Master Theses:
Properties and Applications of the Student T Copula

Joanna Gatz

July 30, 2007

Contents

Chapter 1

Introduction

The main focus of this theses is on the properties and applications of the T copula. This distribution was initially considered as a good alternative to the Normal copula used in non parametric continuous BBN's. However, it was discovered quite early, that it is not possible as the t distribution does not have the independence property, which is the key assumption in modeling BBNs. Despite that, the t distribution appeared to be a very interesting one itself. And the project was carried out to learn more about properties of this distribution as well as its use in analysis a data. The project was not meant initially to analyze a certain dataset. However, a small dataset was chosen later to illustrate challenges in estimating the t distribution. The t distribution is often used in financial data returns such as equities, assets or foreign exchange rates that tend to be uncorrelated but dependent. They have heavy tails and extremes appear in clusters. The t distribution is often suggested, because of its properties following mostly from the fact that it belongs to the family of the elliptical distributions. It enables to capture correlations in the extreme market movements, since it has the tail dependence property. It is also easy to simulate and the corresponding copula allows combining different margins using dependence function that exhibits tail dependence property. The Student t distribution can be parameterized by the correlation, which is easily estimated from the data, as well as the degrees of freedom parameter that controls degree of tail dependence. The t distribution is very interesting and offers a convenient way for building models.

The multivariate t-distribution allows however a single degrees of freedom parameter. This may be seen as significant restriction. The pairwise relations between variables can be preserved by applying graphical depen-

dence model called vines. It is a set of nested trees that can be specified by rank correlations, distributions or copulas. Vine is a structure that allows to construct a model based on different bivariate T copulas. In this way dependences between variables are preserved more appropriately, especially tail dependences.

We will start with introducing the Student t distribution We limit ourself to presenting properties that are important for applications. Chapter 2 treats univariate t distribution as well as the multivariate one. As an example, we present the heavy tails analysis to determine degrees of freedom parameter for the univariate foreign exchange rates data. For higher dimensions, there are estimators for computing parameters of the t distribution. We can also use semi-parametric likelihood function, meaning that we can estimate the correlation matrix using Kendall's τ, and maximize likelihood with respect to the remaining parameter. Doing so, we automatically assume that the marginals are the same, which is not necessarily a true assumption. Instead, we can estimate only a dependence structure - copula. Chapter 3 focuses on T copulas. This mathematical functions separate the marginal behavior from the underlying dependence structure and are based on the t distribution. The methodology for estimating T copulas that is shown.

It is not easy to capture pairwise dependences between variables, when a multivariate model is built. Application of the multivariate T copula gives bivariate marginals that have the same degrees of freedom parameter. In practice, however, it is rare that pairs of analyzed variables have the same tails. In order to overcome this difficulty we apply vines. Chapter 4 provides a basic introduction about vines. This model builds a joint distribution from bivariate and conditional bivariate distributions. Using bivariate T copulas in the vine model allows better modeling flexibility than given by multivariate T copula.

We compare the model with the three dimensions T copula estimated for the same dataset.

We end this work with a discussion about the results and further research.

Chapter 2

Student T Distribution

In recent years more attention has been paid to the Student T Distribution. It is considered to be an appealing alternative to the normal distribution. Both the multivariate T distribution and the multivariate normal distribution are members of the general family of elliptical distributions. The multivariate T distributions generalize the classical univariate Student T distribution.

It can have many different forms with different characteristics. Application of the multivariate T distributions is a very promising approach in multivariate analysis. The multivariate T distribution is more suitable for real world data than normal distribution, particularly because its tails are heavier and therefore realistic.

This chapter contains the background information about the best known form of the T distribution, the canonical T distribution.

We start with introducing basic properties of the univariate Student t distribution. Since it is a heavy tail distribution, we present the tail index estimators and show how it can be used in order to estimate the degrees of freedom parameter. This is an alternative approach to maximum likelihood method for estimating this parameter. Student t distribution is commonly used to model financial data. We fit using tail index estimators univariate Student t distributions to the foreign exchange returns datasets.

Further, we present some chosen properties of the bivariate Student t distribution. These will be extended to the multivariate case.

The notion of the copula will be presented in this chapter as well. We will show the construction of the t copula together with its properties. We will present method for the inference for this copula, which is based on the semi

parametric pseudo - likelihood method. We will show how it works in practice by fitting the t copula to the foreign exchange returns datasets. We will see that this methodology is quite effective.

2.1   Univariate Student T distribution

The student T distribution ([2],[3]) is an absolutely continuous probability distribution given by the density function

$$f_v(x) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\Gamma(\frac{v}{2})}\left(1+\frac{x^2}{v}\right)^{-(\frac{v+1}{2})}, \qquad -\infty < x < \infty, \qquad (2.1)$$

where $v > 0$ is a degrees of freedom parameter, called also shape parameter, and $\Gamma(.)$ is the Euler gamma function of the form

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt.$$

The Student t distribution with v degress of freedom is denoted as t_v, and the density by dt_v.

In particular, when $v = 1$, then the density (2.1) takes the form:

$$f_{t,1}(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty$$

which corresponds to the Cauchy distribution. The T distribution is symmetrical around zero. The odd moments vanish, when they exist. The kth moment exists if and only if $k < v$

$$E(X^k) = \begin{cases} 0 & k \quad \text{odd}, 0 < k < v, \\ \frac{\Gamma(\frac{k+1}{2})\Gamma(\frac{v-k}{2})v^{\frac{k}{2}}}{\sqrt{\pi}\Gamma(\frac{v}{2})} & k \quad \text{even}, 0 < k < v, \\ NaN & k \quad \text{odd}, 0 < v \leq k, \\ inf & k \quad \text{even}, 0 < v \leq k. \end{cases}$$

The variance and kurtosis are given by the formulas:

$$\frac{v}{v-2}, \qquad \text{for} \quad k > 2,$$

$$3 + \frac{6}{v-4}, \qquad \text{for} \quad k > 4,$$

respectively.

The density (2.1) can be constructed, (see [1]), by letting $Z_0, Z_1, \ldots, Z_v$ be independent standard Normal random variables and setting

$$\chi_v^2 = Z_1^2 + \ldots + Z_v^2,$$

where χ_v^2 is a chi-squared random variable with v degrees if freedom, mean v, variance $2v$ and density given by

$$q_v(z) = \frac{1}{2\Gamma(\frac{v}{2})} e^{-(\frac{z}{2})} \left(\frac{z}{2}\right)^{(\frac{v}{2}-1)}.$$

Then the Student T random variable is defined as follows:

$$T = \frac{Z_0}{\sqrt{\chi_v^2 \setminus v}}.$$

In order to obtain the density $f_v(t)$ it suffices to notice that the conditional density of T given χ_v^2 is the normal density function with mean zero and variance $\frac{v}{u}$ of the form:

$$f_v(t|\chi_v^2 = u) = \sqrt{\frac{u}{2\pi v}} exp\left(-\frac{t^2 u}{2v}\right), \tag{2.2}$$

In order to derive the joint density function of T and χ_v^2, the conditional density (2.2) is multiplied by the chi squared density, $q_v(z)$. Finally, to extract the univariate density (2.1) for T, we integrate the following expression:

$$\int_0^\infty f_v(t|\chi_v^2 = u)q_v(u)du \equiv \int_0^\infty \frac{du}{2\Gamma(\frac{v}{2})} \sqrt{\frac{u}{2\pi v}} \left(\frac{u}{2}\right)^{(\frac{v}{2}-1)} e^{-\left(\frac{u}{2}+\frac{t^2 u}{2v}\right)}.$$

The resulting density is of the univariate Student t distribution given by (2.1).

The cumulative distribution function (cdf) of the T distribution is given by the formula:

$$F_v(x) = \int_{-\infty}^x f_v(t)dt = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\Gamma(\frac{v}{2})} \int_{-\infty}^x \left(1+\frac{t^2}{v}\right)^{-(\frac{v+1}{2})} dt.$$

It can be written in terms of geometric functions:

$$F_v(x) = \frac{1}{2} + \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\Gamma(\frac{v}{2})} x \; {}_2F_1\left(\frac{1}{2}, \frac{v+1}{2}, \frac{3}{2}, -\frac{x^2}{v}\right),$$

7

where $_2F_1$ is a Gaussian hypergeometric function. We can rewrite it also in terms of β functions:

$$F_v(x) = \frac{1}{2}\left(1 + sign(x)\left(1 - I_{v/x^2+v}\left(\frac{v}{2}, \frac{1}{2}\right)\right)\right).$$

The regularized β - function is given by

$$I_x(a,b) = \frac{B_x(a,b)}{B(a,b)},$$

where $B(a,b)$ is the ordinary β function and $B_x(a,b)$ is the incomplete β function of the form:

$$B_x(a,b) = \int_0^x t^{a-1}(1-t)^{(b}-1)dt.$$

The inverse cumulative distribution function is:

$$F_v^{-1}(u) = sign\left(u - \frac{1}{2}\right)\sqrt{v\left(\frac{1}{I_{if[u<1/2,2u,2(1-u)]}^{-1}(n/2,1/2)} - 1\right)}.$$

This result is not easy to implement. Generally it is not used in practice, as there are other, faster methods for calculating the quantile function. Some of them are discussed by William Shaw ([3]).

Sampling from T distribution can be accomplished in a simple way. A survey of the classical methods for simulation is given in Section $IX.5$ by Devroey ([4]). We sample T distribution, first by using $v + 1$ samples from the standard Normal distribution, assumed that v is an integer. (Deegres of freedom parameter v can take also real values in its range.) Alternatively, the normal variate can be divided by scaled sample from the χ^2 distribution. Figure 2.1 shows the results of sampling the student T distribution with 3 degrees of freedom, using two methods:

- We take 1000 samples for four independent standard normal variables, Z_1, Z_2, Z_3, Z_4, and calculate:

$$\frac{Z_1 \times \sqrt{(3)}}{Z_2^2 + Z_3^2 + Z_4^2},$$

which has the student T distribution with 3 degrees of freedom. (Figure 2.1 a)).

8

- 1000 independent samples are taken from the standard normal variable Z_1 and chi-squared variable, χ_3^2, with 3 degrees of freedom. We calculate:

$$\frac{Z_1 \times \sqrt{(3)}}{\chi_3^2},$$

  which is the sample from the student T distribution with 3 degrees of freedom, (Figure 2.1 b)).

Both approaches follow straight forward from the construction of the T distributed random variable.

To see that the tails of the T distribution are heavier than those of the normal distribution, the standard normal distribution was sampled as well 2.1 c). The univariate normal and T student density functions are plotted in Figure 2.1 d). We can observe that the T density function approaches normal density, as the parameter of the degrees of freedom grows.

In fact, the student T distribution converges to normal distribution as parameter of degrees of freedom goes to infinity and we have following relations:

$$\lim_{v \to \infty} f_{t,v}(x) = \phi(x),$$

$$\lim_{v \to \infty} P_{t,v}(X \leq x) = \Phi(x).$$

2.1.1 Parameter Estimation

In order to draw conclusions about a population we use information obtained from data. That involves parameter estimation, hypothesis testing and modeling.

A standard way of estimating the density function of the population is to identify appropriate characteristics, such as symmetry, range and so on. Then we choose some well known parametric distribution that has those characteristics and estimate the parameters of that distribution. For instance, if the probability density of interest has an infinite support on both ends, then a normal or a Student's T distribution may be an useful approximation.
One of the most common ways of estimating parameters is the maximum likelihood method,([5]).
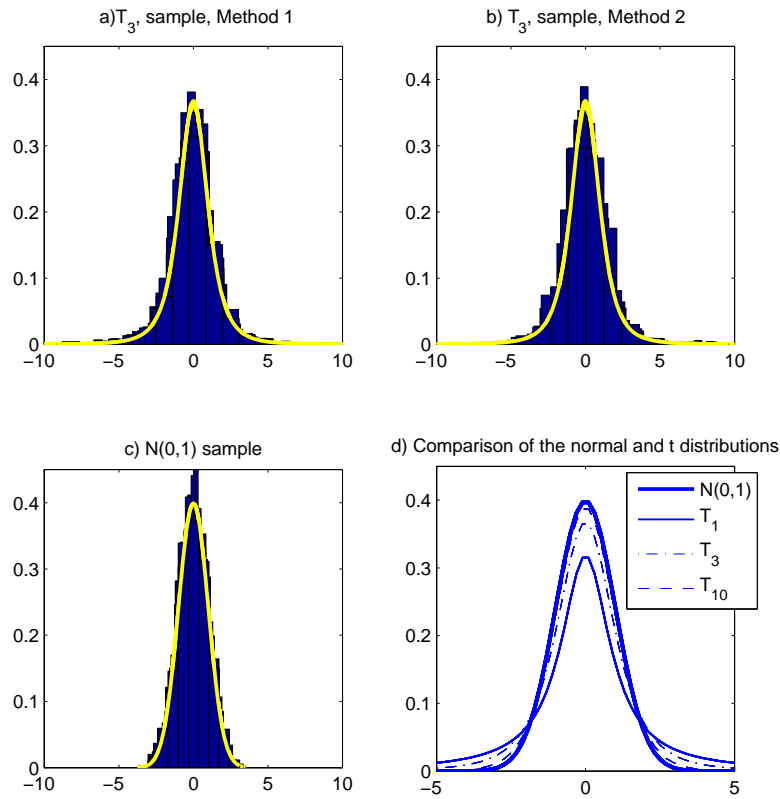
Figure 2.1: a),b)-Sampling the univariate t distribution with the comparative plot of the normal density; c)-Sampling of the normal distribution; d)- plot of the standard normal density and student t densities with $1, 3$ and 10 degrees of freedom.

Maximum Likelihood

The method of maximum likelihood involves the use of a *likelihood function* that is the joint density for a random sample. Suppose that $f(y|\theta)$ is the density function of variable Y, where θ denotes the vector of parameters. And supppose that we have n independent samples from Y. Then the likelihood function is the function of the parameter vector θ and is defined as follows:

$$L(\theta; y_1, \ldots, y_n) = \Pi_i f(y_i|\theta).$$

10

The maximum likelihood estimate of θ for the given data is the value of θ for which $L(\theta; y_1, \ldots, y_n)$ attains its maximum value. The data, which are realizations of the variables in the density function, are considered as fixed and parameters are considered as variables of the optimization problem, ([5]). In many cases it is more convenient to work with natural logarithm of function $L(\theta; y_1, \ldots, y_n)$ than with the function $L(\theta; y_1, \ldots, y_n)$ itself.

