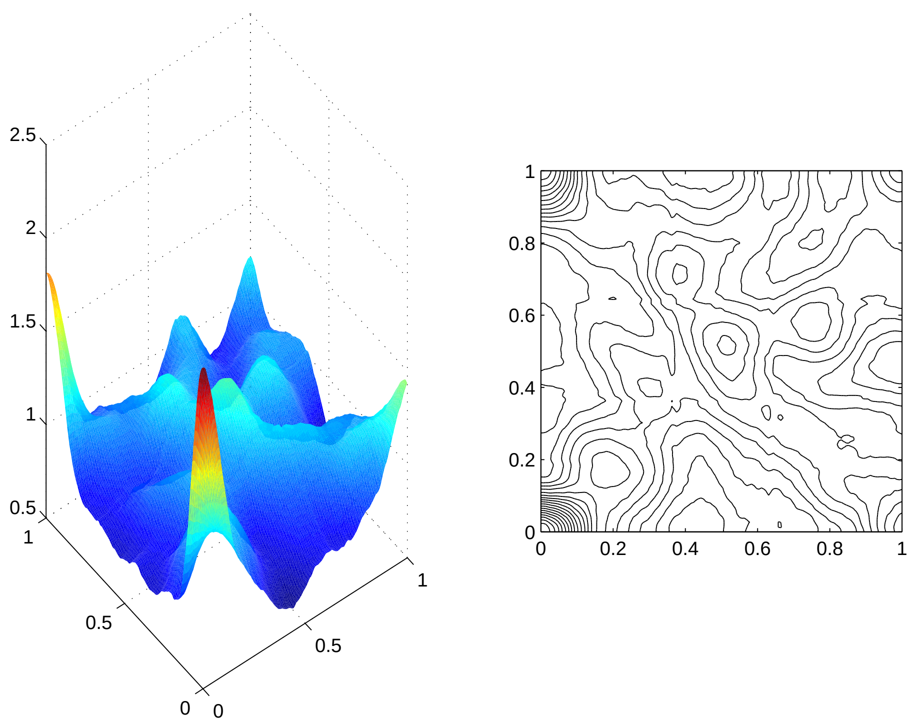

Figure 6.3 Copula density function of time dependence structure for logreturn CTL.

holds for all long term data sets. This means that parameter estimation with standard maximum likelihood method basically is inaccurate.

Filtering of Data By investigating data sets it can be verified that the volatility is not constant, so the Bachelier-Samuelson Black-Scholes model, see Appendix F, is improper. To avoid this property of a changing market we employ a devolatilization of the data set.

Ome may generalize the Bachelier-Samuelson Black-Scholes model by making the volatility time dependent and the noise process a Lévy process L_t, i.e.

$$dS(t) = (\mu + \frac{\sigma_t}{2})S(t)dt \, + \, \sigma_t S(t)\, dL_t. \tag{6.1}$$

The logreturn X_t of the stock price is then, if assumed that σ_t moves slowly compared to L_t,

$$X_t = \log(S_t) - \log(S_{t-\Delta}) = \mu\xi + \sigma_t L_t - \sigma_{t-\Delta} L_{t-\Delta} \approx \mu\xi + \sigma_t(L_t - L_{t-\Delta}) = \mu\Delta + \sigma_t A_t,$$

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

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

By devolatilizing the changing volatility, the data is made independent of market changes. The time dependent volatility is estimated by the Nadaraya-Watson algorithm, see Nadaraya (1964) and Watson (1964), together with Bengtsson and Olsbo (2002) and Drees and Starica (2002).

6.2 Simulation of Copula

The parameters in each copula function control the degree of dependence. In Figure 6.4, we can find how the parameters change, for example, a PQD copula to a NQD copula, or the Π copula to the W copula, and so on.

6.3 Fitting Copulas to Data

We propose two different diagnostics: A numerical method and a graphical method.

Now, we consider distances between fitted copuls and empirical copula, see Appendix D, and the corresponding p-value, see Appendix E, of the three devolatilized logreturn series of CTL, SBC and AT, labeled series A_i, $i \in 1, 2, 3$. First, we look at the dependence structure among A_1, A_2 and A_3. There are higher crash dependence than boom dependence, see Figure 6.5.

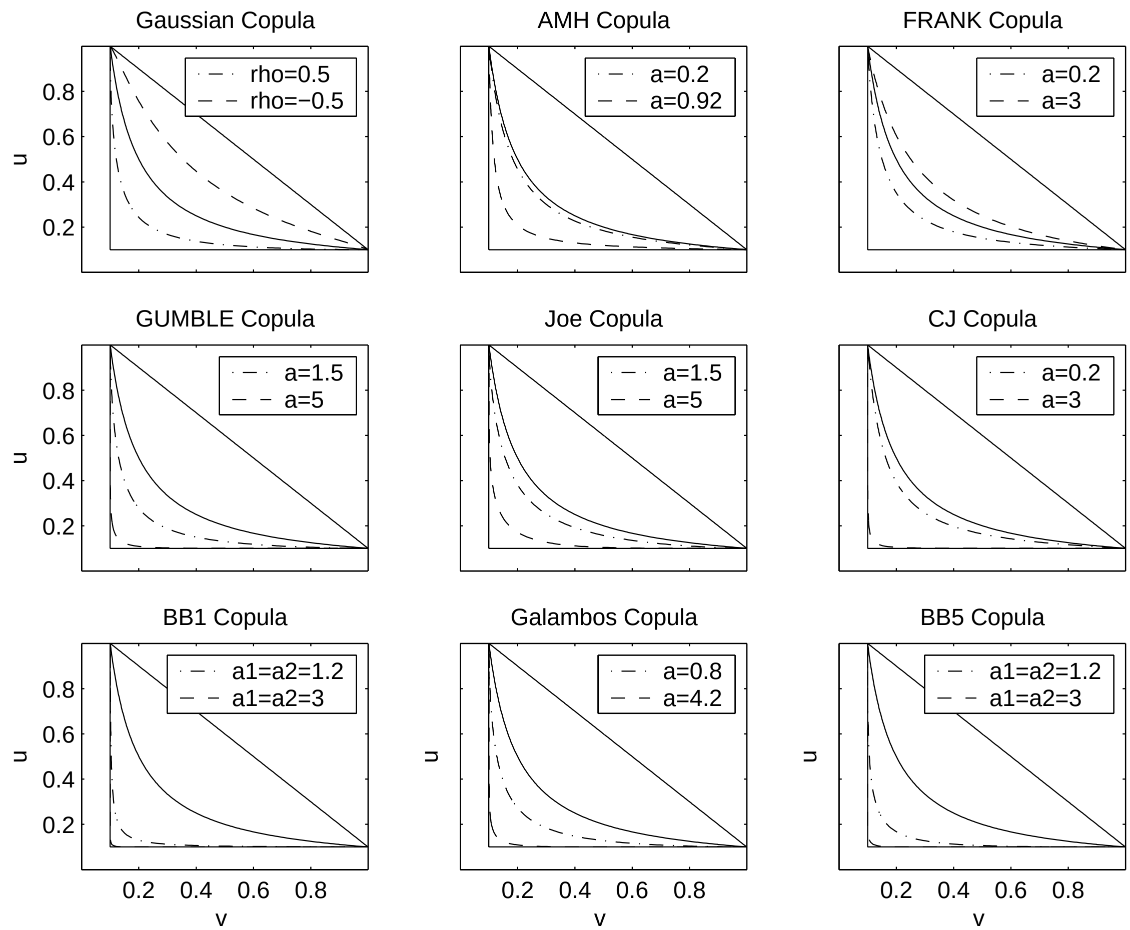

Figure 6.4 Simulation Copulas's PQD and NQD properties by changing parameters.

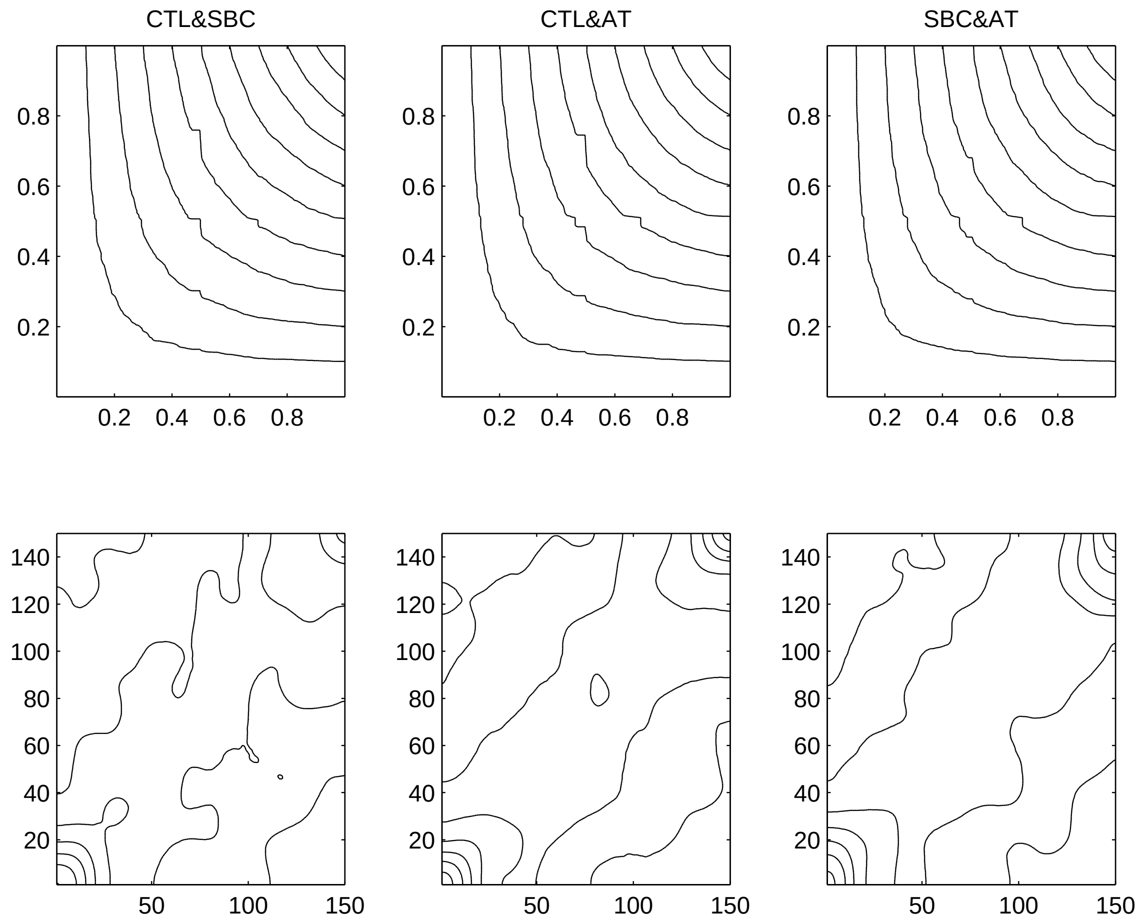

Figure 6.5 Empirical copulas and its density copula functions of dependence structure among CTL, SBC and AT.

Combined with Figure 6.5, it can be seen that there are many sensitive regions in the central part of the copulas. Kolmogorov-Smirnov distance (D.1) and Kuiper distance (D.3) are sensitive to all data, while the Anderson-Darling distance (D.2) emphasizes the tails. We selected to use the Kupier distance (D.3), which considers the greatest deviations upwards as well as downwards.

For the various copulas and each pair of logreturn series, by calculating the minimal Kuiper distance between the empirical copulas and theoretical models, the copula's parameters are found. Appendix B and Figure 6.6 show that Gumbel survival copula is superior to all of the copulas that we investigated, which means there is a lower tail dependence in our data set.

The p-values should be checked to see if the Gumbel survival copula is fit for the data sets. When calculating p-values for the three pairs, $A_1\&A_2$, $A_1\&A_3$ and $A_2\&A_3$, we found that the Gumbel copula is not a good model, because all of them are smaller than 5%.

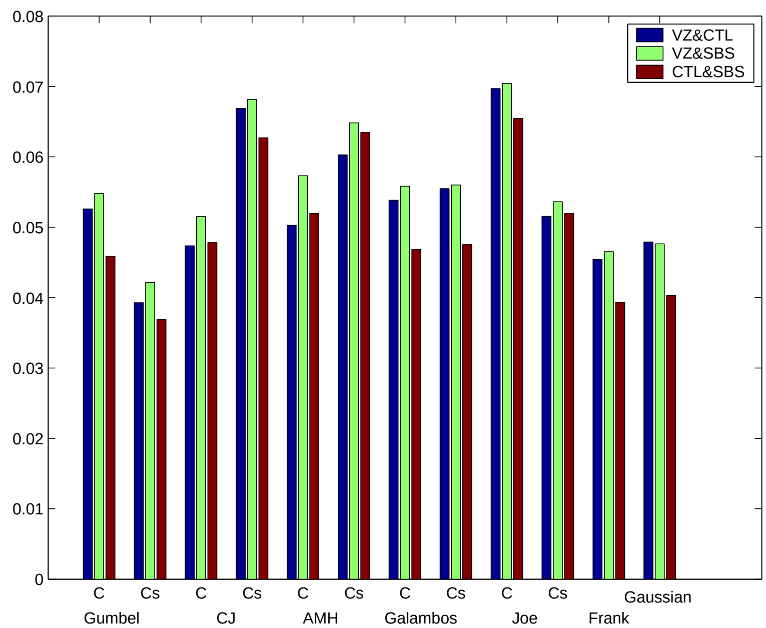

Figure 6.6 Kuiper distances of copulas to the empirical copulas of $A_1\&A_2$, $A_1\&A_3$ and $A_2\&A_3$.

Statistical Presentation of Short Term Data Sets Here, three five-minutes short term tick data sets (Ericson B, Volvo and H&M, see Figure 6.7), catched from the OMX homepage of the variates in the SEB Sverigefond 1 have been investigated. The short term tick data sets are not Gaussian distributed, see Figure 6.8. And the data sets have discrete distributions because the time interval is too short to project exterior events. The short term data sets are not i.i.d as well as the long term data, see Figure 6.10. Moreover, the short term data sets feature a stronger negative time dependence than long term data.