The maximum likelihood function for Student T distribution has the following form:

$$L(v; x_1, \ldots, x_n) = \Pi_{i=1}^{n} \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\Gamma(\frac{v}{2})}\left(1 + \frac{x_i^2}{v}\right)^{-(\frac{v+1}{2})}.$$

Taking the logarithm results in:

$$\log L(v; x_1, \ldots, x_n) = \log\left(\frac{\Gamma^n(\frac{v+1}{2})}{\sqrt{\pi v}^n \Gamma^n(\frac{v}{2})}\right) - \frac{v+1}{2}\Sigma_{i=1}^{n}\log\left(1 + \frac{x_i^2}{v}\right).$$

Taking the derivative with respect to v and equating to zero gives:

$$\frac{d\log L(v; x_1, \ldots, x_n)}{dv} =$$

$$\frac{\frac{\frac{1}{2}\Gamma^n\left(\frac{v+1}{2}\right)n\Psi\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}^n\Gamma^n\left(\frac{v}{2}\right)} - \frac{\frac{1}{2}\Gamma^n\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}^n\Gamma^n\left(\frac{v}{2}\right)\frac{n}{v}} - \frac{\frac{1}{2}\Gamma^n\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}^n\Gamma^n\left(\frac{v}{2}\right)n\Psi\left(\frac{v}{2}\right)}}{\Gamma^n\left(\frac{v+1}{2}\right)\sqrt{\pi v}^n\Gamma^n\left(\frac{v}{2}\right)} - \frac{1}{2}\left(\Sigma_{i=1}^{n}\log(1 + \frac{x_i^2}{v})\right)$$

$$-\frac{v+1}{2}\Sigma_{i=1}^{n} - \frac{x_i^2}{v}\frac{1}{1 + \frac{x_i^2}{v}} = 0,$$

where $\Psi(x)$ has the form:

$$\Psi(x) = \frac{d\log\Gamma(x)}{dx}.$$

Solving this equation for v is very difficult.
In order to obtain an estimate for v we can calculate the $\log L(v; x_1, \ldots, x_n)$ for various v and find the maximum among those. As an example, let's take a 100 samples from the Student T distribution with $v = 4$ degrees of freedom. The picture 2.2 presents the $\log L(v; x_1, \ldots, x_n)$ computed for $v = 1$
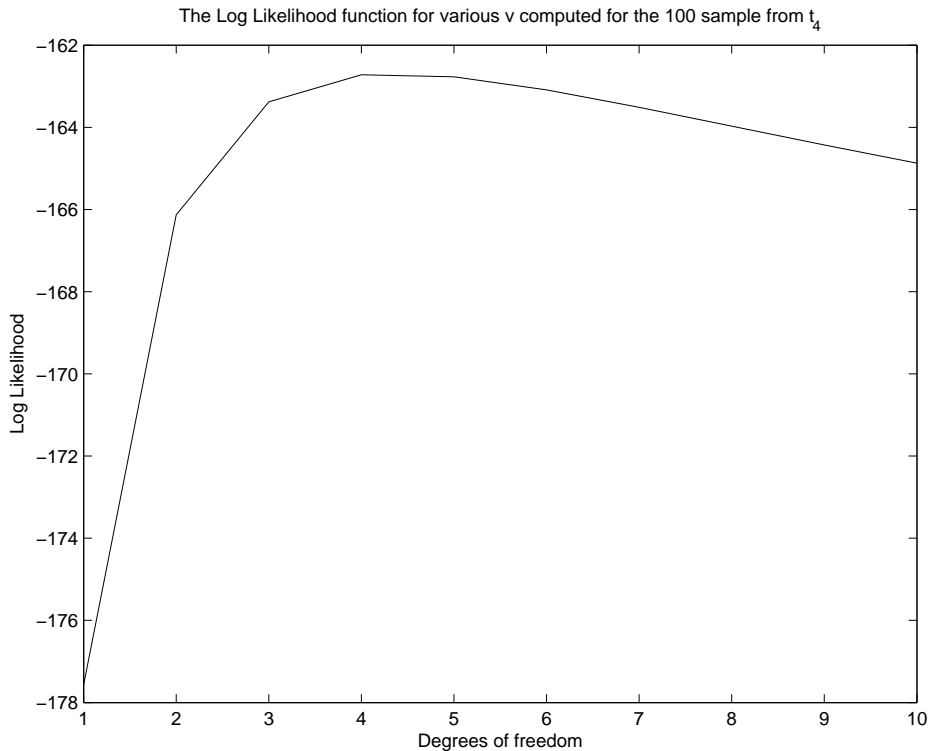
11

Figure 2.2: The log-likelihood function for Student T distribution for $v = 1, \ldots, 10$ degrees of freedom.

to 10 with step 1. The maximum is obtained for $v = 4$. Since Student t distribution has heavy tails governed by the degrees of freedom parameter v, we can apply alternative methods for estimating this parameter.

Proceeding section provides background informations about the theory behind the tail index estimators. We explain the relation between the tail index and the degrees of freedom parameter. By means of simulation we examine the performance of presented estimator for the Student t distribution. We will see that they are quite accurate.

Estimation of the Tail Index

The basic ideas behind the methodology are presented. The various estimators for tail index are provided, however without rigorous mathematical treatment. These can be found in [6], [7],[8],[9]. The bootstrap method is

12

applied in order to improve performance of the Hill estimator ([7]).

Heavy tail is the property of a distribution for which the probability of a large values is relatively big. Insurance losses, financial log-returns, file sizes stored on a server are examples of heavy tailed phenomena. The modeling and statistic of such phenomena are tail dependent and much different than classical modeling and statistical analysis. Heavy tail analysis does not give the primacy to central moments, averages and the normal density with light tail as the classical approach does. Heavy tail analysis is the study of systems whose behavior is governed by large values which shock the system periodically. In heavy - tailed analysis the asymptotic behavior of descriptor variables is determined by the large values.

Roughly speaking, a random variable T has a heavy tail (right) if there exists a positive parameter $\alpha > 0$ such that

$$P(T > t) \sim x^{-\alpha}, \quad x \to \infty. \tag{2.3}$$

Examples of such random variables are those with Cauchy, Pareto, Student t, F or stable distribution. An elementary observation is that a heavy -tailed random variable has a relatively large probability of exhibiting a large value, compared to random variables, which have exponentially bounded tails such as normal, Weibull, exponential, or gamma random variables. The concept of heavy-tail distribution should not be confused with the concept of a distribution with infinite right support. For instance, both normal and Pareto distributions have positive probability of achieving a value bigger than any preassigned threshold. However, Pareto variables have much bigger probability of exceeding the threshold.

The theory of regularly varying functions is the appropriate mathematical analysis tool for proper discussion of heavy - tail phenomena. We can think of regularly varying functions as functions which behave asymptotically like power functions. The formal definition of regularly varying function is given below:

Definition 2.1 *A measurable function* $U : R_+ \to R_+$ *is regularly varying at* ∞ *with index* $p \in R$ *if for* $x > 0$

$$\lim_{t \to \infty} \frac{U(tx)}{U(t)} = x^p.$$

We call p the exponent of variation.

The canonical p-varying function is x^p. The following functions are not regularly varying: $e^x, \sin(x)$. In probability applications we are concerned with distributions whose tails are regularly varying. Examples are

$$1 - F(x) = x^{-\alpha}, \quad x \geq 1, \quad , \alpha > 0,$$

where $F(x)$ is the cumulative distribution function:

- The tail of the Student t distribution is regularly varying,[3]. Additionally we have that a decay in probability density function of a Student t distribution with v degrees of freedom is

$$O(t^{-v-1}),$$

  which means that its tails are bounded by a polynomial with degree $(-v - 1)$. The decay for the cumulative distribution function is

$$O(t^{-v}).$$

  Because of the symmetry of this distribution, also $1 - F(x)$ has the same decay. If we estimate the tail index α we can take parameter v equal to α. Figure 2.3 shows $1 - F(x)$ for t_3 and function x^{-3}, $c > 0$. As x increases, the $1 - F(x)$ behaves like the x^{-3} function.

- The extreme-value distribution

$$\Phi_\alpha(x) = \exp(-x^{-\alpha}), \quad x \geq 0.$$

  $\Phi_\alpha(x)$ has the property

$$1 - \Phi_\alpha(x) \sim x^{-\alpha} \quad as \quad x \to \infty.$$

- A stable law with index $\alpha, 0 < \alpha < 2$ has the property

$$1 - G(x) \sim cx^{-\alpha}, \quad x \to \infty, \quad c > 0.$$

  The detailed discussion the theory of regularly varying functions can be found in [6].

14

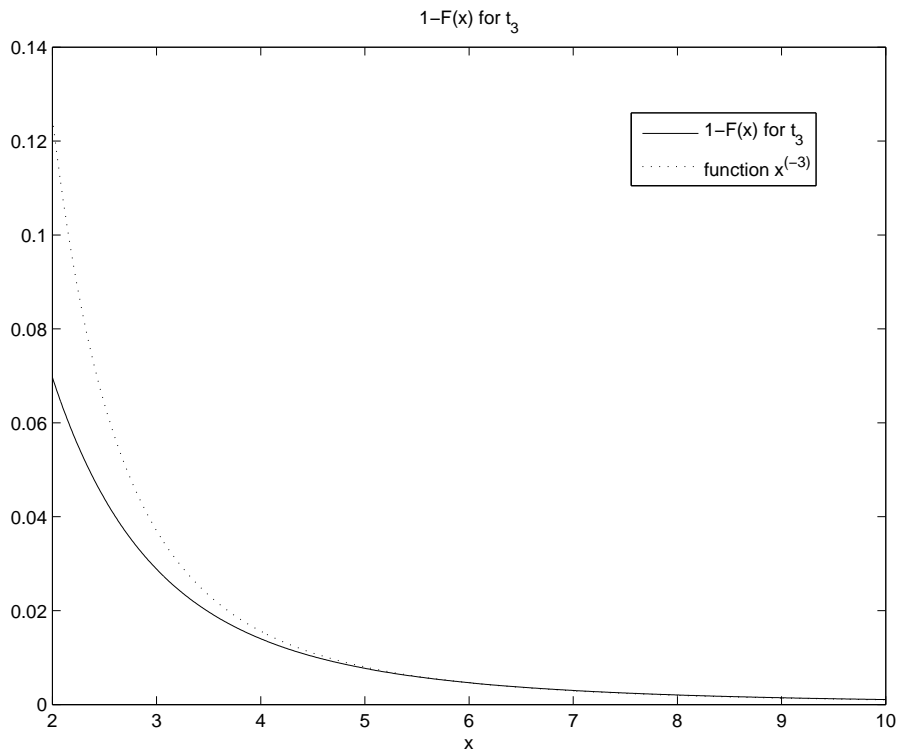Figure 2.3: The 1-F(x) function for Student t distribution with $v = 3$ degrees of freedom and x^{-3} function.

Estimators for Heavy Tail Index α

There is a number of estimators that can be applied in order to obtain an estimate for tail index α, and therefore for parameter v for Student t distribution. We present three of them. The formal mathematical proves that they are consistent is omitted here. These can be found in [6],[9],[8].

Majority of the tail index estimators are actually estimators for $1/\alpha$. They are based on the two competing models described below:
Suppose $X, X_1, \ldots, X_n$ have the same distribution $F(x)$ and that inference is to be based on $x_1, \ldots, x_n$ denoting the observations for each variable. There are at least two competing heavy-tail models:

1. We can assume that F has a Pareto right tail from some point on. This

means that there exists some $x_l > 0$, and $\alpha > 0$ such that

$$P(X > x) = cx^{-\alpha}, \quad x > x_l.$$

So we assume an exact Pareto tail from x_l onwards. The form of the tail for $x < x_l$ may or may not be specified in this approach.

2. Assume that F has regularly varying right tail with index $(-\alpha)$,

$$P(X > x) = 1 - F(x) = x^{-\alpha}L(x),$$

where $L(x)$ is a slowly varying function, that is with $p = 0$.
The expression above is the semi-parametric assumption of regular variation. The focus is on estimating the index of regular variation α.

List below contains three estimator for $1/\alpha$ together with their short description.
Suppose we have independent and identically distributed variables that are observed $X_1, \ldots, X_n$. For $1 \leq i \leq n$, write $X_{(i)}$ for the ith largest value of $X_1, \ldots, X_n$, so that

$$X_{(1)} \geq X_{(2)} \geq \ldots X_{(n)}$$

are order statistics.

- **The Hill estimator** is a popular estimator of $1/\alpha$, (see [9],[6].) It is based on k upper-order statistics and is defined as follows:

$$H_{k,n} := \frac{1}{k}\Sigma_{i=1}^{k}log\frac{X_{(i)}}{X_{(k+1)}}.$$

To understand the idea behind it, suppose for a moment that instead of semi-parametric assumption (2), we have independent and identically distributed data from the Pareto parametric family:

$$1 - F(x) = P(X_i > x) = x^{-\alpha}, \quad x > 1, \quad \alpha > 0.$$

Thus F is a Pareto with support $[1, \infty)$ and corresponding density function is given by:

$$f(x) = \frac{\alpha}{x^{\alpha+1}}.$$

16

Then the maximum likelihood estimator for $1/\alpha$ is

$$\overline{\alpha}^{-1} = \frac{1}{n}\Sigma_{i=1}^{n}logX_i.$$

This follows from the fact that $logX_i, 1 < i < n$ is a random sample from the distribution with tail

$$P(logX_1 > x) = P(X_1 > e^x) = e^{-\alpha x}, \quad x > 0,$$

which is the exponential distribution tail. The mean of this distribution is α^{-1} and the Maximum Likelihood function is $\overline{X}$, which in this case is the given estimator.

The strong assumption about the Pareto distribution of the population in most cases needs to be weakened. It means that we assume a Pareto tail from some point onwards, as in model (1), rather than the exact model.

In practice we make a plot of the Hill estimator $H_{k,n}$ of $1/\alpha$ for different number of upper - order statistics k :

$$\left((k, H_{k,n}^{(-1)}), 1 \leq k \leq n\right).$$

Then, as long as the graph looks stable, we can pick out the value of α. Sometimes it is not easy. The difficulties when using the Hill estimator are the following:

1. It is sensitive to the choice of the number of upper order statistic k since we have to choose the estimate for α looking at the graph.

2. The graph may exhibit volatility, so finding the true value for α becomes difficult.

3. It gives optimal results for underlying distributions if they are close to the Pareto distribution. If they are not, then the error may be big.

4. The Hill estimator is not location invariant. A shift in location does not theoretically affect the tail index but it may seriously affect the Hill estimator.

- **The Pickands estimator** is based on the assumption that the distribution function F of the sample belongs to the family of extreme value

distributions, denoted by $D(G_\gamma)$. The extreme value distributions can be defined as one parameter family:

$$G_\gamma(x) = exp\left(-(1+\gamma x)^{-(1/\gamma)}\right), \quad \gamma \in R, \quad , 1+\gamma x > 0.$$

Now, if we consider support (S) of this family with respect to the parameter γ, we can observe the following:

$$S = \begin{cases} \left(-\frac{1}{\gamma}, \infty\right) & if \quad \gamma \geq 0, \\ ((-\infty, \infty)) & if \quad \gamma = 0, \\ \left(-\infty, \frac{1}{|\gamma|}\right) & if \quad \gamma \leq 0. \end{cases}$$

The heavy-tailed case corresponds to $\gamma \geq 0$, and then $\gamma = 1/\alpha$. The Pickands estimator is a semi-parametric estimator of γ. It uses differences of quantiles and is based on three upper-order statistics, $Z_{(k)}, Z_{(2k)}, Z_{(4k)}$ from a sample of size n. The estimator is given by the formula:

$$\overline{\gamma}_{k,n}^{Pickands} = \frac{1}{log2}log\left(\frac{Z_{(k)} - Z_{(2k)}}{Z_{(2k)} - Z_{(4k)}}\right).$$

The advantage over the Hill estimator is that it is location invariant. It is also scale invariant. If estimated parameter γ is negative or equal to zero, it indicates that the heavy tail model is inappropriate. The Pickands estimator is sensitive to the choice of number of upper order statistics k. The estimation of the parameter is read from its plot vs number of upper order statistics k, (see [6],[8]).

- **The Deckers - Einmahl - de Haan Estimator, (DEdH), for $\gamma = 1/\alpha$.** This estimator can be thought of as the extended Hill's estimator. Tt performs well for the class of extreme distribution functions. It is defined as follows:

$$\overline{\gamma}_{k,n}^{DEdH} = 1 + H_{k,n}^{(1)} + \frac{1}{2}\left(\frac{(H_{k,n}^{(1)})^2}{H_{k,n}^{(2)}} - 1\right)^{(-1)},$$

where
$$H_{k,n}^{(1)} := \frac{1}{k}\Sigma_{i=1}^{k}log\frac{X_i}{X_{k+1}}$$

18

is the Hill's estimator and

$$H_{k,n}^{(2)} := \frac{1}{k}\Sigma_{i=1}^{k}\left(log\frac{X_i}{X_{k+1}}\right)^2.$$

Because $H_{k,n}^{(1)}$ and $H_{k,n}^{(12)}$ can be interpreted as empirical moments, this estimator is also referred to as a moment estimator of γ, (see [8]).
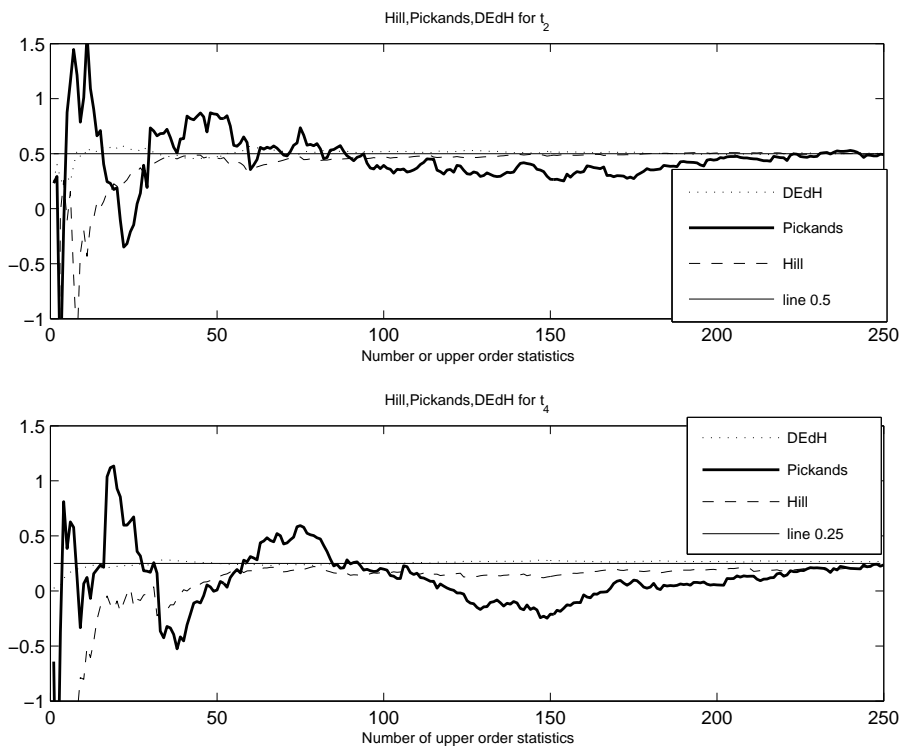

Figure 2.4: Hill, Pickands and DEdH estimators for $1/\alpha = 1/v$ for student t distribution with $v = 2, 4$ degrees of freedom.

Presented estimators are proven to be consistent. Except for Pickands estimators, all the others are applicable only in case of heavy tail distributions. To perform well, they require that $k(n) \to \infty$. It is not known how to choose k optimally for a finite sample.

In order to see how good these estimators are for the Student t distribution with $v = 2$ and $v = 4$, they are applied to 10000 random samples. The

samples are sorted in descending order. The $k = 250$ upper order statistics are used in estimation. Figure 2.4 shows the results.

It can be observed that the Pickands estimator is very noisy. The hill and DEdH estimators oscillate around the true value of $1/\alpha \sim 1/v$. Figure 2.5 shows how sensitive these estimators are to the data. The variation of these estimators over 100 simulations is presented by plotting the average values and their standard deviations. It can be observed that there are strong fluctuations when the number of upper ordered statistics goes to zero. It means that for low values of k the estimator should not be trusted. The plots of means of the estimators indicate the true values for $1/\alpha \sim 1/v$. The estimators are all biased.

In the literature, the Bootstrap method, jackknife method and variants of the Hill plot are proposed to improve the estimation, ([6],[7]). We shall take a close look at the bootstrap method for the Hill estimator.
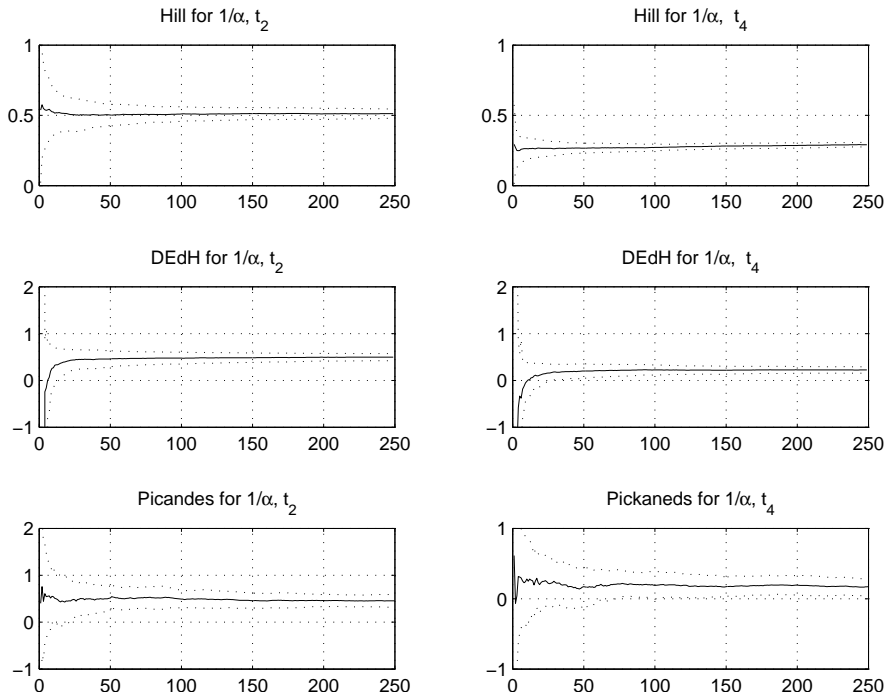

Figure 2.5: The variation of the Hill, Pickands and DEdH estimators over 100 samples from student t distribution with $v = 2, 4$ degrees of freedom.

The bootstrap method for Hill estimator.

   Picket, Dacorogna and Muller in [7], proposed a method of improving the performance of the Hill estimator. They suggested using bootstrap method to find an appropriate number of upper order statistics k^* and compute the Hill estimator of the tail index for this chosen k^*.

We describe their work and perform numerical simulations for the Student t distribution.

The Hill estimator is normally distributed. Authors of the method computed the expectation of the Hill estimator in asymptotic expansion and showed that it is biased:

$$E(H_{k,n}) = \frac{1}{\alpha} + B,$$

where B is the bias or systematic error. It is equal to:

$$B = E(H_{k,n}) - \frac{1}{\alpha}.$$

Further, they computed the variance of the Hill estimator in asymptotic expansion. These results were used to compute the error of the Hill estimator. The error has two components: the systematic error B and a stochastic error with the variance computed from:

$$E([H_{k,n} - E(H_{k,n})]^2).$$

The **total error** has the following variance:

$$E([H_{k,n} - \frac{1}{\alpha}]^2) = B^2 + E([H_{k,n} - E[H_{k,n}]]^2) =$$

$$= \frac{1}{\alpha^2}\frac{\beta^2 b^2}{(\alpha+\beta)^2}a^{-\frac{2\beta}{\alpha}}\left(\frac{k}{n}\right)^{\frac{2\beta}{\alpha}} + \frac{1}{\alpha^2 k}, \tag{2.4}$$

   where α, β, a, b are positive constants that were obtained while asymptotic expansion of the expectation and variance of the Hill estimator.

This total error is large both for large k (where the first term dominates) and very small k (where the second term dominates). The error becomes smaller in a middle of region for moderately small k. The total error of the Hill estimator can be minimized with respect to k. Alternatively it can be

constructed from the Hill estimator in order to obtain a bias - corrected estimator. The second approach is more difficult. Therefore the authors concentrated on the minimization of the Hill estimator with respect to k.

In order to find a minimum of the error given in (2.4) with respect to k, zero of the derivative of the error with respect to k is determined. The resulting $\overline{k}$ is given by:

$$\overline{k} = \left(\frac{\alpha(\alpha + \beta)^2}{2\beta^3 b^2} \right)^{\frac{\alpha}{\alpha+2\beta}} (a \times n)^{\frac{2\beta}{\alpha+2\beta}}. \tag{2.5}$$

It means that the expected value of the total error, $E([H_{k,n} - \frac{1}{\alpha}]^2)$ as a function of k, has a horizontal tangent at $\overline{k}$. Then the neighboring integer values are almost as good as $\overline{k}$.

The resulting $\overline{k}$ may be reinserted into the total error equation to give the minimum error variance that can be achieved by Hill estimator. However, the interest is only in the dependence of the error variance on the sample size n, and we have that $E([H_{k,n} - \frac{1}{\alpha}]^2)$ is proportional to $n^{-\frac{2\beta}{\alpha+2\beta}}$. Increasing the sample size n leads to smaller error of the Hill estimator.

The best k can not be computed from equation 2.5, because we do not know the parameters. In order to overcome this situation, authors follow approach introduced by Hall, [10]. He proposed using bootstrap re-samples of a small size n_1 and k_1 values, which differ from n to m. These new samples are then used for computing $\overline{H_{k_1,n_1}}$. Hall suggested finding the optimal $\overline{k_1}$ for the subsamples from

$$min_{k_1} E([H_{k_1,n_1} - H_0]^2 | F_n), \tag{2.6}$$

where F_n is the empirical distribution and $H_0 = H_{k_0,n}$ is some initial full sample estimate for $1/\alpha$ with a reasonably chosen but non- optimal k_0. Equation (2.6) is a new approximate version of equation (2.4).
The quantity H_0 is a good approximation of $1/\alpha$, for subsamples, since we know that the error is proportionally larger for n_1 than for n observations. The value $\overline{k_1}$ is found by recomputing $\overline{H_{k_1,n_1}}$ for different values of k_1 and then empirically evaluating and minimizing equation (2.6). Given $\overline{k_1}$, the $\overline{k}$ for the full sample size can be found by taking the ratio $\frac{\overline{k}}{k_1}$ and using the equation (2.4). It results in the following:

$$k* = \overline{(k_1)} \left(\frac{n}{n_1} \right)^{\frac{2\beta}{2\beta+\alpha}}, \tag{2.7}$$

22

where $k*$ is the bootstrap estimate of $\overline{k}$ and $\alpha* = 1/H_{k*,n}$ is the final estimate of the tail index α.

To apply this procedure, we need to know the parameters α and β. For α_0 we use the initial estimate by means of H_0, that is $1/H_0$. The authors of the article [7] used in their investigation $\beta = 0.01, 0.5, 1, 2$. As it turned out in their studies, the results of the estimation is insensitive to the choice of this parameter, among those that were tested.

In order to test the quality of the method for estimating parameter of degrees of freedom for the Student t distribution, the random samples are taken from the Student t distribution with parameter $v = 2, 3, 4, 5, 6$. The sample sizes are $1000, 5000, 10000$. We apply the estimating method to each sample for each parameter degrees of freedom. In order to improve the method, also negative tails are considered. The final estimation is the average taken from the estimation of the right and left tail indexes. To obtain results independent of the particular choice of the initial seed of the random generator, we take the average results computed over 10 realizations for t distribution. The results presented in Table 1 are calculated with the bootstrap method under the following conditions:

- For the initial H_0 in equation (2.6), the k_0 value is chosen as 10% of the full sample size n.

- The re-sample size is chosen as $n_1 = n/40$. The number of re-samples is 100. The re-samples are randomly picked from the full data set. One data point can occur in several re-samples.

- The search space for k_1 in order to find $\overline{k_1}$ through the equation (2.6) is restricted to 10% of subsample size of n_1.

- We apply the equation (2.7) with $\alpha = 1/H_0$ and $\beta = 0.1$ to obtain $k*$ for the full sample size.

- The 95% confidence interval is computed from the stochastic variance, and is given by: $\pm 1.96[\frac{1}{\alpha*^2 k*}]^{1/2}$.

Table 1.

Sample	$v = 2$	$v = 3$	$v = 4$	$v = 5$	$v = 6$
1000	2.05 ± 0.67	3.14 ± 0.44	3.8 ± 0.36	4.5 ± 0.45	4.9 ± 0.26
5000	1.97 ± 0.28	2.83 ± 0.21	3.7 ± 0.15	4.5 ± 0.13	4.7 ± 0.12
10000	2.09 ± 0.2	2.97 ± 0.13	3.93 ± 0.1	4.5 ± 0.09	5.3 ± 0.08

Table 1 above presents estimated tail indexes for the Student t distribution with different degrees of freedom for different sample sizes. As expected the results are better when the sample size increases. The estimation is not very accurate for small samples, especially when the degrees of freedom parameter of theoretical distribution increases. However, we can always use a goodness of fit test for population with estimated parameters to confirm the results. We apply this methodology to the foreign exchange rates in the next section.

2.1.2   Foreign Exchange rates

Knowledge of the distribution of the exchange rates is important in studies where the uncertainty regarding exchange rates movements must be measured, for instance, in modeling the foreign exchange transaction costs.
Many analysts find the Student t distribution to be among others a good model for exchange rates ([11],[12]). We will see, that indeed these data have heavy tails and that we can use tail index estimators to fit an appropriate Student t distribution.

We examine three daily foreign exchange rates versus the U.S. dollar: Canadian dollar, German mark and Swiss francs for the time period 2 of January 1973 to 8 of August 1984. The samples size is 2909. No special adjustment was made for holidays, so the differencing interval between daily series represents changes between business days. In the analysis that follows we model returns of the exchange rates. First, define the return process $\overline{R_i}$ as:

$$\overline{R_i} = (S_i - S_{i-1})/S_{i-1},$$

where (S_i) is the stochastic process representing the exchange rate at times i. It gives the relative difference of rates at each time point i. If the returns are small, then the differenced log - price process approximates the return process

$$R_i = \log S_i - \log S_{i-1} = log\frac{S_i}{S_{i-1}} = log\left(1 + \left(\frac{S_i}{S_{i-1}} - 1\right)\right)$$

$$\approx \frac{S_i}{S_{i-1}} - 1 = \overline{R_i},$$

what follows from the fact that for $|x|$ small,

$$log(1 + x) \sim x, \quad x \to 0.$$

24

Therefore, now we call the process (R_i) the returns process. This process has stationary property. It is scale free and independent of the units as well as the initial rate value.

We transform the data to obtain the return process R_i. Figure 2.6 presents the time series plots of daily exchange rates and corresponding return series for the three datasets.
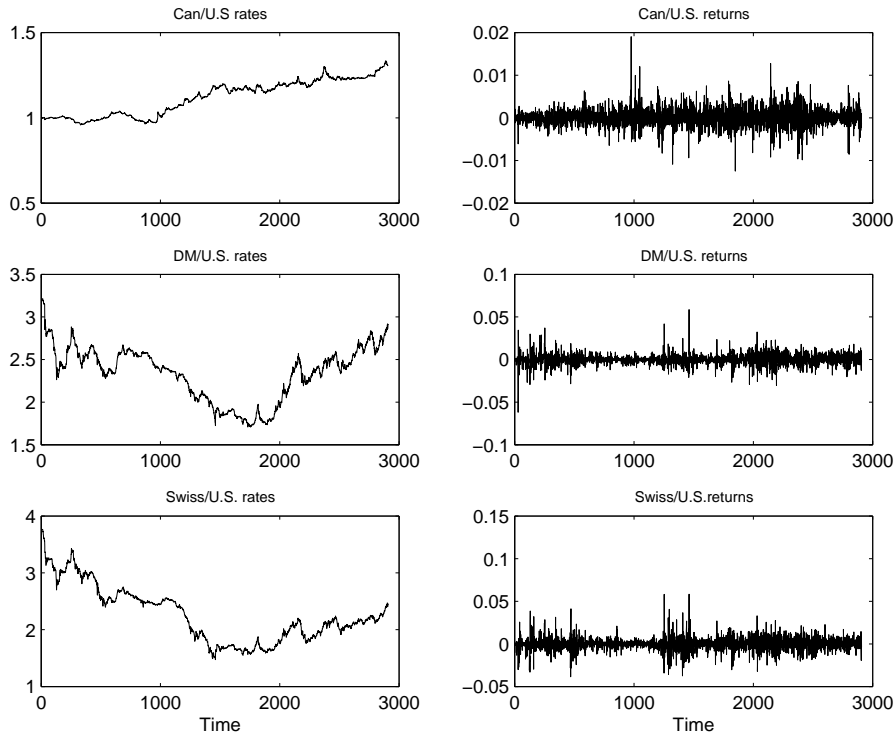

Figure 2.6: Plots of the daily exchange rates and returns of the Canadian dollar, German Mark and Swiss francs vs American dollar.

The following descriptive statistics were calculated for all datasets: location[1]

---

[1]Location parameter-μ, determines the origin of the underlying distribution.

(mean), scale [2] (standard deviation), skewness [3] and kurtosis [4]. Table 2 below presents values of these statistics for three datasets:

	location,	scale	skewness	kurtosis
Candian dollar/U.S.dollar	9.24e-005	0.0022	0.32	5.02
German Mark/U.S.dollar	-3.37e-005	0.0064	-0.015	9.3
Swiss Franc/U.S. dollar	-1.48e-004	0.0076	0.23	5.76

We can observe that on average the Canadian dollar declined against U.S. during the sample period - its mean is positive. The German mark and Swiss Franc rose on average, both have negative mean. There is a slight positive skewness for the Canadian dollar and Swiss francs. For German mark skewness is small and negative. Since skewness is small, we can assume symmetry of the distributions, because it is not strongly contradicted by the data. All three currencies have positive excess kurtosis. This means that their densities are more peaked and heavy tailed than normal density. We compute the tail indexes α for them using estimators described in the previous section. It will give an estimation of the degrees of freedom parameter v for Student t distribution for each dataset.

Figure 2.7 presents plots of the $1/(Hillestimator)$ and $1/(DEdHestimator)$.

---

[2]Scale parameter, standard deviation-σ, indicates the concentration of the underlying density around the origin.