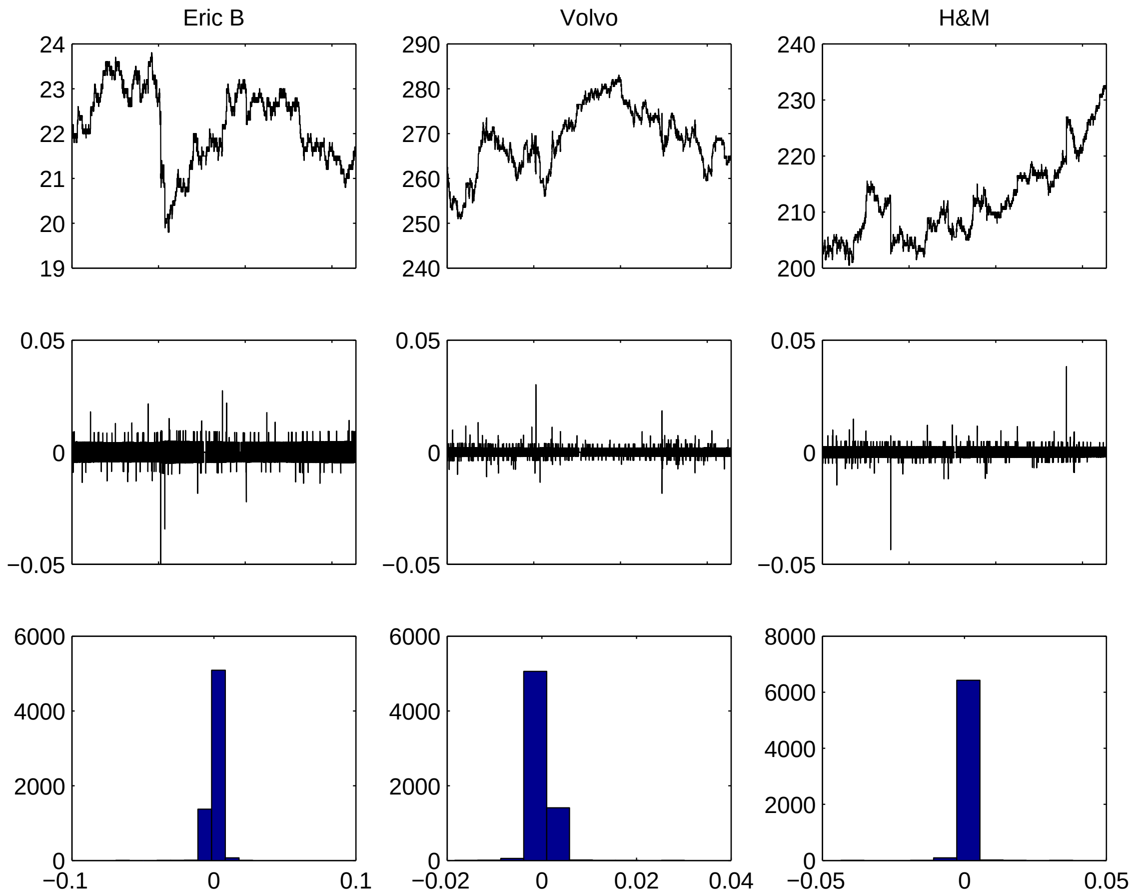

Figure 6.7 First row displays stock prices of Ericson B, Volvo and H&M 02-Oct-04 to 14-Dec-04, second row displays logreturn time series and third row displays histogram of data sets.

By Figure 6.10, the dependence structure is almost the Π copula except for in the tails. This means that the variates are indepdent everywhere except for extreme events.

The distance measure and p-value work here as for the long term data sets. The Joe survival copula is better than all other copulas that we investigated, which means there is a lower tail dependence in the short term data set, see Figure C.1.

We check the p-value to see if the Joe survival copula is fit for the data sets. It

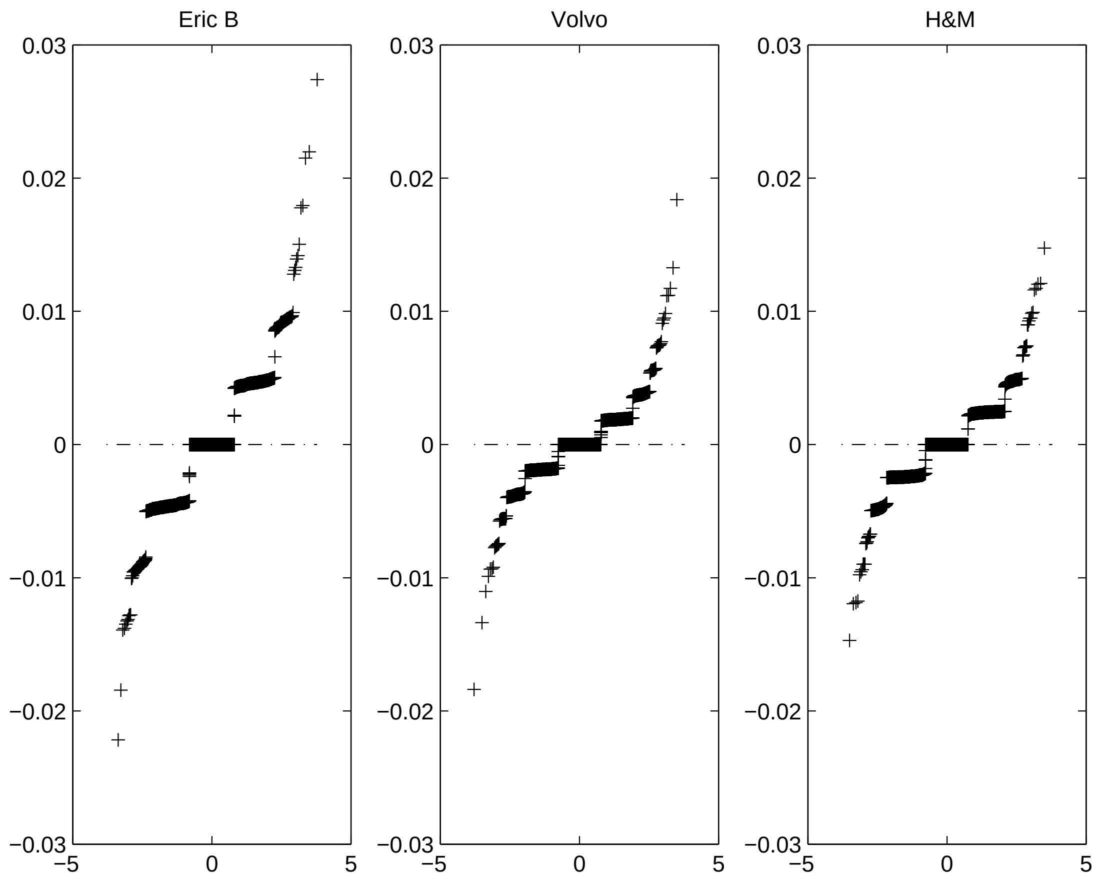

Figure 6.8 Normal quantile-quantile plots for Ericson B, Volvo and H&M.