[3]Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. The skewness of a distribution is defined as

$$S = \frac{E(x - \mu)^3}{\sigma^3},$$

where μ is the of x and σ is the standard deviation of x. Skewness for symmetric distributions, such as normal and Student T distribution should be equal to zero.

[4]Kurtosis is a measure of how outliers-prone a distribution is. Higher kurtosis means more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations. A distribution with positive kurtosis has a higher probability than a normally distributed variable of values near the mean. It also has heavy tails, that is a higher probability than a normally distributed variable of extreme values. A distribution with negative kurtosis has opposite properties. Kurtosis is defined as follows:

$$K = \frac{E(x - \mu)^4}{\sigma^4} - 3,$$

where μ is the of x and σ is the standard deviation of x. Kurtosis defined in this way is also called excess kurtosis. The excess kurtosis of the normal variable equals 0.
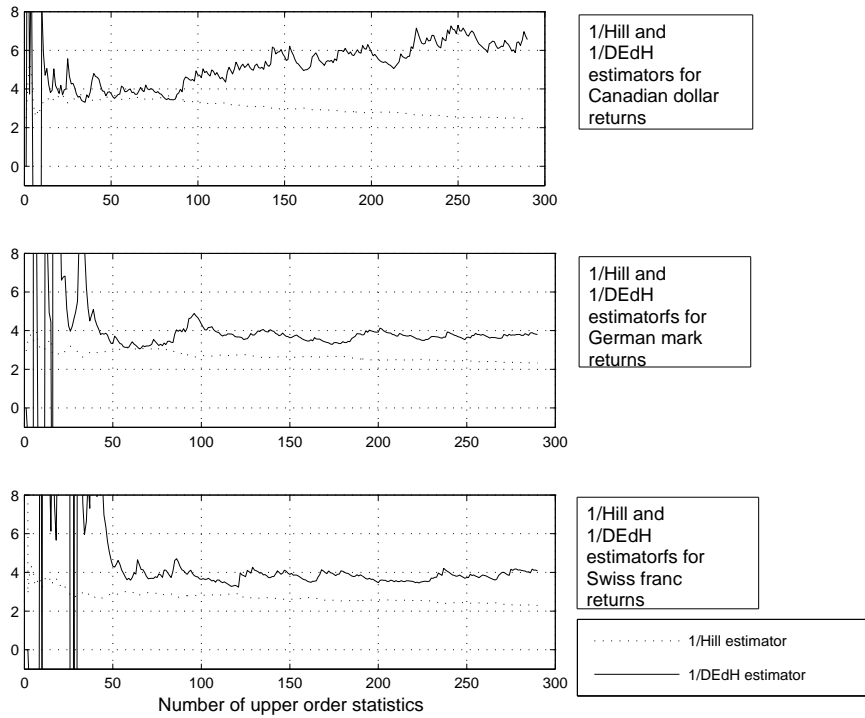
Figure 2.7: $1/Hill$ and $1/DEdH$ estimators of the degrees of freedom parameter for Canadian dollar, German mark and Swiss franc.

The Pickand estimator is very noisy and uninformatively and therefore it was not plotted.

The value of parameter $\bar{v}$ for Canadian dollar indicated by the Hill and DEdH estimators is in the interval $[3.5, 4]$ when number of upper order statistics is smaller than 100. Estimators depart from each other as the number of upper order statistics is greater than 100. The Hill estimator decreases towards 2 and DEdH estimator increases to over 6.

Estimators behave in stable ways for German mark and Swiss franc and indicate the values of the degrees of freedom parameter $\bar{v}$ in the interval $[2, 4]$. It is not easy to pick one value for the degrees of freedom parameter, therefore, we apply the bootstrap method for Hill estimator. It will give us just one value. Next, we perform the Kolmogorov-Smirnov goodness of fit test[5]

---

[5]Kolmogorv-Smirnov goodness of fit test is used to determine whether two underlying one-dimensional probability distributions differ, or whether an underlying probability dis-

for Student t distribution with estimated parameter $\bar{v}$. Additionally, we also test null hypothesis that the currency returns are Student t distributed with the degrees of freedom parameter from the interval $[2, 8]$.

We apply Kolmogorv-Smirnov test to the standardized returns data, see [13]. It means that we divide $T - \mu$ by $\sqrt{var(T)}$, where T is random variable representing our random sample of size N.

If $t_1, \ldots, t_N$ is a sequence of observations, then the unbiased estimators for μ is computed from:

$$\overline{T} = \frac{1}{N} \sum_{1=i}^{N} (t_i)$$

The unbiased estimator for variance for the univariate t distribution, when parameter v is assumed to be known, is given by the formula:

$$V = \frac{(v-2)}{v} \frac{1}{(N-1)} \sum_{1=i}^{N} (t_i - \overline{T})(t_i - \overline{T}).$$

In terms of scale parameter:

$$V = \frac{(v-2)}{v} \sigma^2.$$

We use this transformation with the estimated degrees of freedom parameter $\bar{v}$ in order to obtain a sample from a univariate canonical distribution with $\bar{v}$ degrees of freedom. Then we use the Kolmogorov Smirnov test to confirm whether, data is t distributed with estimated degrees of freedom parameter.

Table 3 below shows the results:

	Canadian dollar	German mark	Swiss franc
$\bar{v}$	4.3 ± 0.19	3.2 ± 0.22	4.1 ± 0.2
Accepted v :	$[3.7, 7.6]$	$[3, 4.5]$	$[3.8, 6.2]$

As we can see, the Kolmogorov Smirnov test accepted null hypothesis that Canadian dollar, German mark and Swiss franc returns are t distributed with estimated degrees of freedom parameter. The null hypothesis for degrees of

---

tribution differs from a hypothesized distribution, in either case based on finite samples. It uses the statistics:

$$max(|F(x) - F_e(x)|),$$

where $F(x)$ is hypothetic distribution function - null hypothesis and $F_e(x)$ is empirical distribution function of the sample.

freedom parameters relatively close to the estimated one can not be rejected either.

We can conclude that datasets are Student t distributed with parameters v equal to the results of the improved Hill estimators for each dataset. Student t distribution is a reasonable statistical model for foreign exchange rates.

Figure 2.8 presents density histograms of the three currencies returns. It shows density histograms of rescaled data together with plots of fitted Student t density functions with estimated parameters $\overline{v}$.
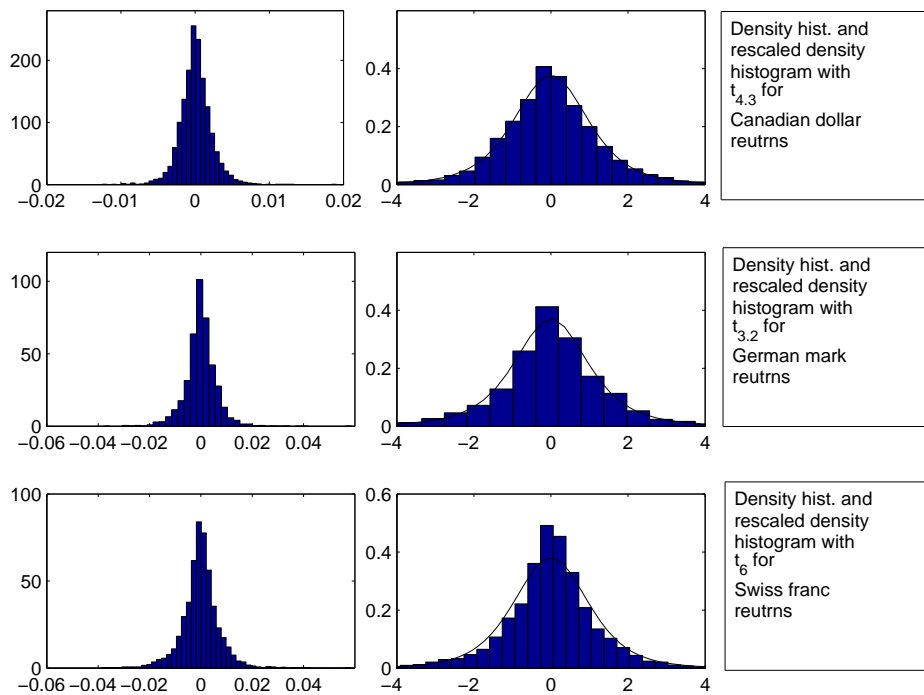

Figure 2.8: Density histograms, rescaled density histograms and fitted t densities functions for returns of the Canadian dollar, the German Mark and the Swiss Francs vs American dollar.

2.2 Multivariate Student t distribution

This section is devoted to the multivariate Student t distribution. Many different constructions of this distribution are available in the literature. They have different features and properties. Kotz and Nadarajah collected most of them in the book *Multivariate t Distributions and Their Applications*, see [14]. We describe here the most common and natural form, that directly generalizes the univariate Student t distribution. We call it canonical multivariate Student t distribution.

We devote a separate subsection for the bivariate case. We rewrite some of the general results for the bivariate Student t distribution such as conditional distribution. It will be used later on to obtain the conditional copula, which is of practical importance. We will also present the T distribution derived by William Shaw with the independence property.
Let us start with defining a multivariate canonical Student t distribution.

The p-dimensional random vector $T = (T_1, \ldots, T_p)$ is said to have multivariate t distribution ([15]) with v degrees of freedom, mean vector $\mu = (\mu_1, \ldots, \mu_p)^T$, location parameter, and positive definite correlation matrix R^6, if its density is given by

$$f_{R,v}(\mathbf{t}) = \frac{\Gamma(\frac{v+p}{2})}{\Gamma(\frac{v}{2})(\pi v)^{p/2}|R|^{1/2}} \left(1 + \frac{(\mathbf{t} - \mu)^T \times |R|^{-1}(\mathbf{t} - \mu)}{v} \right)^{-(\frac{v+p}{2})}. \quad (2.8)$$

We denote this by $T \sim td_p(v, \mu, \mathbf{R})$. The cumulative distribution function will be denoted by $t_p(v, \mu, R)$.
The degrees of freedom parameter v is also referred to as the shape parameter. The peakedness of the density may be diminished, preserved or increased by varying v. The distribution is said to be central if $\mu = 0$, otherwise is said to be non-central. The particular case of (2.8) for $\mu = 0$ and $R = I$ is a mixture of the normal density with zero means and covariance matrix vI - in the scale parameter v.

The Student t distribution (2.8) can also be written as

---

[6]For the purposes of this report R is assumed to be the correlation matrix. In general, however, it is a positive definite matrix of scale parameters, sometimes called also a dispersion matrix. When multiplied by $\frac{v}{v-2}$ it becomes a covariance matrix. To get the correlation matrix, the covariance matrix Σ is multiplied by the following matrix D, $D\Sigma^* D$. On the diagonal of the D, $\frac{1}{\sigma x}$ is put and the rest of the entries are zeros.

$$f_{R,v}(\mathbf{t}) = \int_0^\infty \frac{|\omega^2 R|^{-\frac{1}{2}}}{(2\pi)} exp\left(-(\mathbf{t}-\mu)^T(\omega^2 R)^{(-1)}\left(\frac{\mathbf{t}-\mu}{2}\right)\right)h(\omega)d\omega \quad (2.9)$$

which is a mixture of the multivariate normal distribution $N_p(\mu, \omega^2 R)$ and ω has the inverted gamma distribution with probability density function:

$$h(\omega) = \frac{2(\frac{v}{2})^{\frac{v}{2}}}{\Gamma(\frac{v}{2})}\omega^{-(v+1)}exp\left(\frac{-v}{2\omega^2}\right)$$

where v is the degree of freedom of inverted gamma distribution. Equivalently, $v\omega^{-2}$ has a chi-square distribution with v degrees of freedom. Thus for a given ω, the random vector T has a multivariate normal distribution

$$(T|\Omega = \omega) \approx N_p(\mu, \omega^2 R).$$

As $v \to \infty$, the random variable Ω becomes a degenerate random variable with all the non-zero mass at the point unity and the probability density function of the multivariate T distribution in (2.8) converges to that of the multivariate normal distribution $N_p(\mu, R)$. The uncorrelatedness of the $T_1, \ldots, T_p$ does not imply that they are independent of each other, unless $v \to \infty$. The Student t distribution does not posses the independence property.
Taking $\mu = [0,0]^T$, the equation (2.8) simplifies to

$$f_{R,v}(\mathbf{t}) = \frac{\Gamma(\frac{v+p}{2})}{\Gamma(\frac{v}{2})(\pi v)^{p/2}|R|^{1/2}}\left(1 + \frac{\mathbf{t^T}|R|^{-1}\mathbf{t}}{v}\right)^{-(\frac{v+p}{2})}. \quad (2.10)$$

When the correlation matrix R is an identity matrix I, then the expression above simplifies to

$$f_{R,v}(\mathbf{t}) = \frac{\Gamma(\frac{v+p}{2})}{\Gamma(\frac{v}{2})(\pi v)^{p/2}}\left(1 + \frac{\mathbf{t^T t}}{v}\right)^{-(\frac{v+p}{2})}, \quad (2.11)$$

which is not a product of the p univariate Student t density functions. Because of the lack of the independence property, it is more appropriate for some applications to use the uncorrelated t model , which is presented later on.

31

2.2.1 Representation

The multivariate Student t distribution with v degrees of freedom, mean vector μ and correlation matrix R can be represented at least in three different ways. Each representation results in the canonical multivariate Student t distribution. Representation number 1 is more convenient for computing moments of the t distribution, whereas representation 3 is used to sample the t distribution. They are shown below, see [14],[22].

1. $\mathbf{Y}$ is a p-variate normal random vector with mean 0 and covariance matrix R, and if vS^2/σ^2 is the chi-squared random variable with degrees of freedom v, independent of $\mathbf{Y}$, then

$$\mathbf{T} = S^{-1}\mathbf{Y} + \mu \qquad (2.12)$$

   is a Student t distributed random variable with parameters μ, v, R. This implies that $\mathbf{T}|S = s$ has a p-variate normal distribution with mean vector μ and covariance matrix $(1/s^2)R$.

2. If $\mathbf{V}^{1/2}$ is the symmetric square root of $\mathbf{V}$, that is,

$$\mathbf{V}^{1/2}\mathbf{V}^{1/2} = \mathbf{V} \sim W_2(R^{-1}, v+2-1),$$

   where $W_2(\Sigma, n)$ denotes the p-variate Wishart distribution [7] with degrees of freedom n and covariance matrix Σ, and if $\mathbf{Y}$ has the p-variate normal distribution with zero means and covariance matrix vI_p (I_p is the p-dimensional identity matrix), independent of $\mathbf{V}$, then

$$\mathbf{T} = (\mathbf{V}^{1/2})^{-1}\mathbf{Y} + \mu.$$

   That implies that $\mathbf{T}|\mathbf{V}$ has p-variate normal distribution with mean vector μ and covariance matrix $v\mathbf{V}^{-1}$.

---

[7]Wishart distribution- is an multivariate analogue to a chi-squared distribution. If $x_1, \ldots, x_k$ are a sequence of independent identically distributed random variables, each having distribution $N(0, \sigma^2)$, then $1/\sigma^2 \Sigma x_i^2$ has a chi-square distribution with k degrees of freedom. The multivariate analogue occurs when $x_1, \ldots, x_k$ form a sequence of independent p-variable random vectors each with distribution $N(0, \Sigma)$ and the matrix C is defined by $C = \Sigma_{i=1}^k x_i x_i^T$. The $(i, i)^{th}$ element of the matrix C is the sum of squares of the ith elements of the vectors $x_1, \ldots, x_k$, while the $(i, j)^{th}$ element of $C, i \neq j$ is the sum of products of ith and j^{th} elements of these vectors. The joint distribution of all elements of C is said to be a Wishart distribution based on p variables, with k degrees of freedom and parameter Σ. It is denoted by $W_p(k, \Sigma)$, see [16]

3. If **T** has the stochastic representation

$$\mathbf{X} = \mu + \frac{\sqrt{v}}{\sqrt{S}}\mathbf{Z}, \qquad (2.13)$$

where $\mu \in R^p$, $S \sim \chi_v^2$ and $\mathbf{Z} \sim \mathbf{N_p}(\mathbf{0}, \mathbf{R})$ are independent, then X has an p-variate Student t distribution with mean μ and covariance matrix $\frac{v}{v-2}R$. This representation is commonly used for sampling purposes. The sampling procedure will be provided later on.