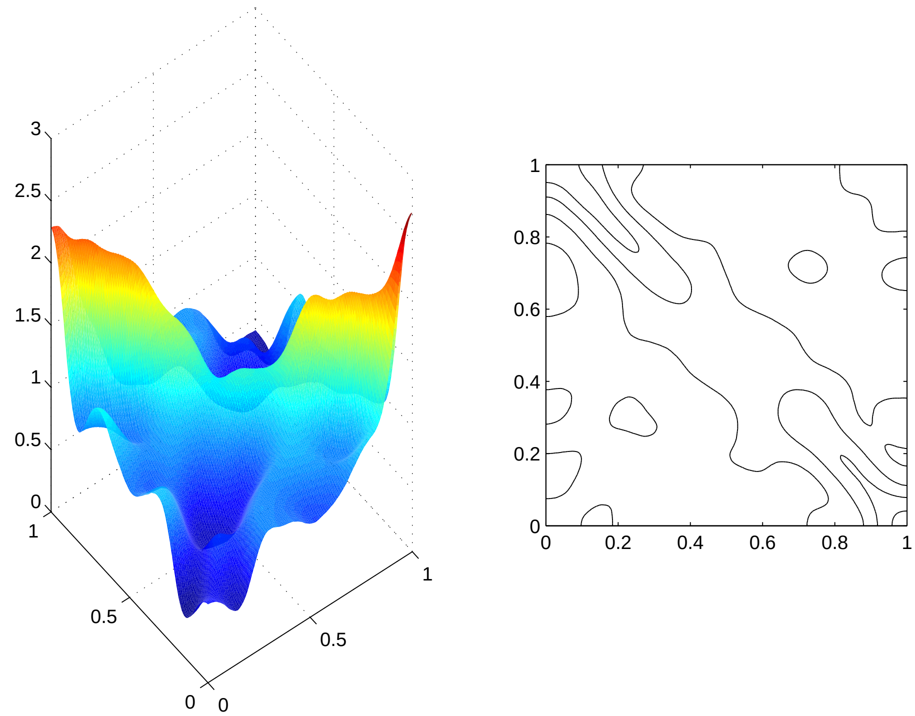

Figure 6.9 Copula density function of time dependence structure for short term data set Ericson B.

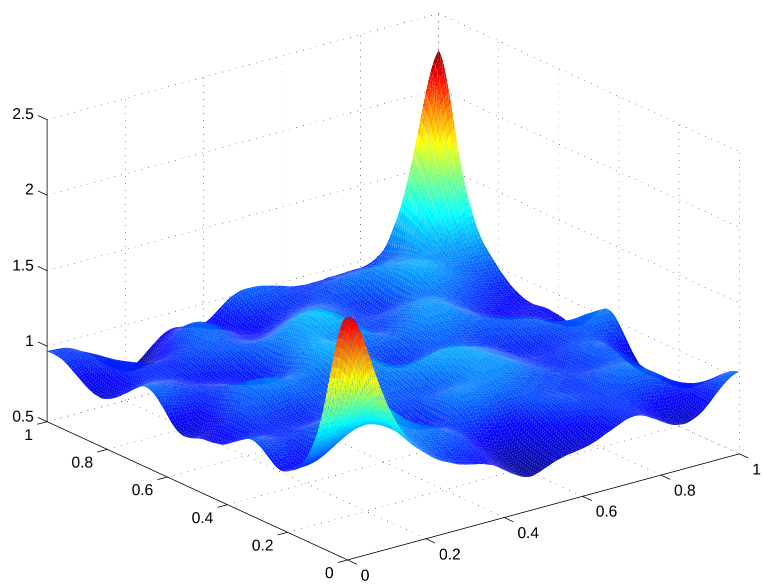

Figure 6.10 Empirical copula density function between short term data sets Ericson B and Volvo.

turned out that the p-value of the three pairs were smaller than 5%, so that the Joe survival copula does not fit the short term data.

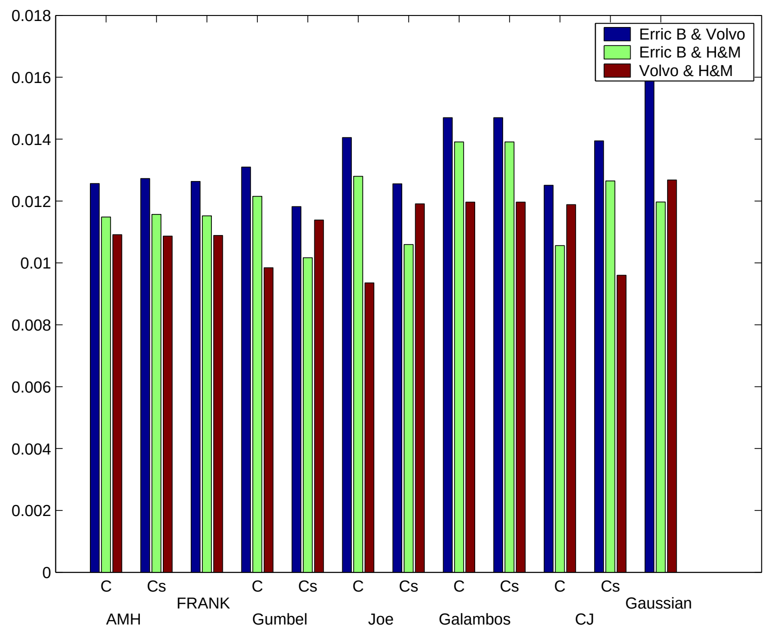

Figure 6.11 Kuiper distances of copulas to empirical copula of short term data sets Ericson B, Volvo and H&M.

6.4 Correlation as Measure of Dependence

For the observed data sets we now check how good correlation is to describe the dependence structure.

We introduce the new variate of non devolatilized logreturns

$$X_i^{\rho_j} = X_i - \frac{\mathrm{Cov}(X_i, X_j)}{\mathrm{Var}(X_j)} X_j,$$

and then checked if X_j is independent from $X_i^{\rho_j}$.

There are too many outliers from the independent distribution for it to be the dependence structure of the variates, see Figure 6.12. So we used the copula method to find the dependence structure of X_1 and $X_2^{\rho_1}$, see Figure 6.13.

Clearly, the density function of the copula is not flat, and the result displays that the dependence is much more complex than what the linear measure describes.

By the bootstrap method, the distances of the simulated data from the Π copula is found, see Figure 6.14. The distances of the three empirical copulas of the transformed

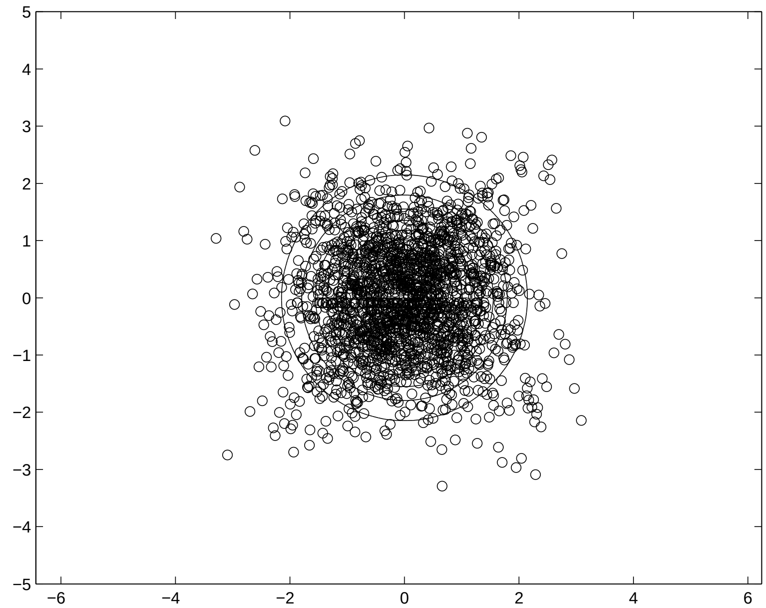

Figure 6.12 Logreturns with normal marginals of X_1 and $X_2^{\rho_1}$ and the density of the independent distribution with standard normal marginals.

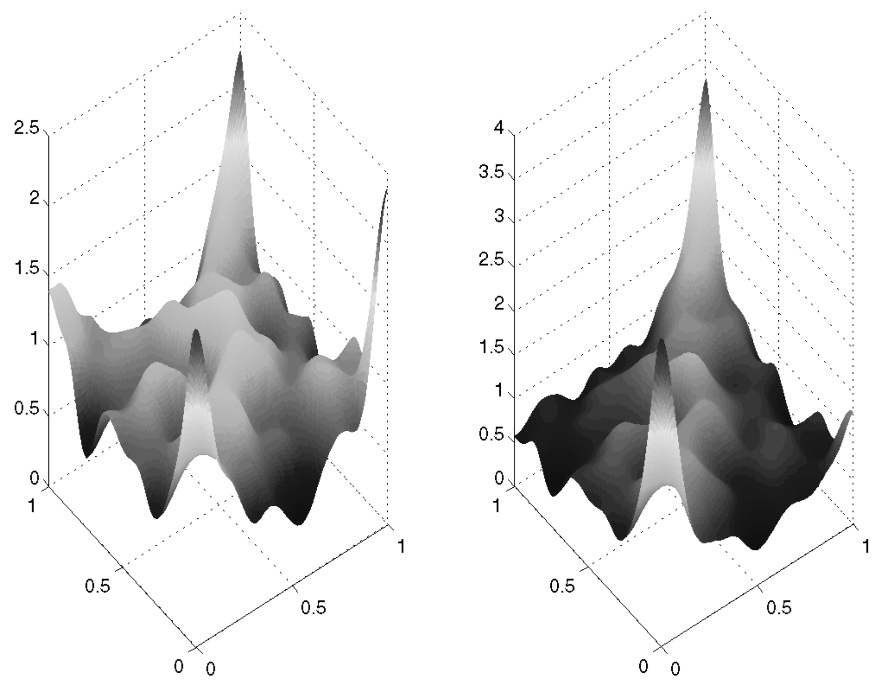

Figure 6.13 Copula density functions of X_1 & $X_2^{\rho_1}$ and X_1 & X_2.

data sets are $D_{12} = 0.103$, $D_{13} = 0.111$ and $D_{23} = 0.108$. By inspection of Figure 6.14, this shows that there still is considerable dependence between the variates.

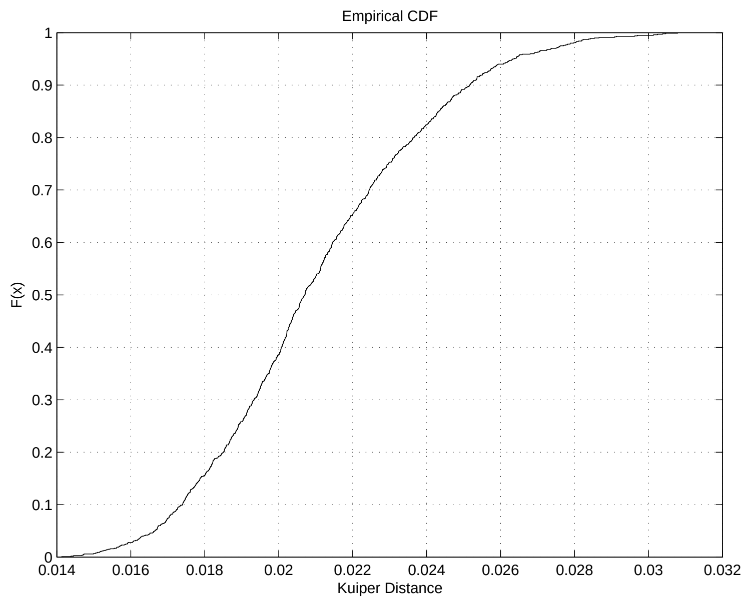

Figure 6.14 Empirical distribution for Kuiper distances between empirical copula for 1000 simulations of independent data and Π copula.

Although correlation says something about dependence, correlation is not enough to describe the dependence structure between two variates. Equal correlation does not imply same dependence structure. Consider two copula models, Gumbel survival copula and Gaussian copula. Generate two random normal vectors X_1 and X_2 from Gumbel survival copula with the parameter α, giving $\alpha = 1.5$. From (X_1, X_2), calculate the correlation ρ and then generate two vectors Y_1 and Y_2 of joint normal distribution with correlation ρ, see Figure 6.15.

Define market crash as the event when both variates simultaneously is in the lowest 5% of their marginal distributions.

Then the true risk for market crash is given by

$$C^{1.5}_{\text{Gumbel survival}}(0.05, 0.05) = 2.2\ \%.$$

And by simulating the variables (Y_1, Y_2) the estimated risk is found.

By Figure 6.16, it is shown that the risk of market crash is far larger than what is estimated by the normal assumption. This means that the risk of market crash is underestimated.

This kind of information is very important in, for instance, the field of hedging, specially when the crash dependence property is displayed in the market.

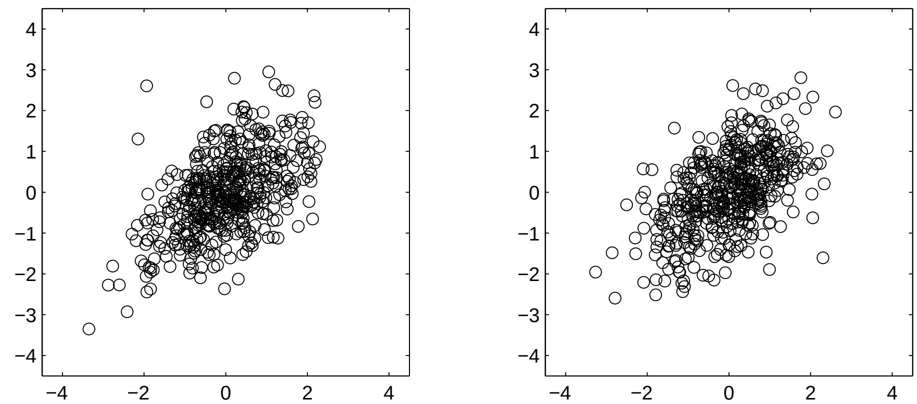

Figure 6.15 On left is 500 simulations of the Gumbel survival copula and on right is 500 simulations of the Gaussian model with equal correlation.

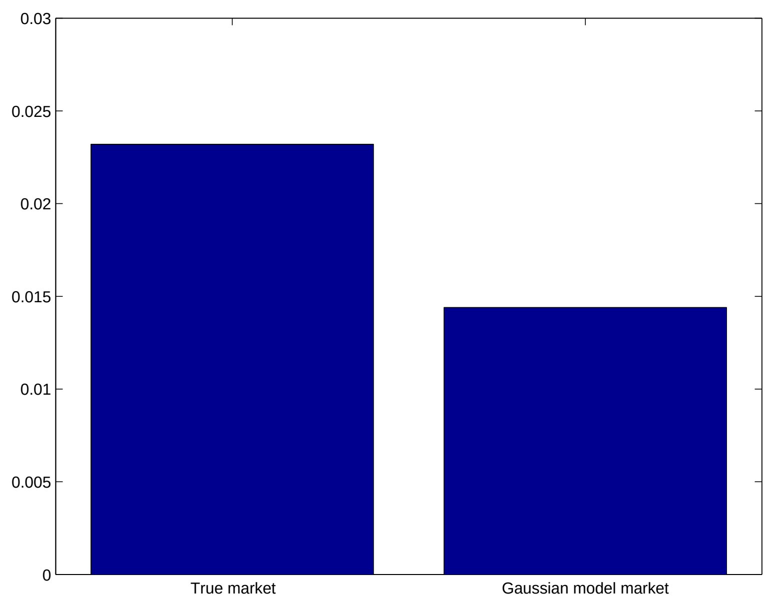

Figure 6.16 Calculated risk of market crash for the true market and the model of the market by the Monte Carlo algorithm.