2.2.2 Moments

Since variables $\mathbf{Y}$ and S in the representation 1 (2.12) are independent, the conditional distribution of (T_i, T_j), given $S = s$, is the bivariate normal with means (μ_i, μ_j), common variance σ^2/s^2 and correlation coefficient $r_{i,j}$. Thus,

$$
\begin{aligned}
E(T_i) &= E[E(T_i|S = s)] \\
&= E(\mu_i) \\
&= \mu_i.
\end{aligned}
$$

We use the classical identity

$$Cov(X_i, X_j) = E[Cov(X_i, X_j)|S = s] + Cov[E(X_i|S = s)E(X_j|S = s)],$$

for $i, j = 1, \ldots, p$, in order to calculate the second moments. Using the assumptions above, we can write:

$$E[Cov(T_i, T_j)|S = s] = \sigma^2 r_{i,j} E\left(\frac{1}{S^2}\right)$$

and

$$Cov[E(T_i|S = s)E(T_j|S = s)] = 0.$$

$E(\frac{1}{S^2})$ exists if $v > 2$. It is computed by integrating the chi square distribution in representation (2.1) multiplied by $\frac{1}{s^2}$. It equals to

$$\frac{v}{\sigma^2(v-2)}.$$

By choosing $i = j$ and $i < j$, respectively, one obtains

$$Var(T_i) = \frac{v}{v-2}$$

and

$$Cov(T_i, T_j) = \frac{v}{v-2} r_{i,j}.$$

(For further results see [14].)

2.2.3 Characteristic function

Literature provides characteristic functions for univariate and multivariate t distributions derived by different authors. We present a derivation due to Joarder and Ali (1996), which is relatively recent one, [14].

The characteristic function of $\mathbf{T}$ following a multivariate Student t distribution (2.8) is given by

$$\phi_{\mathbf{T}}(t) = e^{(it^{\mathbf{T}}\mu)} \frac{||\sqrt{vR}\mathbf{t}||^{v/2}}{2^{v/2-1}\Gamma(v/2)} K_{v/2}(||\sqrt{vR}\mathbf{t}||), \qquad (2.14)$$

where $||\mathbf{t}|| = \sqrt{(\mathbf{t}^{\mathbf{T}}\mathbf{t})}$ and $K_{v/2}(||\sqrt{vR}\mathbf{t}||)$ is the MacDonald function with order $v/2$ and argument $(||\sqrt{vR}\mathbf{t}||)$.
An integral representation of the MacDonald function is

$$K_\alpha(t) = \left(\frac{2}{t}\right)^\alpha \frac{\Gamma(\alpha+1)}{\sqrt{\pi}} \int_0^\infty (1+u^2)^{-(\alpha+1/2)} cos(tu) du,$$

where $t > 0$ and $\alpha > -1/2$.
For the univariate Student t distribution with the density function (1.1) the characteristic function in terms of the MacDonald function is of the form:

$$\phi_X(t) = \frac{v^{v/4}|t|^{v/2}}{2^{v/2-1}\Gamma(v/2)} K_{v/2}(\sqrt{v}t).$$

For details see [14].
The characteristic function of $\mathbf{T}$ in (2.14) can be written as follows:

$$\phi_{\mathbf{T}}(t) = E(e^{i\mathbf{t}^T T}) = e^{i\mathbf{t}^T \mu} \psi(\mathbf{t}^T R\mathbf{t}), \qquad (2.15)$$

for some function ψ.

34

2.2.4 Elliptical Distribution

The class of the elliptical distributions can be defined in a number of different ways. We present one of the possible definitions.

Definition 1 *If $\mathbf{X}$ is a p-dimensional random vector and, for some $\mu \in R^p$ and some $p \times p$ nonnegative definite symmetric matrix Σ, the characteristic function $\phi_{\mathbf{X}-\mu}(\mathbf{t})$ of $\mathbf{X} - \mu$ is a function of the quadratic form $\mathbf{t^t \Sigma t}$, $\phi_{\mathbf{X}-\mu}(\mathbf{t}) = \psi(\mathbf{t^T t})$, we say that $\mathbf{X}$ has an elliptical distribution with parameters μ, Σ and ψ, and we write $\mathbf{X} \sim E_p(\mu, \Sigma, \psi)$.*

Therefore, elliptical distribution is a family of distributions whose characteristic function takes the form:

$$\phi_X(t) = e^{i\mathbf{t^T}\mu}\psi(\mathbf{t}^T\Sigma\mathbf{t}),$$

for some function ψ, vector μ and nonnegative definite matrix Σ, see [17]. Looking at the characteristic function given in (2.14), we can see that Student t distribution belongs to the elliptical distribution family and therefore enjoys some useful properties of this family, such as:

1. Marginal distributions of elliptically distributed variables are elliptical.

2. Conditional distribution of X_1 given the value of X_2 is also elliptical, but in general not of the same type.

3. Any linear combination of elliptically distributed variables is elliptical.

4. For multivariate elliptical distributions partial and conditional correlations are equal. Zero conditional correlation does not necessarily indicate conditional independence. This is the property of the normal distribution.
   Baba, Shibata and Sibuya in [20] provided a necessary and sufficient conditions for coincidence of the partial covariance with the expectation of the conditional covariance. Essentially, that is the linearity of the conditional expectation. A corollary of their main result shows that the partial correlation is equal to the conditional correlation if the conditional correlation is independent of the value of the condition, and also if the conditional expectation is linear.
   They showed that this holds for elliptical distributions and therefore for the Student t distribution.

35

5. The relation between product moment correlation and Kendall's tau for bivariate elliptical distribution with correlation ρ is given by the formula:
$$\rho = \sin\left(\frac{\pi}{2}\tau\right).$$

This is a useful result especially for statistical purposes. It can be used to build a robust estimator of linear correlation for elliptically distributed data, see [19]

For proves of these result, we refer to [22], [17],[18].

2.2.5 Marginal Distributions

Suppose that $\mathbf{T}$ is a Student t distributed with parameters μ,v and correlation matrix R. They can be partitioned as follows:

$$\mathbf{T} = \begin{pmatrix} \mathbf{T_1} \\ \mathbf{T_2} \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \mathbf{R} = \begin{pmatrix} R_{1,1} & R_{1,2} \\ R_{2,1} & R_{2,2} \end{pmatrix},$$

where $\mathbf{T_1}$ is $p_1 \times 1$ and $R_{1,1}$ is $p_1 \times p_1$. Then T_1 has the p_1-variate t distribution with v degrees of freedom, mean vector μ_1 and correlation matrix $R_{1,1}$ with the joint density function given by

$$f_{\mathbf{T_1}}(\mathbf{t_1}) = \frac{\Gamma((v+p_1)/2)}{(v\pi)^{p_1/2}\Gamma(v/2)|R_{1,1}|^{1/2}} \left(1 + \frac{1}{v}(\mathbf{t_1} - \mu_1)^T R_{1,1}^{-1}(\mathbf{t_1} - \mu_1)\right)^{-(v+p_1)/2}.$$

Moreover, X_2 also has the $(p - p_1)$- variate t distribution with v degrees of freedom, mean vector μ_2, correlation matrix $R_{2,2}$, and with joint density function

$$f_{\mathbf{T_2}}(\mathbf{t_2}) = \frac{\Gamma((v+p-p_1)/2)}{(v\pi)^{p-p_1/2}\Gamma(v/2)|R_{2,2}|^{1/2}} \left(1 + \frac{1}{v}(\mathbf{t_2} - \mu_2)^T R_{2,2}^{-1}(\mathbf{t_2} - \mu_2)\right)^{-(v+p-p_1)/2},$$

see [14].

2.2.6 Conditional Distribution

In order to provide a conditional distribution we use now a slightly different notation. Instead of the correlation matrix R, we consider a covariance matrix defined as $\frac{v}{v-2}R = v * R$, where $v* = \frac{v}{v-2}$. Then we denote the t

distribution with these altered parameters as $t_p(v, \mu, v * R)$. With the above partition, the conditional distribution of $\mathbf{T_1}$ given $\mathbf{T_2} = \mathbf{t_2}$ has the p_1-variate Student t distribution with mean $\mu_{\mathbf{1.2}}$, and covariance matrix $v *_{1.2} R_{1,1.2}$ given by

- $\mu_{1.2} = \mu_1 + R_{1,2} R_{2,2}^{-1} (\mathbf{t_2} - \mu_{\mathbf{2}})$ is conditional mean,

- $v_{1.2} = \frac{v}{v+p-p_1+2}$,

- $R_{1.2} = (1 + (\mathbf{t_2} - \mu_{\mathbf{2}})^T (vR_{2,2})^{-1} (\mathbf{t_2} - \mu_{\mathbf{2}})) R_{1,1.2}$ is the conditional correlation matrix, where

$$R_{1,1.2} = R_{1,1} - R_{1,2} R_{2,2}^{-1} R_{2,1}.$$

For the details of this result see [14] and references in [15].

2.2.7 Distribution of a Linear Function

If $\mathbf{T}$ has a multivariate t distribution with degrees of freedom v, mean vector μ, and correlation matrix R, then for any nonsingular scalar matrix A and for any b, the variable

$$AT + b$$

is multivariate Student t distributed with degrees of freedom parameter v, mean vector $A\mu + b$ and correlation matrix AXA^T. The degree of freedom parameter remains the same, see [14], [15].

2.2.8 Dependence Measures

An important issue in practical applications is to determine the dependence between variables. To do so, several measures are applied. The best known is the following:

Definition 2. **Product Moment Correlation** *of random variables X, Y with finite expectations $E(X), E(Y)$ and finite variances σ_X^2, σ_Y^2 is:*

$$\rho(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}$$

This coefficient is a measure of linear dependence between two variables. The bivariate T distribution can be parameterized by this coefficient.

There are also other dependence coefficients, which are based on ranks of the variables. In contrast to the product moment correlation, they are independent of the marginal distributions and invariant under continuous increasing transformations.

Definition 3. Rank Correlation -Spearman's ρ_r *of random variables* X, Y *with cumulative distribution functions* F_X *and* F_Y *is defined as*

$$\rho_r(X, Y) = \rho(F_X(X), F_Y(Y)).$$

There is no explicit relation between rank correlation and linear correlation for elliptical distributions, except for normal one. It is possible, though, to compute analytically the rank correlation corresponding to a given linear correlation. We are not showing such calculations here. Instead we explain this relationship vie Monte Carlo simulations.

1. 1000 samples are taken for each ρ from the student T distribution with fixed v.

2. The rank correlation is computed for each sample;

3. Steps 1 and 2 are repeated 1000 times;

4. The average of the computed rank correlations is taken for each product moment correlation.

Figure 2.9 shows the results. It looks like the rank correlation does not depend on the degrees of freedom parameter. The plot of the rank correlation, solid line, lies below the product moment correlation. It tends to be a bit lower than the corresponding linear correlation.

Another popular measure of dependence is Kendall's tau, τ. The following definition of τ is taken from the [22].

Definition 3. **Kendall's** τ *for a random vector* (X, Y) *is defined as follows:*

$$\tau(X, Y) = P(X - \overline{X})(Y - \overline{Y}) > 0 - P(X - \overline{X})(Y - \overline{Y}) < 0,$$
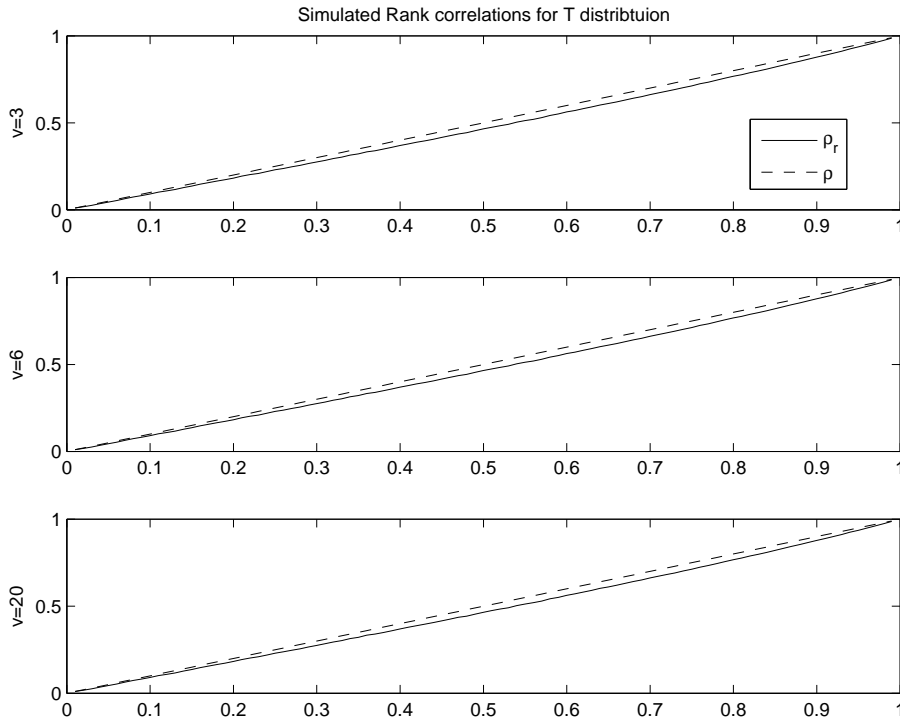
Figure 2.9: Rank Correlations for the T distribution, obtained by the Monte Carlo method.

where $(\overline{X}, \overline{Y})$ is an independent copy of (X, Y), and X, Y are random vectors of length n.

For elliptical distributions we know the relation between τ and ρ, that was given in section 2.2.4. We can find τ from samples.

The sample estimator $\overline{\tau}$ is given by the following. In a sample of length n there are $\frac{n(n-1)}{2}$ pairs of points $\{(x_i, y_i), (x_j, y_j)\}$, where a point can not be paired with itself and two points in either order count s one pair. Let c denote the number of pairs such that $(x_i - x_j)(y_i - y_j) > 0$, concordant pairs, and d the number of pairs such that $(x_i - x_j)(y_i - y_j) < 0$, discordant pairs. Then the estimator for τ is given by:

$$\overline{\tau} = \frac{c - d}{c + d}.$$

In practice however, it might occur that $x_i = x_j$ and/or $y_i = y_j$. In this case, tie, the pair is neither concordant nor discordant. If there is a tie in $x's$ the

39

pair will be called an "extra y pair", e_y, and an "extra x pair", e_x, if there is a tie in the $y's$. The adjusted sample version of Kendall's tau is then given by

$$\overline{\tau} = \frac{c - d}{\sqrt{c + d + e_y}\sqrt{c + d + e_x}}.$$

It is an unbiased estimator of τ. Since the $\overline{\tau}$ is an unbiased estimator of τ and from the relation

$$\rho = \sin\left(\frac{\pi}{2}\tau\right),$$

the we have that

$$\overline{\rho} = \sin\left(\frac{\pi}{2}\overline{\tau}\right).$$

is a natural estimator of linear correlation for elliptical distributions, and therefore for Student t distribution. This estimator is, however, not unbiased, but it is an robust estimator of the linear correlation. This result is of great practical importance and will be used in order to estimate the correlation matrix for T copulas later on. For details see [23]

2.2.9   Uncorrelated T Model

In practice, one may be confronted with the situation where the observed data has a symmetrical distribution with tails which are fatter than those predicted by the normal distribution. In such cases, we can use the multivariate t model. We introduce here the independent and uncorrelated t model. Since we want sometimes to estimate the scale parameters matrix Σ instead of the correlation matrix R, we present the uncorrelated t model with scale parameter matrix Σ. We will give also the unbiased estimators for Σ and mean.

The joint probability density function of k independent observations that have p-variate t distribution with v degrees of freedom, mean μ and the covariance matrix Σ can be written as:

$$f(t_1, t_2, \ldots, t_k) = f(t_1)f(t_2)\ldots f(t_k).$$

We may call it independent t model. However, many authors (see references in [14],[15],[21]) found it more instructive to consider dependent but

uncorrelated t distributions model:

$$f(t_1,\ldots,t_k) = \frac{\Gamma((v+p)/2)}{(\pi^k v)^{p/2}\Gamma(v/2)|\Sigma|^{k/2}}\left(1 + \frac{1}{v}\Sigma_{i=1}^k(t_i-\mu_i)^T\Sigma^{-1}(t_i-\mu_i)\right)^{-(v+kp)/2}$$

(2.16)

It was proven that the tails of this model are thicker than those of the independent t model. As degrees of freedom parameter $v \to \infty$, the observations in uncorrelated model are independent and the uncorrelated t model becomes a product of k independent p-dimensional random variables each having normal distribution $N_p(\mu, \Sigma)$, see [15] and [21]. This model can be used to obtain the likelihood estimators of the parameters μ and Σ, given in the next section.