All of this show that tail dependence is stronger than what the correlation take notion of. It is fairly obvious that data sets may contain unsuspected relations, but it is not known which are the strongest. An intelligent dependence structure analysis can lead to good understanding of data sets. Copula model is one of the intelligent methods.

Chapter 7
Mixture of Copulas

By collecting the copulas that estimate lower tail, upper tail and central dependence and mixing them toghether in a mixing copula, we found our model copula. We considered three copulas, Joe survival copula, Gumbel copula and AMH copula.

Appendix B displays the numerical results corresponding to the minimal Kuiper distance between empirical copula and these three copulas, individually.

With these three copulas, we are ready to define a mixture model. Take $\beta_1, \beta_2 \in [0, 1]$, $\beta_3 = 1 - \beta_1 - \beta_2$ with $\beta_1 + \beta_2 \leq 1$ and define a mixed copula as

$$C_{\text{mix}}(u, v; \alpha, \beta) = \beta_1 C_{\text{Joe survival}}(u, v; \alpha_1) + \beta_2 C_{\text{Gumbel}}(u, v; \alpha_2) + \beta_3 C_{\text{AMH}}(u, v; \alpha_3),$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ are association parameters in mixture which reflect the degree of dependence, and $\beta = (\beta_1, \beta_2)$ are weight or shape parameters which reflect the dependence structures. In Figure 7.1, we plot some mixtures as examples. For the first row $\beta_1 = 1/2$, $\beta_2 = 0$; for the middle row $\beta_1 = \beta_2 = 1/3$; for the last row $\beta_1 = 0$, $\beta_2 = 1/2$; for all figure, $\alpha_1 = 1.8$, $\alpha_2 = 1.4$, $\alpha_3 = 1$

We use a two-stage parameter estimation approach: First, we estimate the association parameters by finding the minimum Kuiper distance of every single copula. Then we use the minimum Kuiper distance to estimate the shape parameters of the mixture copula. This approach makes optimization process quite simple. The results ares shown in Table 7.1.

Two different diagnostics are used to test the goodness-of-fit; a Kuiper test and a chi-square test.

We calculated the Kuiper distances of the mixture copula for every pair of data sets, and got $D_{12} = 0.036$, $D_{13} = 0.036$ and $D_{23} = 0.030$.

To find the approximate p-value of the data set coming from the copula model, the bootstrap method was used. The results are displayed in Figure 7.2, and shows that the distances $D_{12} = 0.036$, $D_{13} = 0.036$ and $D_{23} = 0.030$ all are large enough to reject the fitted model.

The other method to test a goodness of fit is a chi-square test, see Table 7.2. Because the p-values of paired CTL & AT and SBC & AT are bigger than 5%, the model is not rejected, that is, the mixture model fits the real market data sets well enough.

The mixture copula is clearly better than the Gumbel copula, see table 7.2.

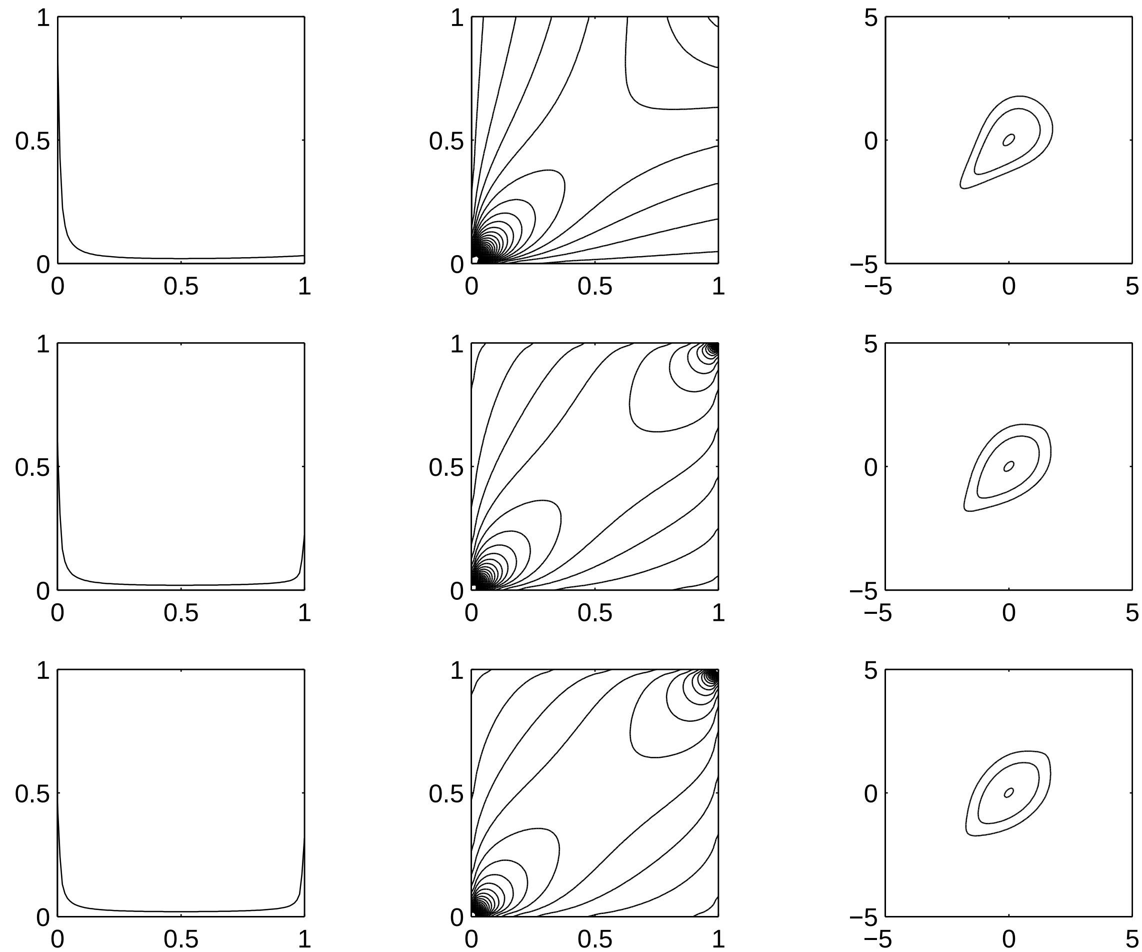

Figure 7.1 The first column is the density of three copulas cross sections on the diagonal u=v, the second column display copula density function and the third column show the joint density function with standard normal marginals.

	CTL & SBC	CTL & AT	SBC & AT
α_1	1.426	1.451	1.257
α_2	2.292	1.637	1.670
α_3	0.994	1.000	0.878
β_1	0.605	0.287	0.027
β_2	0.275	0.544	0.476
β_3	0.120	0.169	0.497

Table 7.1 Parameters of mixture model copula for three paired stocks.

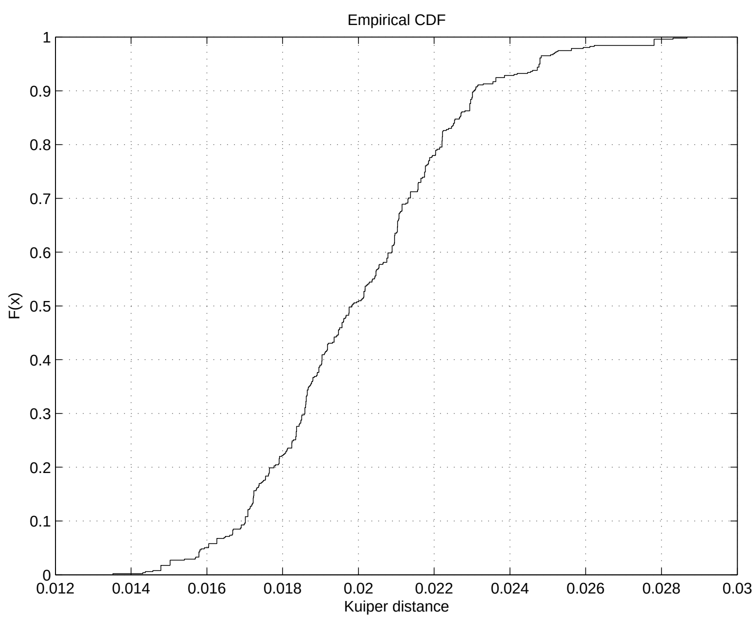

Figure 7.2 Empirical distribution of Kuiper distances between empirical copula and empicrial copuls for 1000 observations of the model copula.

	CTL & SBC	CTL & AT	SBC & AT
C_{mixture} with devolatilization	$3.909e-05$	0.018	0.587
C_{mixture} without devolatilization	0.015	0.054	0.145
C_{Gumbel} with devolatilization	$9.854e-07$	$1.510e-05$	$8.390e-04$
C_{Gumbel} without devolatilization	0.004	$6.796e-05$	$9.389e-04$

Table 7.2 p-values of mixture copula and Gumbel copula.

Chapter 8
Conclusions

By employing empirical copulas, we have seen stronger negative time dependence structure for short term data than for long term data. The dependence structure of short term data may be affected by the discontinuous price setting (in Sweden the market only uses a resolution as low as 0.5SEK). Both short term data and long term data exhibit stronger crash than boom dependence in time. This means good news are momentare while bad news are consistent.

The dependence structure between two variates in long and short term is not equal. Short term data sets exhibit equal crash and boom dependence while long term data sets clearly display stronger crash than boom dependence.

Actually it seems very unlikely to find any dependence between two short term variates. From the tick data, two variates are almost independent except at extrem events. So the dependence in short term is probably entire market changes.

A Gaussian assumption for dependence structures of variates is a bold one: We have found that the dependence structure is far more complex. Dependence in the tails is stronger than in central regions of data. Unlike multivariate gaussian distributions, the true market has asymmetric dependence. This means that good news are for minority while bad news are for majority.

By simulation it is clear that correlation is not a bad method to fir Gaussian processes. However, we know that in the real market, the data comes from other more complex distributions. So correlation cannot describe dependence among market stocks, and we have shown that copula models give far better dependence descriptions.

Assume that the dataset come from a Gumbel survival copula, and that someone makes the mistake to think that it is jointly Gaussian distributed. Then a great risk is taken, since for the real data's distribution one must considered its heavy lower tail, which is crash dependent. And so risk become greatly underestimated. The correlation is a linear estimate that is not versatile enough to take study the concept of risk.

All copula models, which we have been able to find in the literature, have been used to investigate the dependence structure for real stock data sets. And although a couple of copulas fit better than the Gaussian copula, none of them fit the dependence structure well enough to not show significant deviations from the empirical copula. This means the stock market dependence structure is more complex than a single copula models' dependence structure. Copulas in the literature just have association parameter and only control the degree of dependence. So a mixture copula is used to fit the data and work well because it has weight parameters, which have an effect on the structure of dependence, besides association parameters.

Since the dependence structure for logreturns is stronger before devolatilization than after devolatilization, real-world portfolios have time dependence structures more complicated that suggested by the Bachelier-Samuleson Black-Scholes model. Hence some exterior event seem to affect stocks prices then and then.

Appendix A
Table of Tail Dependence Coefficents

Copula	λ_U	λ_L
Gaussian	0	0
AMH	0	0
Frank	0	0
Gumbel	$2 - 2^{1/\alpha}$	0
Joe	$2 - 2^{1/\alpha}$	0
Galambos	$\frac{1}{2^{1/\alpha}}$	0
CJ	0	$\frac{1}{2^{1/\alpha}}$
BB1	$2 - 2^{\frac{1}{\alpha_2}}$	$\frac{1}{2^{1/(\alpha_1 \alpha_2)}}$

Table A.1 Tail dependence of copulas.

Appendix B
Table of Distances and Parameters for Long Term Data

Copula	AT & CTL	AT & SBC	CTL & SBC
Gaussian	$D = 0.048$ $\rho = 0.366$	$D = 0.048$ $\rho = 0.420$	$D = 0.040$ $\rho = 0.449$
AMH	$D = 0.050$ $\alpha = 1.000$	$D = 0.057$ $\alpha = 1.000$	$D = 0.052$ $\alpha = 1.000$
AMH Survival	$D = 0.060$ $\alpha = 0.803$	$D = 0.065$ $\alpha = 0.916$	$D = 0.064$ $\alpha = 1.000$
Frank	$D = 0.045$ $\alpha = 0.110$	$D = 0.047$ $\alpha = 0.064$	$D = 0.039$ $\alpha = 0.056$
Frank Survival	$D = 0.045$ $\alpha = 0.110$	$D = 0.047$ $\alpha = 0.064$	$D = 0.039$ $\alpha = 0.056$
Gumbel	$D = 0.053$ $\alpha = 1.441$	$D = 0.055$ $\alpha = 1.439$	$D = 0.046$ $\alpha = 0.896$
Gumbel Survival	$D = 0.040$ $\alpha = 1.483$	$D = 0.042$ $\alpha = 1.476$	$D = 0.037$ $\alpha = 0.793$
Joe	$D = 0.070$ $\alpha = 1.735$	$D = 0.070$ $\alpha = 1.671$	$D = 0.065$ $\alpha = 1.807$
Joe Survival	$D = 0.052$ $\alpha = 1.740$	$D = 0.054$ $\alpha = 1.849$	$D = 0.052$ $\alpha = 1.795$
Galambos	$D = 0.054$ $\alpha = 0.722$	$D = 0.056$ $\alpha = 0.726$	$D = 0.047$ $\alpha = 0.896$
Galambos Survival	$D = 0.056$ $\alpha = 0.773$	$D = 0.056$ $\alpha = 0.759$	$D = 0.048$ $\alpha = 0.793$
CJ	$D = 0.047$ $\alpha = 0.879$	$D = 0.052$ $\alpha = 0.946$	$D = 0.048$ $\alpha = 0.978$
CJ Survival	$D = 0.067$ $\alpha = 0.833$	$D = 0.068$ $\alpha = 0.756$	$D = 0.063$ $\alpha = 0.920$

Table B.1 Minimal Kuiper distances and corresponding copula parameters for long term data.