Parameters Estimation for One Population.

The maximum likelihood estimators of the μ and Σ of the uncorrelated t-model in (2.16) are given by

$$\mu = \overline{T} = \frac{1}{n}\Sigma_{i=1}^n T_i$$

and

$$\Sigma = A/n = \left(\Sigma_{i=1}^n(T_i-\overline{T})(T_i-\overline{T})\right)/n,$$

where n is the sample size and $A\Sigma_{i=1}^n(T_i-\overline{T})(T_i-\overline{T})$ is the sum of product matrix based on the uncorrelated t-model. For details see [15] and [21]. However, maximum likelihood estimators in this case are not appealing because most important properties of maximum likelihood follow from the independence of the observations. This is not the case for t-model 2.16 for finite value of the degrees of freedom parameter v. The sample mean is obviously an unbiased and consistent estimator of μ. The unbiased estimator of Σ is given by

$$\Sigma = A\frac{v-2}{v}/(n-1).$$

2.2.10 Sampling the Multivariate Student t Distribution

Simulation of the distribution is an important issue. The sampling procedure for multivariate t distribution is based on its stochastic representation 3.

There are also other methods for sampling t distribution, see [21],[4].
The general algorithm for sampling canonical p-variate t distribution with v
degrees of freedom and correlation matrix R is the following.

Algorithm 1

- Find Choleski decomposition A of the correlation matrix R,

- Simulate p independent random variates $z_1, \ldots, z_p$ from $N(0,1)$,

- Simulate a random variate s from χ^2_v independent of $z_1, \ldots, z_p$,

- Set $\mathbf{y} = A\mathbf{z}$. In this way we obtain a p-variate normal random variable with correlation matrix R.

- Set $\mathbf{x} = \frac{\sqrt{v}}{\sqrt{s}}\mathbf{y}$.
  $\mathbf{x}$ is a random sample from p-variate t distribution with correlation matrix R and v degrees of freedom. It is based on the representation 3.

2.2.11 Bivariate Student t distribution

All the results that were given for multivariate Student t distribution hold
naturally for the bivariate case. We rewrite some of them here, as they will
be used in next section to construct conditional copula. The construction of
the Student t distribution with the independence property will be presented
as well. This distribution is not canonical. It was derived by William Shaw,
see [1].

The two dimensional random vector $T = (T_1, T_2)$ is said to have bivariate
T distribution ([15]) with v degrees of freedom, mean vector $\mu = (\mu_1, \mu_2)^{\mathbf{T}}$
and positive definite correlation matrix R if its density is given by

$$f_{t,v}(t_1,t_2) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v}{2})\pi v\sqrt{(1-\rho^2)}}\left(1 + \frac{(t_1-\mu_1)^2 + (t_2-\mu_2)^2 - 2\rho(t_1-\mu_1)(t_2-\mu_2)}{v(1-\rho^2)}\right)^{-(\frac{v+2}{2})}. \tag{2.17}$$

Taking $\mu = [0,0]^T$, it becomes

$$f_{t,v}(t_1,t_2) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v}{2})\pi v\sqrt{(1-\rho^2)}}\left(1 + \frac{(t_1)^2 + (t_2)^2 - 2\rho(t_1)(t_2)}{v(1-\rho^2)}\right)^{-(\frac{v+2}{2})}. \tag{2.18}$$

When the correlation is zero, this expression simplifies further to:

$$f_{t,v}(t_1, t_2) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v}{2})\pi v}\left(1 + \frac{1}{v}(t_1^2 + t_2^2)^{-\frac{v+2}{2}}\right),$$

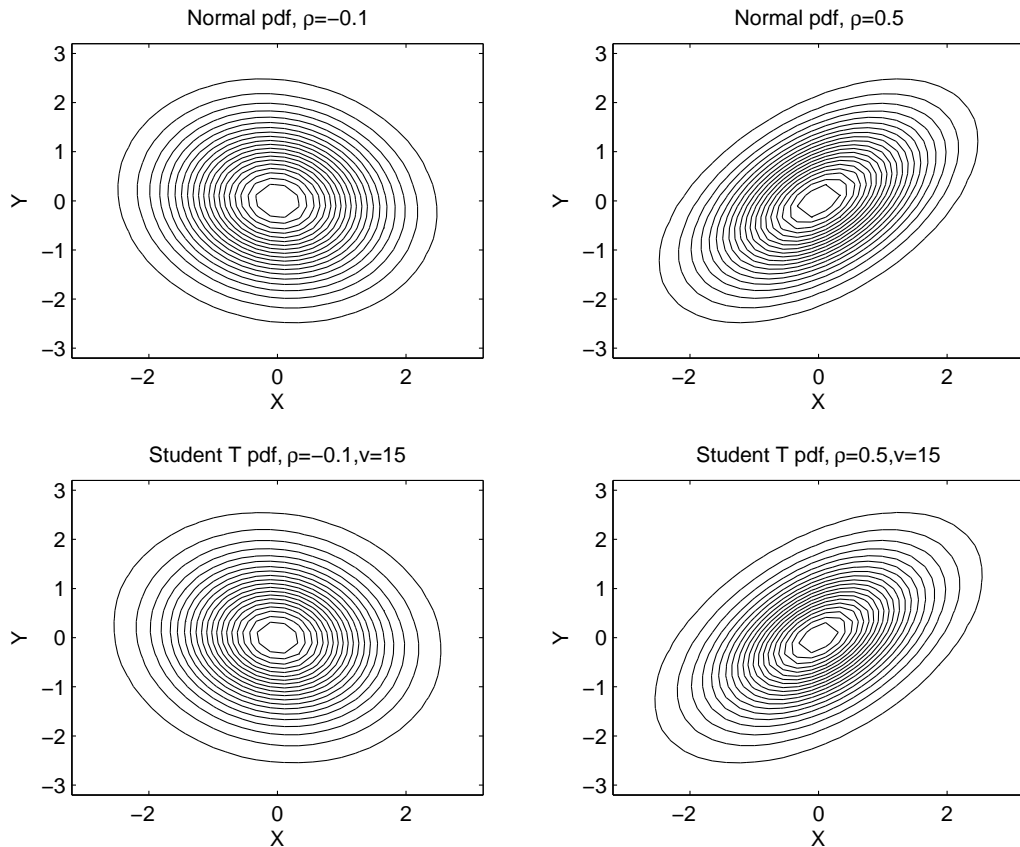which clearly is not the product of the two T student density functions.

Figure 2.10:

Figures 2.10 and 2.11 show how this density looks likes like compared to the bivariate standard normal probability density function, which is given by the equation

$$p(x_1, x_2; \rho) = [2\pi\sqrt{1-\rho^2}]^{-1}exp[-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)]. \quad (2.19)$$
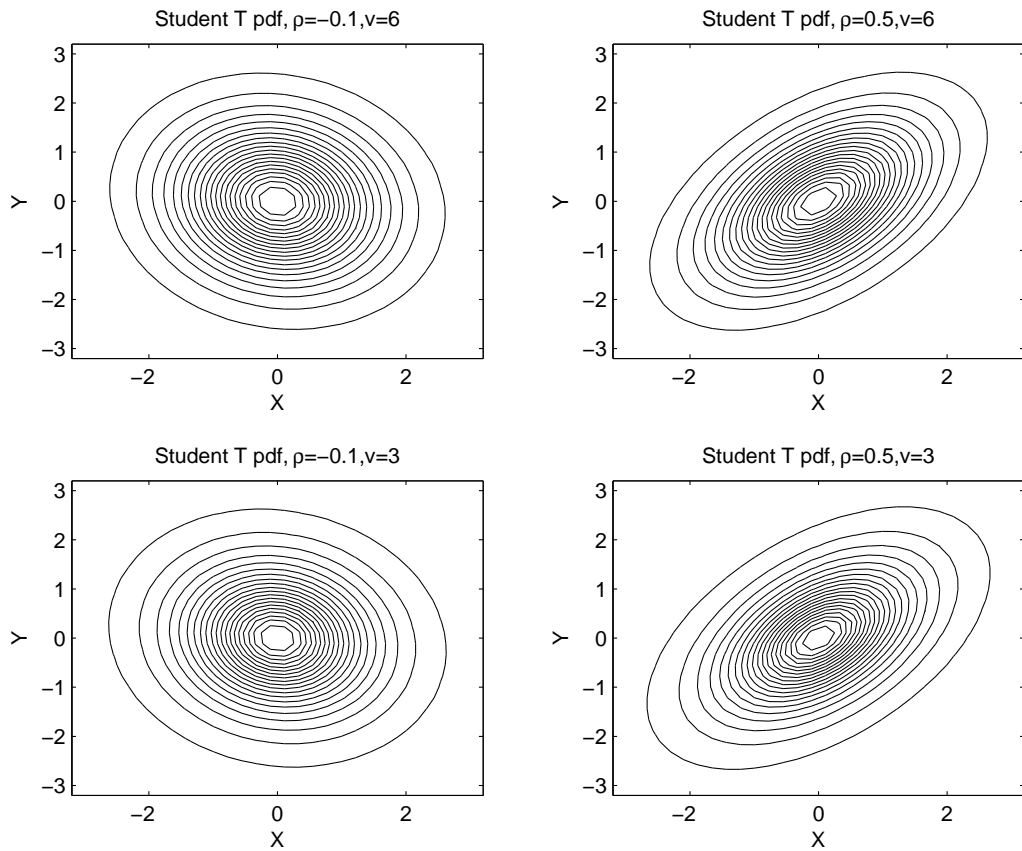
43

Figure 2.11: Contour plots of the standard bivariate Normal density with $\rho = -0.1, 0.5$ and student T density functions with for $v = 3, 6, 15$ for $\rho = -.1$ and 0.5

Plots shows normal and student t densities with $\rho = -0.1$. and 0.5. It can be observed that, as the degrees of freedom decreases, the tail of the student T densities becomes heavier.

We can derive central bivariate distribution given by (2.18) is as follows:

- Let W_i be independent standard Gaussian random variables;
- Let

$$Z_{01} = \alpha W_1 + \beta W_2, \qquad Z_{02} = \gamma W_1 + \delta W_2,$$

where α, β, γ and δ satisfy:

$$\alpha^2 + \beta^2 = 1 = \gamma^2 + \delta^2,$$

which ensures that Z_{01} and Z_{02} have standard normal distributions. It gives:

$$Z_{01} = W_1, \qquad Z_{02} = W_1 \sin\theta + W_2 \cos\theta,$$

where θ is a mixing angle.

- Construct variables

$$T_1 = Z_{01}\sqrt{\frac{v}{C^2}}, \qquad T_2 = Z_{02}\sqrt{\frac{v}{C^2}},$$

where C^2 is a sample from the χ_v^2 with v degrees of freedom.

- The inversion of these relationships results in:

$$W_1 = \sqrt{\frac{v}{C^2}}T_1 \qquad W_2 = \frac{1}{\cos\theta}\sqrt{\frac{C^2}{v}}\left(T_2 - T_1\sin\theta\right).$$

- The conditional density of T given a fixed value of $C^2 = z$ is

$$f(t_1, t_2 | C^2 = z) = \frac{z}{2\pi v \cos(\theta)} exp\left\{-\frac{z}{2v\cos^2(\theta)}(t_1^2 + t_2^2 - 2t_1 t_2 \sin(\theta))\right\}.$$

It was obtained by changing variables and exploiting the independence of W_1 and W_2. The product of their densities is a normal density. Then integrating over the χ_v^2 density of z results in the density form (2.18).

We provide detailed derivation of the t distribution with independence property and since it is similar to that of the canonical one, all the details will become more clear.

Conditional Distribution If $(T_1, T_2)^T$ has a bivariate t distribution with mean zero, correlation ρ and v degrees of freedom parameter, then $T_2|T_1 = t$ is t-distributed with $v + 1$ degrees of freedom and

$$\mu_{2.1} = E(T_2|T_1 = t) = \rho t, \qquad Var(T_2|T_1 = t) = \left(\frac{v + t^2}{v + 1}\right)(1 - \rho^2).$$

45

It is denotes as

$$dt_{v+1} \left(\frac{t_1 - \rho t}{\sqrt{\frac{(v+t^2)(1-\rho^2)}{v+1}}} \right)$$

and the inverse of the conditional distribution is also t distributed:

$$dt_{\rho,v} \left(t_1 \sqrt{\frac{(v+t^2)(1-\rho^2)}{v+1}} + \rho t \right).$$

Sampling bivariate canonical Student t distribution To sample the bivariate t distribution we use the algorithm presented in section (2.2.10). In order to show the difference between the normal and Student t distribution, we show in Figure 2.12 two samples generated for $t_{\rho=.5,v=3}$ and normal distribution with correlation $\rho = 0.5$. The vertical and horizontal lines mark true theoretical 0.005 and 0.995 quantiles for bivariate normal and t distribution. We can observe that number of points below and above these lines for the normal distribution is much smaller than for the t distribution. This comes as a consequence of the tail dependence of the t distribution. Tail dependence expresses the expectation of the joint extreme events to occur together. This concept will be explained in the next section.

Bivariate Student t Distribution with the Independence Property.

The independence property is a desired distribution property in practical applications, such as Bayesian Belief Nets. We present here the t distribution, which possess this property. It was derived recently, by William Shaw, see([1]). The density function of this distribution is complicated and it is difficult to derive the conditional distribution. The correlation of the variables depends strongly on the degrees of freedom parameter. As we will see, for small degrees of freedom the correlation does not cover the interval $[-1, 1]$. Therefore we are unable to apply this distribution to Bayesian Belief Nets. The Student t distribution with the independence property is constructed as follows:

- Let W_i be independent standard Gaussian random variables;

- Let
$$Z_{01} = \alpha W_1 + \beta W_2, \qquad Z_{02} = \gamma W_1 + \delta W_2,$$
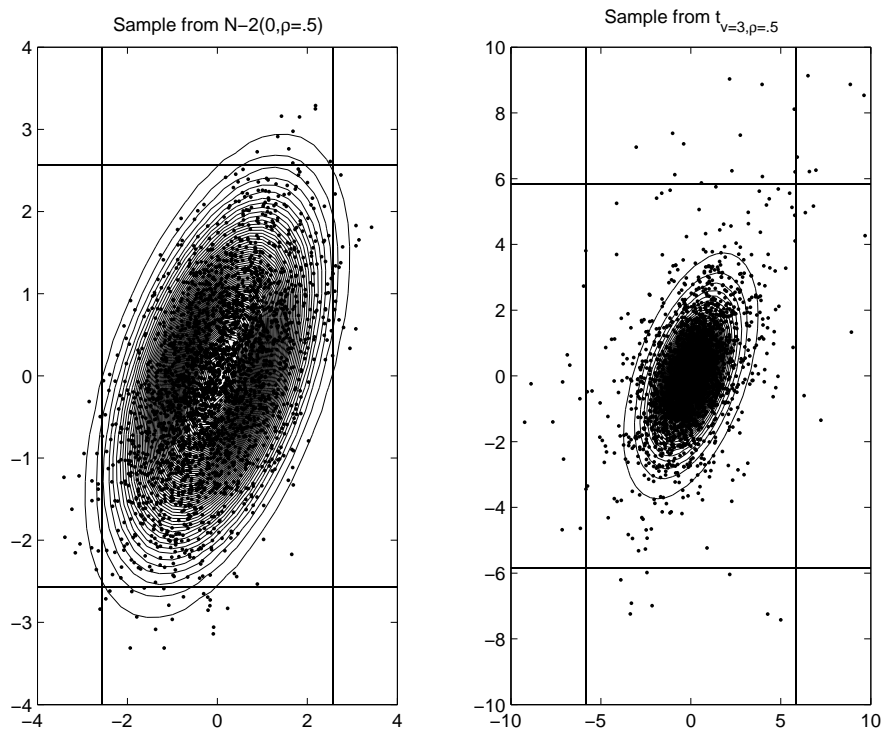
46

Figure 2.12: Contour plots of the standard bivariate Normal density with $\rho = 0.5$ and student T density functions with for $v = 3$ and $\rho = 0.5$ together with 2000 samples from each distribution. Horizontal and vertical lines mark the 0.005 and 0.995 quantiles.

where α, β, γ and δ satisfy:

$$\alpha^2 + \beta^2 = 1 = \gamma^2 + \delta^2,$$

which ensures that Z_{01} and Z_{02} have standard normal distributions. It gives:

$$Z_{01} = W_1, \qquad Z_{02} = W_1 \sin \theta + W_2 \cos \theta,$$

where θ is an mixing angle.

- Construct variables

$$T_1 = Z_{01} \sqrt{\frac{v}{C_1^2}}, \qquad T_2 = Z_{02} \sqrt{\frac{v}{C_2^2}},$$

47

where C_1, C_2 are independent samples from the χ^2_v with v degrees of freedom. This gives:

$$T_1 = \sqrt{\frac{v}{C_1^2}}W_1 \qquad T_2 = \sqrt{\frac{v}{C_2^2}}(W_1 \sin\theta + W_2 \cos\theta). \qquad (2.20)$$

- The product moment correlation of the variables Z_{01} and Z_{02} is to be found. By definition we get:

$$\rho(Z_{01}, Z_{02}) = E(Z_{01}Z_{02}).$$

Because of the assumption about the independence of the W_1 and W_2, the following relation holds:

$$Z_{01} = W_1 \perp Z_{02} - W_1 \sin(\theta) = \cos(\theta)W_2.$$

Now, we use the property of the normally distributed variable, that its linear combination remains normally distributed. Multiplying densities of the independent random variables resulting from the relation above and substituting appropriate variables we get the joint density of Z_{01} and Z_{02}:

$$\frac{1}{2\pi\sqrt{1-sin^2(\theta)}}e^{-\frac{1}{2(1-sin^2(\theta))}(z_1^2+z_2^2-2sin(\theta)z_1z_2)}.$$

It is clear now, that the product moment correlation between Z_{01} and Z_{02} is $\rho = \sin\theta$.

- The inversion of the relationship (2.20) results in:

$$W_1 = \sqrt{\frac{v}{C_1^2}}T_1 \qquad W_2 = \frac{1}{\cos\theta}\left(\sqrt{\frac{C_2^2}{v}}T_2 - \sqrt{\frac{C_1^2}{v}}T_1 \sin\theta\right).$$

By fixing values z_1, z_2 of C_1^2, C_2^2, it can be deduced that:

$$W_1 = \sqrt{\frac{z_1}{v}}T_1|_{z_1} \implies T_1|_{z_1} \sim N(0, \frac{v}{z_1}).$$

Because of the fact that $W_2 \sim N(0,1)$, we can deduce that:

$$\left(\frac{1}{cos(\theta)}T_2|_{z_2}\sqrt{\frac{z_2}{v}} - W_1 sin(\theta)\frac{1}{cos(\theta)}\right) \sim N(0,1).$$

48

Then, knowing that:

$$W_1 sin(\theta)\frac{1}{cos(\theta)} \sim N\left(0, \frac{sin^2(\theta)}{cos^2(\theta)}\right),$$

the distribution function of the variable $T_2|_{z_2}$ can be found. Solution of the following equation provides the variance of the normally distributed $T_2|_{z_2}$:

$$\frac{1}{cos^2(\theta)}x\frac{z_2}{v} - \frac{sin^2(\theta)}{cos^2(\theta)} = 1,$$

it yields that $T_2|_{z_2} \sim N(0, \frac{v}{z_2})$.

- The standard normal density for W_1, W_2 has the form

$$\frac{1}{2\pi}e^{\left(-\frac{1}{2}(w_1^2+w_2^2)\right)}.$$

- Since variables

$$\sqrt{\frac{v}{z_1}}W_1 = T_1|_{z_1} \qquad \text{and}$$

$$\sqrt{\frac{v}{z_2}}W_2 cos(\theta) = T_2|_{z_2} - \sqrt{\frac{v}{z_2}\frac{z_1}{v}}T_1|_{z_1}sin(\theta)$$

are independent and by substitution:

$$w_1 = t_1,$$

$$w_2 = t_2 - \sqrt{\frac{v}{z_2}\frac{z_1}{v}}t_1 sin(\theta),$$

the conditional density of the T_1, T_2 given fixed values z_1, z_2 of C_1^2, C_2^2 is given by:

$$f_v(t_1, t_2|C_1^2 = z_1; C_2^2 = z_2) =$$

$$= \frac{\sqrt{z_1 z_2}}{2\pi v \cos\theta}\exp\left(-\frac{1}{2\pi \cos^2\theta}(z_1 t_1^2 + z_2 t_2^2 - 2t_2 t_1 \sin\theta\sqrt{z_1 z_2})\right).$$

- This conditional density is integrated with respect to the product density of z_1, z_2 with v degrees of freedom, which is given by

$$\frac{1}{2^v \Gamma^2(\frac{v}{2})}(z_1 z_2)^{\frac{v}{2}} e^{(-\frac{1}{2}(z_1+z_2))}.$$

- After integration, the density function of the T distribution with independence property has the form:

$$C * (\alpha_1 \alpha_2)^{-\frac{v}{2}-1} \left(\Gamma \left(\frac{v+1}{2} \right)^2 {}_2F_1 \left(\frac{v+1}{2}, \frac{v+1}{2}; \frac{1}{2}; \frac{\gamma^2}{4\alpha_1\alpha_2} \right) \sqrt{\alpha_1}\sqrt{\alpha_2} \right)$$

$$-\gamma \Gamma \left(\frac{v}{2} + 1 \right)^2 {}_2F_1 \left(\frac{v}{2}+1, \frac{v}{2}+1; \frac{3}{2}; \frac{\gamma^2}{4\alpha_1\alpha_2} \right), \qquad (2.21)$$

where ${}_2F_1$ is Gaussian hypergeometric function and

$$\alpha_1 = 1 + \frac{t_1^2}{v \cos^2 \theta}, \quad , \alpha_2 = 1 + \frac{t_2^2}{v \cos^2 \theta}, \quad \delta = \frac{2 t_1 t_2 \sin \theta}{v \cos^2 \theta}$$

are intermediate variables and

$$C* = \frac{1}{\cos(\theta) \pi v \Gamma(\frac{v}{2})^2}$$

is the normalizing constant.

- Later on we will see, that the correlation of this distribution is dependent on the $\sin(\theta)$ and has the form:

$$\rho(T_1 T_2) = \frac{\sin(\theta)}{\Gamma^2 \left(\frac{v}{2} \right)} \Gamma^2 \left(\frac{v-1}{2} \right) \left(\frac{v}{2} - 1 \right)$$

Hence, when $\theta = 0$ the correlation is zero and the density function simplifies to the expression:

$$\frac{\Gamma(\frac{v+1}{2})^2}{\Gamma(\frac{v}{2})^2 v \pi} \left(\frac{1}{1+\frac{t_1^2}{v}} \right)^{\frac{v+1}{2}} \left(\frac{1}{1+\frac{t_2^2}{v}} \right)^{\frac{v+1}{2}},$$

which is the product of the two marginal T densities functions. Therefore, presented T distribution has indeed the independence property.
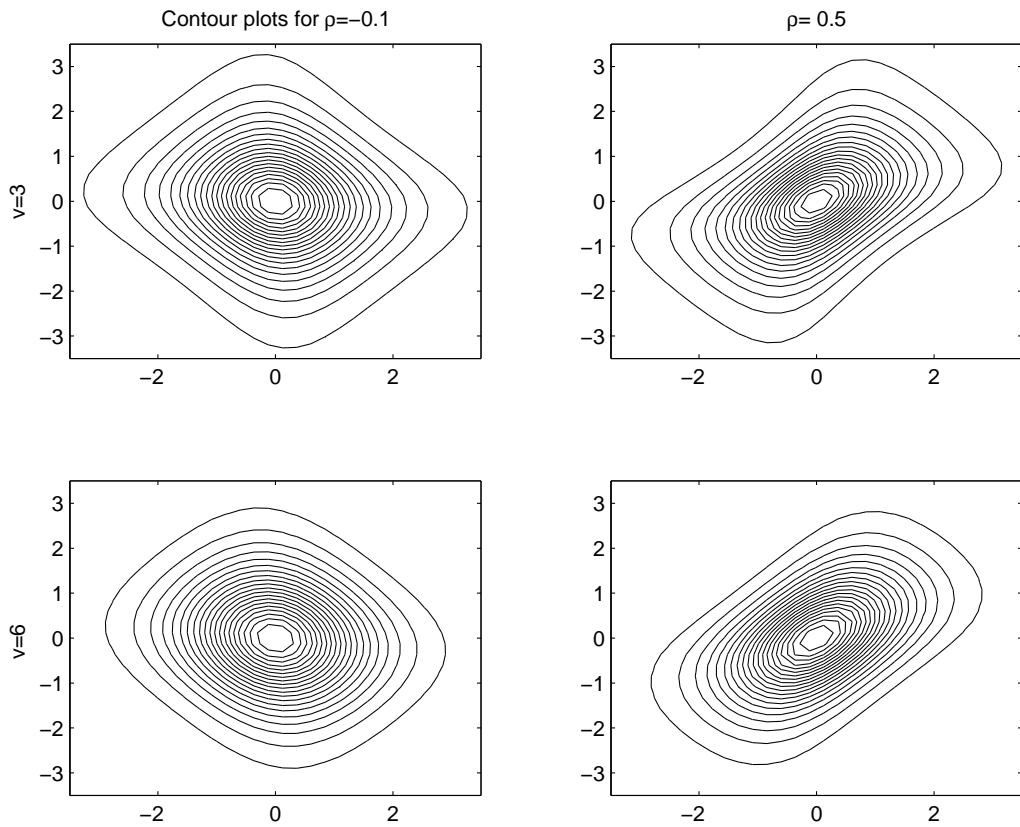
50

Figure 2.13: Contour plots of the t student density with the independence property.

Shaw and Lee are probably the first, who introduced the T density formulated in this way. Figure 2.13 shows contour plots for this density.

The shape of this density is different than the shape of the canonical T distribution. It does not look like an ellipse. The mass of the density is more concentrated, when the degrees of freedom parameter increases.

We calculate now the product moment correlation for the student T distribution with the independence property. In calculations we use the conditionally Gaussian nature of this distribution. Let the values z_1, z_2 of C_1^2, C_2^2

respectively to be fixed,then:

$$\rho(T_1, T_2 | z_1, z_2) = \frac{E(T_1 T_2 | z_1, z_2) - E(T_1 | z_1)E(T_2 | z_2)}{\sigma_{T_1 | z_1} \sigma_{T_2 | z_2}},$$

knowing that

$$T_1 \sim N(0, \frac{v}{z_1}) \quad \text{and} \quad T_2 \sim N(0, \frac{v}{z_2}),$$

we can write:

$$E(T_1 T_2 | z_1, z_2) = sin(\theta) \frac{v}{\sqrt{z_1 z_2}},$$

where $sin(\theta)$ is a correlation between T_1, T_2. Integrating over the product of the univariate gamma distributions with parameters $\alpha = \frac{v-1}{2}, \beta = 2,$ (χ_v^2 is a special case of the gamma distribution), we obtain:

$$E(T_1 T_2) = \frac{sin(\theta)}{\Gamma^2 \left(\frac{(}{v})(2)\right)} \Gamma^2 \left(\frac{v-1}{2}\right) \frac{v}{2}.$$

Knowing that the marginal distributions are also Student t, it suffices to multiply this result by standard deviations of T_1, T_2 that are equal to $\sqrt{\frac{v}{v-2}}$:

$$\rho(T_1 T_2) = \frac{sin(\theta)}{\Gamma^2 \left(\frac{v}{2}\right)} \Gamma^2 \left(\frac{v-1}{2}\right) \left(\frac{v}{2} - 1\right). \tag{2.22}$$

Figure 2.14 presets relationship between $\rho = sin(\theta)$ and $\theta \in [0, 2\pi]$ for different degrees of freedom. Function $sin(\theta)$ that corresponds to correlation of the underlying normal variables Z_{01}, Z_{02} is also plotted.

Dotted fat line is the sin function corresponding to the correlation between underlying normal variables. The solid lines, are the correlations curves that are obtained from (2.22) for θ and for different degrees of freedom. We marked the correlation curve for the distribution with 3 degrees of freedom. Clearly, it does not cover the interval $[0, 1]$, but its subset $[-0.63, 0.63]$. When degrees of freedom increase, this range increases as well.

There are several generalizations of the above construction for marginal distributions with different degrees of freedom. We present one that is based on the grouped T copula developed by Demarta ([25]). The density function is obtained in a similar way. The main difference is that the χ_v^2 variables in
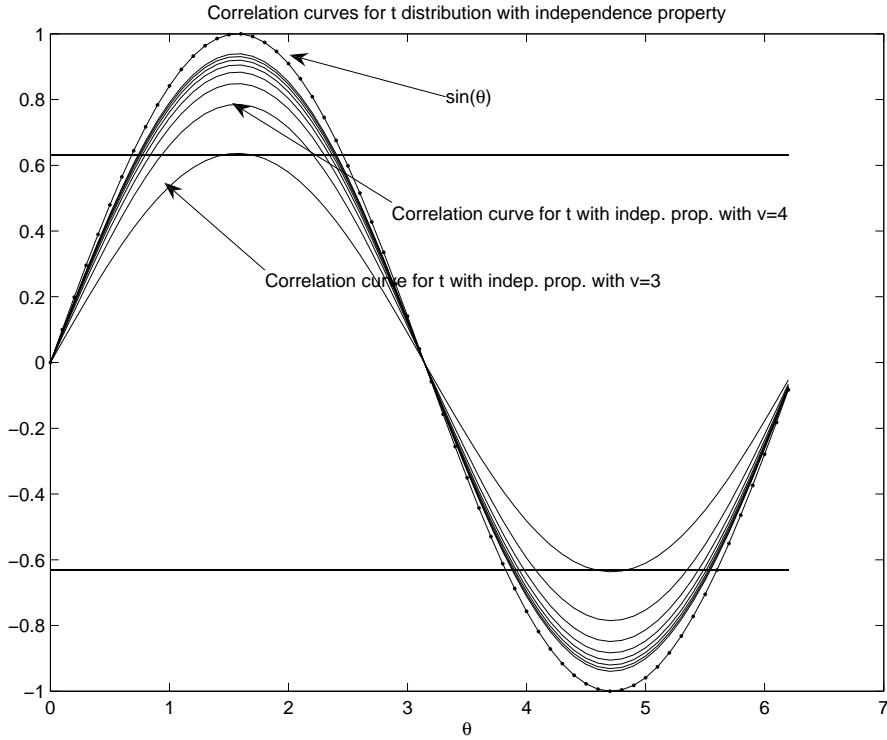
Figure 2.14: Dotted fat line is the sin function and the curves below it represents the correlations that can be reached starting with degrees of freedom equal 2 (the lowest) till 15.

the construction have different degrees of freedom, v_1, v_2. For fixed values z, w of C_1^2, C_2^2 and change of variables the conditional density has the form:

$$f(t_1, t_2 | C_1^2 = z, C_2^2 = w) =$$

$$\frac{\sqrt{zw}}{2\pi \cos\theta \sqrt{v_1, v_2}} exp\left(-\frac{1}{cos^2\theta}\left(\frac{z}{v_1}t_1^2 + \frac{w}{v_2}t_2^2 - 2sin\theta t_1 t_2 \sqrt{\frac{zw}{v_1 v_2}}\right)\right).$$

Integrating over the product of the densities of C_1^2, C_2^2 results in the density formula:

$$C*(\alpha_1)^{-\frac{v}{2}-1}(\alpha_2)^{-\frac{v_2}{2}-1}\left(\Gamma\left(\frac{v_1+1}{2}\right)\Gamma\left(\frac{v_2+1}{2}\right) {}_2F_1\left(\frac{v_1+1}{2}, \frac{v_2+1}{2}; \frac{1}{2}; \frac{\gamma^2}{4\alpha_1\alpha_2}\right)\sqrt{\alpha_1}\sqrt{\alpha_2}\right)$$

$$-\gamma\Gamma\left(\frac{v_1}{2}+1\right)\Gamma\left(\frac{v_2}{2}+1\right) {}_2F_1\left(\frac{v}{2}+1, \frac{v}{2}+1; \frac{3}{2}; \frac{\gamma^2}{4\alpha_1\alpha_2}\right), \qquad (2.23)$$

53

where $_2F_1$ is the Gaussian hypergeometric function and

$$\alpha_1 = 1 + \frac{t_1^2}{v_1 \cos^2\theta}, \quad \alpha_2 = 1 + \frac{t_2^2}{v_2 \cos^2\theta}, \quad \delta = \frac{2t_1 t_2 \sin\theta}{\sqrt{v_1 v_2} \cos^2\theta}$$

are intermediate variables, and

$$C* = \frac{1}{\cos\theta\pi\sqrt{v_1}v_2\Gamma(\frac{v_1}{2})\Gamma(\frac{v_2}{2})}$$

is the normalizing constant.