Appendix C
Table of Distances and Parameters for Short Term Data

	Ericson B & Volvo	Ericson B & H&M	Ericson B & H&M
AMH	$D = 0.013$ $\alpha = 0.095$	$D = 0.012$ $\alpha = 0.063$	$D = 0.011$ $\alpha = 0.047$
AMH Survival	$D = 0.013$ $\alpha = 0.095$	$D = 0.012$ $\alpha = 0.063$	$D = 0.011$ $\alpha = 0.047$
Frank	$D = 0.013$ $\alpha = 0.823$	$D = 0.012$ $\alpha = 0.880$	$D = 0.011$ $\alpha = 0.910$
Gumbel	$D = 0.013$ $\alpha = 1.027$	$D = 0.012$ $\alpha = 1.015$	$D = 0.010$ $\alpha = 1.013$
Gumbel Survival	$D = 0.012$ $\alpha = 1.025$	$D = 0.010$ $\alpha = 1.019$	$D = 0.011$ $\alpha = 1.014$
Joe	$D = 0.014$ $\alpha = 1.039$	$D = 0.013$ $\alpha = 1.024$	$D = 0.009$ $\alpha = 1.023$
Joe Survival	$D = 0.013$ $\alpha = 1.019$	$D = 0.011$ $\alpha = 1.027$	$D = 0.011$ $\alpha = 1.005$
Galambos	$D = 0.015$ $\alpha = 0.000$	$D = 0.014$ $\alpha = 0.000$	$D = 0.012$ $\alpha = 0.000$
Galambos Survival	$D = 0.015$ $\alpha = 0.000$	$D = 0.014$ $\alpha = 0.000$	$D = 0.012$ $\alpha = 0.000$
CJ	$D = 0.013$ $\alpha = 0.028$	$D = 0.011$ $\alpha = 0.033$	$D = 0.012$ $\alpha = 0.006$
CJ Survival	$D = 0.014$ $\alpha = 0.051$	$D = 0.013$ $\alpha = 0.029$	$D = 0.010$ $\alpha = 0.029$

Table C.1 Minimal Kuiper distances and corresponding copula parameters for short term data.

Appendix D
Goodness of Fit

To measure how close, or how far, an empirical distribution is from a theoretical distribution, several distances are use. Among them, we cite three: the Kolmogorov-Smirnov distance, the Anderson-Darling distance and the Kuiper distance.

The Kolmogorov-Smirnov distance is the greatest distance between the empirical distribution and a hypothetical theoretical distribution for the data, i.e. for us in the sense of copulas:

$$D_{\mathrm{KS}} = \max_{u,v \in [0,1]} |C_{\mathrm{emp}}(u,v) - C_{\mathrm{theory}}(u,v)|, \tag{D.1}$$

where C_{emp} is the empirical copula and C_{theory} the theoretical copula function.

The Anderson-Darling distance, see Anderson-Darling (1954), is defined as a scaled version of the Kolmogorov-Smirnov distance:

$$D_{\mathrm{AD}} = \max_{u,v \in [0,1]} \frac{|C_{\mathrm{emp}}(u,v) - C_{\mathrm{theory}}(u,v)|}{\sqrt{C_{\mathrm{theory}}(u,v)(1 - C_{\mathrm{theory}}(u,v))}}. \tag{D.2}$$

The Anderson-Darling distance emphasizes the fit in the tails, which makes it inadequate for our purposes, since we are interested in the entire distrbution.

The Kuiper distance, see Kuiper (1962), considers greatest distances upwards as well as downwards:

$$D_{\mathrm{Kuiper}} = \max_{u,v \in [0,1]} (C_{\mathrm{emp}}(u,v) - C_{\mathrm{theory}}(u,v)) + \max_{u,v \in [0,1]} (C_{\mathrm{theory}}(u,v) - C_{\mathrm{emp}}(u,v)). \tag{D.3}$$

Appendix E
p-values

Each statistical test has an associated null hypothesis, the p-value is the probability that the samples could have been drawn from the model being tested, given the assumption that the null hypothesis is true. A p-value of .05, for example, indicates that you would have only a 5% chance of drawing the sample being tested if the null hypothesis was actually true.

A null hypothesis is typically a statement of no difference. A p-value close to zero signals that the null hypothesis is false, and that a difference is very likely to exist. Large p-values closer to 1 imply that there is no detectable difference for the sample size used. A p-value of 0.05 is a typical threshold used in industry to evaluate the null hypothesis.

To show if copula models fit the logreturn data sets, we can calculate a p-value by the chi-square statistic test.

The chi-square statistic test of k boxes is given by

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j},$$

where O_j is observed frequency for box j, i.e., the number of observations that lies in the box j, and where

$$E_j = n[C(u_j, v_j) - C(u_{j-1}, v_j) - C(u_j, v_{j-1}) + C(u_{j-1}, v_{j-1})]$$

is the expected frequency for box j. Here $u_j > u_{j-1}, v_j > v_{j-1}$, $u_0 = v_0 = 0$ and $u_k = v_k = 1$ must hold.

Appendix F
Bachelier-Samuleson Black-Scholes Model

Several models have been proposed to model the price process, $S(t)$, of an asset. The most widely used model is the Bachelier-Samuleson Black-Scholes model, which gives the stock value at time t as the solution to the stochastic differential equation

$$dS(t) = (\mu + \frac{\sigma}{2})S(t)dt \; + \; \sigma S(t)\, dB_t, \tag{F.1}$$

where B_t is brownian motion.

The solution to (F.1) is

$$S(t) = S(0)e^{\mu t + \sigma B_t}, \tag{F.2}$$

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

By considering the logreturn, $X(t)$, of the asset value, the data set becomes driven simply 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 inteval between sampling points. Hence, for the Bachelier-Samuelson model, the logreturns of stock values ares the increments of an Brownian motion, i.e. they are stationary and independent.

References

1. Albin P. (2000). Stochastic simulation. Unpublished Lecture Notes, Chalmers University of Technology.

   http://www.math.chalmers.se/~palbin/Lecture.ps

2. Anderson, T.W. and Darling, D.A. (1954). A test of goodness of fit. *J. Amer. Statist. Assoc.* **49** 765-769.

3. Andersson, D. and Lernvik, M. (2004). On longe range dependence in the returns of risky assets. Master's Thesis, Chalmers University of Technology.

   http://www.math.chalmers.se/~palbin/daniel-magnus.ps

4. Drees, H. and Starica, C. (2002). A simple non-stationary model for stock returns. Preprint.

   www.math.chalmers.se/Math/Research/Preprints/2002/73.ps.gz

5. Bengtsson, M. and Olsbo, V. (2002). Value at risk using stochastic volatiliy models. Master's Thesis, Chalmers University of Technology.

   http://www.math.chalmers.se/~palbin/mattiasviktor.ps

6. Chen, X. and Fan, Y. (2002). Evaluating density forecast via the copula approach. Preprint.

   http://www.economics.ucr.edu/seminars/fall02/FanY.PDF

7. Embrechts, P., Lindskog, F. and McNeil, A. (2000). Modelling dependence with copulas and applications to risk management. Preprint.

   http://www.risklab.ch/ftp/papers/DependenceWithCopulas.pdf

8. Hu, L. (2004). Dependence patterns across financial markets: a mixed copula approach. Preprint.

   http://swan.econ.ohio-state.edu/ling/papers/copula.pdf

9. Kuiper, N.H. (1962). On a metric in the space of random variables. *Statistica Neerlandica* **16** 231-235.

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

11. Nelsen, R.B. (1999). *An Introduction to Copulas.* Springer, New York.

12. Ornelas, J.R.H., de Farias, A.R. and Barbachan, J.S.F. (2003). Goodness-of-fit tests focus on VaR estimation. Preprint.

http://professores.ibmecrj.br/pepe/research/papers/FOF.pdf

13. Schweizer, B. (1991). Thirty years of copulas. In *Advances in Probability Distributions with Given Marginals* (G. Dall'Aglio, ed.) 13-50. Kluwer, Dordrecht.

14. Sklar, A. (1959). Fonctions de répartition àn dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris.* **8** 229-231.

15. Watson, G.S. (1964). Smooth regression analysis, *Sankhya, Series A* **26** 359-372.

16. Zivot, E. (2002). EVANESCE implementation in S-PLUS FinMetrics module. Preprint.

http://faculty.washington.edu/ezivot/book/QuanCopula.pdf