When $\theta = 0$ this density reduces to the product of the marginal densities:

$$\frac{\Gamma(\frac{v_1+1}{2})\Gamma(\frac{v_2+1}{2})\left(1 + \frac{t_1^2}{v_1}\right)^{-\frac{v_1}{2}-1}\left(1 + \frac{t_2^2}{v_2}\right)^{-\frac{v_2}{2}-1}}{\sqrt{v_1}\sqrt{v_2}\pi\Gamma(\frac{v_1}{2})\Gamma(\frac{v_2}{2})}.$$

The product moment correlation is of the form:

$$\rho = sin\theta\frac{\Gamma(\frac{v_1-1}{2})}{\Gamma(\frac{v_1}{2})}\sqrt{\frac{v_1}{2}-1}\frac{\Gamma(\frac{v_2-1}{2})}{\Gamma(\frac{v_2}{2})}\sqrt{\frac{v_2}{2}-1}.$$

Picture 2.15 presents contour plots of this density. It looks very similar to the density of the T distribution with the independence property with the same marginal. However, it is wider in the direction of the variable that has lower degrees of freedom.

Sampling these distribution is accomplished by constructing variables from (2.20) and extracting the $\sin(\theta)$ from the equation for the correlation (2.22). It is possible to extend this derivation for the higher dimensions.

As we can see from this chapter, the canonical Student t distribution has many interesting properties. We showed here only few of them, that are important for practical purposes of this work. We could also observe, that altering the construction of the t distribution resulted in a distribution with different properties than those of the canonical one. (We refer interested to the book of Kotz, [14].)
It is also worth of mentioning that Student t distribution becomes more and more popular in statistical applications, for example to model errors. It is also a basis of the construction of the T copulas. Copulas are of special
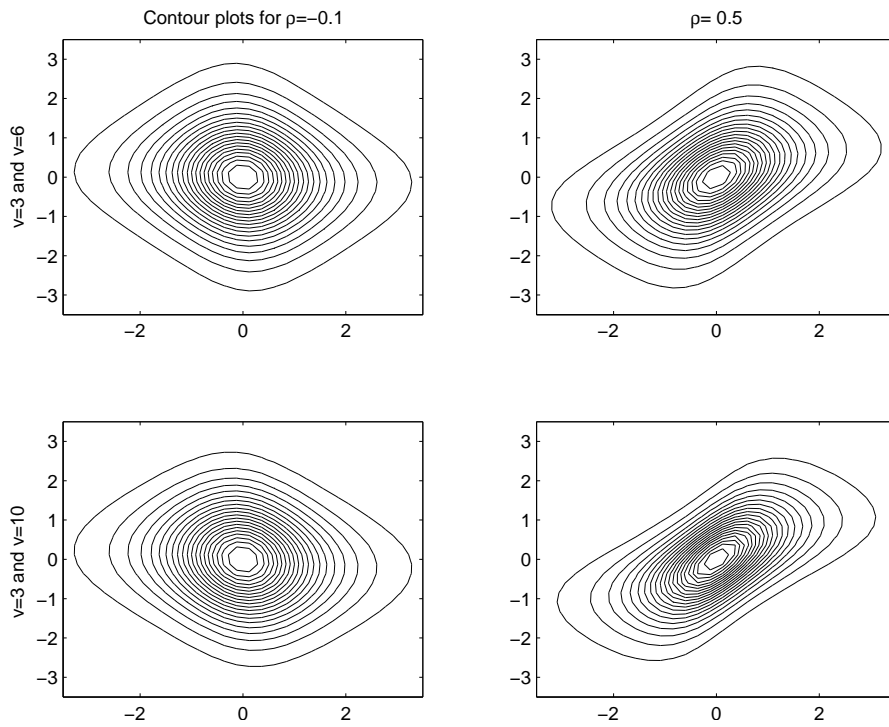
Figure 2.15: Contour plots of the t student density with the independence property.

interest as they separate the dependence structure from margins. Because of its properties, the T copula is used in financial applications as an more realistic alternative to the normal copula.

Next chapter is devoted to these functions.

Chapter 3

T Copula

In recent years copulas has become a popular tool used to model dependencies in risk modeling, insurance and finance. Analysts have to deal in practice with multi-dimensional probability distributions, where there is no clear multivariate distribution. They want to explore the dependence between variables of interest, as well as their marginal behavior. The idea of separating a distribution into a part which describes the dependence structure and a part which describes the marginal behavior only, has led to the concept of copula.

A Copula is a mathematical function that combines marginal probability into a joint distribution ([24]).

Copulas refer to the class of multivariate distribution functions supported on the unit cube with uniform marginals. The formal definition of the copula is given below. (Nelsen, [24], provides broad background for the concept of the copulas.)

Definition 3.1 (Copula) *A function $C : [0,1]^p \to [0,1]$ is a p-dimensional copula if it satisfies the following properties:*

1. *For all $u_i \in [0,1]$, $C(1,\ldots,1,u_i,1,\ldots,1) = u_i$.*

2. *For all $u \in [0,1]^d$, $C(u_1,\ldots,u_d) = 0$ if at least one of the coordinates,u_i, equals zero.*

3. *C is grounded and p-increasing, i.e., the $C-measure$ of every box whose vertices lie in $[0,1]^p$ is non-negative.*

As already mentioned, the importance of the copula comes from the fact that it captures the dependence structure of a multivariate distribution. We

present now the fundamental fact known as Sklar's theorem, which can be found in [22].

Theorem 3.1 *Given a p-dimensional distribution function H with continuous marginal cumulative distribution functions $F_1, \ldots, F_p$, then there exists a unique p-dimensional copula $C : [0,1]^p \to [0,1]$ such that*

$$H(x_1, \ldots, x_p) = C(F_1(x_1), \ldots, F_p(x_p)).$$

From Skar's theorem we can see that for continuous multivariate distribution functions, we can separate the multivariate dependence structure from univariate margins and the dependence structure can be represented by copula.

In order to explain how this unique copula is related to the distribution function, we need the following definition.

Definition 3.2 *Let F be univariate distribution function. The generalized inverse of F is defined as*

$$F^{-1}(t) = inf(x \in R : F(x) \geq t)$$

for all $t \in [0,1]$, using the convention $inf \emptyset = \infty$.

Corollary 3.1 *Let H be a p-dimensional distribution function with continuous marginals $F_1, \ldots, F_p$ and copula C (where C satisfies conditions in definition 1.) Then for any $u \in [0,1]^p$,*

$$C(u_1, \ldots, u_p) = H(F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p)).$$

Without the continuity assumption the above may not hold. Copula summarizes the dependence structure for any multivariate distribution with continuous margins.

The copula has also an important property, that strictly increasing transformations of the underlying random variables result in the transformed variables having the same copula. Therefore, the copula function of the random vector $(X_1, \ldots, X_p)$ is invariant under strictly increasing transformations. It also means that the copula corresponding to $t_p(v, \mu, \Sigma)$ is identical to that of $t_p(v, 0, R)$. This is summarized in the following theorem.

Theorem 3.2 *Consider p continuous random variables $(X_1, \ldots, X_p)$ with copula C. If $g_1, \ldots, g_p : R \rightarrow R$ are strictly increasing on the range of $X_1, \ldots, X_p$, then $(g_1(X_1), \ldots, g_p(X_p))$ also have C as their copula.*

The multivariate T copula corresponding to the canonical T distribution (and also the T student distribution with independence property with equal and unequal degrees of freedom) is constructed as follows:

$$C_{R,v}^t(u_1, \ldots, u_p) = \int_{-\infty}^{t_v^{-1}(u_1)} \cdots \int_{-\infty}^{t_v^{-1}(u_p)} f(\mathbf{t}) d\mathbf{t},$$

where $f(\mathbf{t})$ denotes the Student t density function and t_v^{-1} denotes the quantile function of a standard univariate t_v distribution.

The density of the T copula can be computed by differentiating from

$$C(u_1, \ldots, u_p) = F(F_{T_1}^{-1}(u_1), \ldots, F_{T_p}^{-1}(u_p)).$$

Applying standard integration rules

$$\left(\int_{-\infty}^{g^{-1}(x)} f(s,t) ds \right)' = f(g^{-1}(x), t)(g^{-1}(x))' = f(g^{-1}(x), t) \frac{1}{g'(g^{-1}(x))},$$

the T copula density is obtained:

$$c_{R,v}^t(u) = \frac{f_{v,R}^t(F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p))}{\Pi_{i=1}^p f_i(F_i^{-1}(u_i))}. \tag{3.1}$$

For the bivariate T copula we have

$$C_{\rho,v}^t(u_1, u_2) = \int_{-\infty}^{t_v^{-1}(u_1)} \int_{-\infty}^{t_v^{-1}(u_2)} f(t_1, t_2) dt_1 dt_2,$$

and

$$c_{v,p}^\rho(u) = \frac{f_{v,\rho}^t(F_{v_1}^{-1}(u_1), F_{v_2}^{-1}(u_2))}{f_v^t(F_{v_1}^{-1}(u_1)) f_v^t(F_{v_2}^{-1}(u_2))},$$

where $F_{v_i}^{-1}(u_1)$ denotes the quantile function of the underlying marginal T distribution with v degrees of freedom and $f_{v,\rho}^t$ denotes appropriate joint density function.
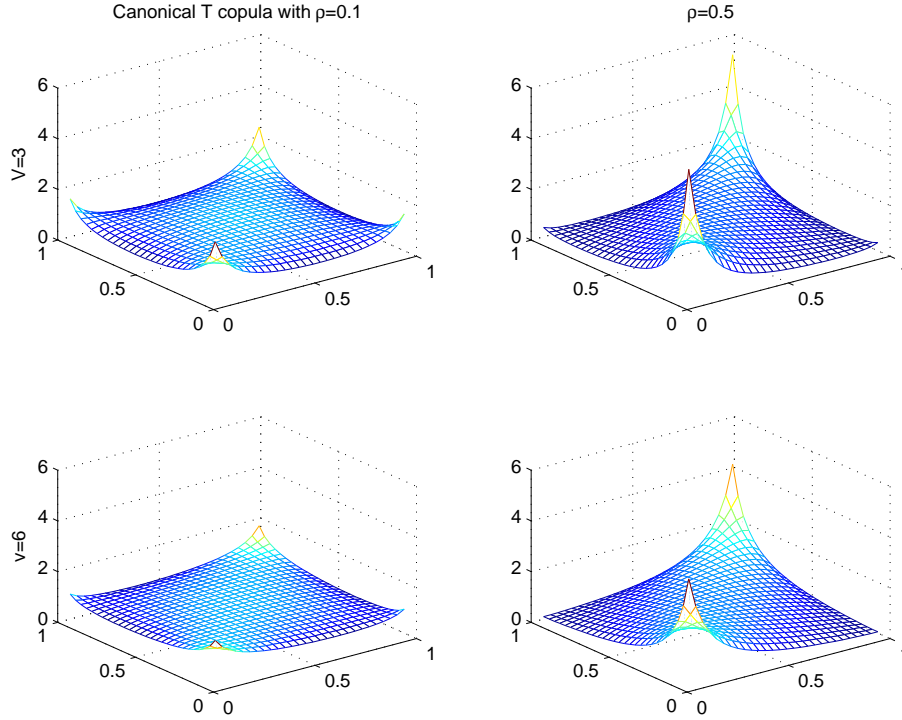
Figure 3.1: Plots of the density functions of the canonical T copula.

Figure 3.1 shows the density functions for the T copula with $\rho = -.1$ and $\rho = .5$ with $v =$ degrees of freedom. We can see that it resembles Gaussian copula. The conditional T copula $C_{v,\rho}^t(u|w)$ is given by the equation:

$$t_{v+1,\rho}\left(\frac{u - \rho \times t_v^{-1}(w)}{\sqrt{\frac{(v+(t_v^{-1}(w))^2)(1-\rho^2)}{v+1}}}\right)$$

and denoted by $h(u, w, \theta)$, where second parameter corresponds to the conditioning variable and θ is a set of parameters. The inverse of the conditional T copula $(C_{v,\rho}^t)^{-1}(u|w)$ is also t distributed:

$$t_{v,\rho}\left(u\sqrt{\frac{(v + (t_v^{-1}(w))^2)(1 - \rho^2)}{v + 1}} + \rho \times t_v^{-1}(w)\right).$$

It is denoted by $h^{-1}(u, w, \theta)$. These equations are very important in practice as we will see later.

3.1 Properties of the T Copula

Most of the properties of the T copula are inherited from the properties of the underlying Student t distribution. They are essentially the same.

- The t copula is a symmetric copula;

- It is based on elliptical distribution and therefore it belongs to the class of the elliptical copulas;

- It does not posses the independence property, but the copula based on the Student t distribution with this property does have it. The plots of this copulas are shown on the Figure 3.2 We can see that they differ from the copulas based on the canonical Student t distribution.
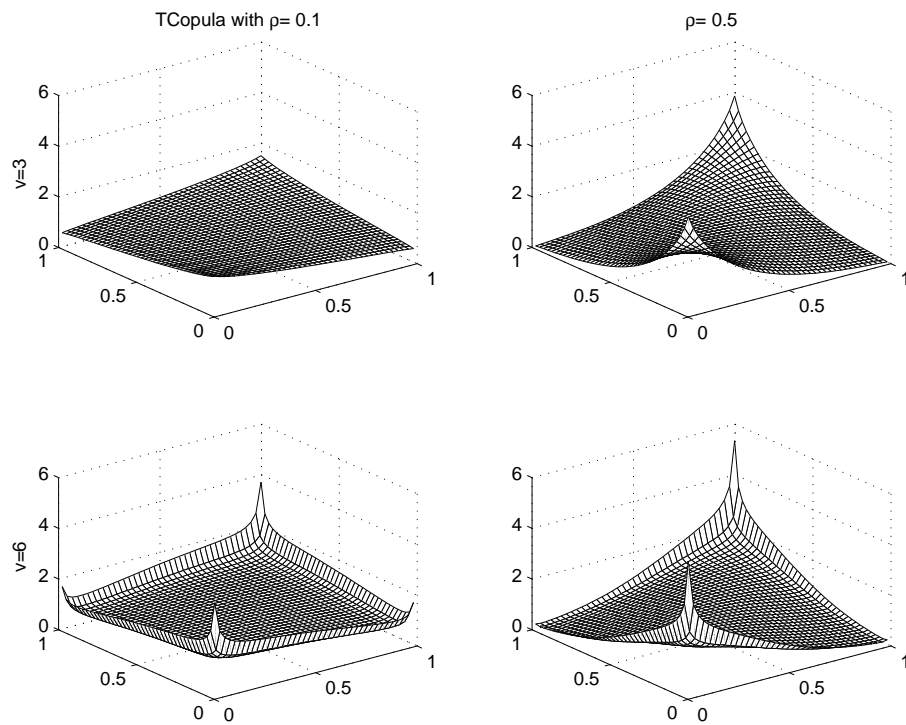

Figure 3.2: Plots of the density functions of the T copula with **equal** degrees of freedom of the marginal distributions.

60

- There is an explicit relation between product moment correlation and Kandall's τ

$$\rho = \sin\left(\frac{\pi}{2}\tau\right)$$

- Partial correlation is equal to conditional correlation. It is important for practical applications to vines. It enables to complete the correlation matrix. The notion of the vines will be explained in the next chapter.

- The T copula exhibits so called tail dependence. We will explain this concept more carefully.

3.1.1 Tail Dependence

The concept of tail dependence relates to the amount of dependence in the upper-right-quadrant tail or lower-left-quadrant tail of a bivariate distribution. This concept is relevant for the study of dependence between extreme values. It reflects the tendency of the two random variables to move together. Tail dependence gives an asymptotic indication of how often we expect to observe joint extreme values. It turns out that tail dependence between two continuous random variables X, Y is a copula property and hence the amount of tail dependence is invariant under strictly increasing transformations of X and Y, see [22].

Definition 3.3 (Tail Dependence) *Let X and Y be a random variables with continuous distributions F and G respectively. The upper tail dependence coefficient of X and Y is given by*

$$\lambda_U := \lim_{u \to 1} P(Y > G^{-1}(u) | X > F^{-1}(u)),$$

provided that the limit $\lambda_U \in [0, 1]$ exists.

The lower tail dependence coefficient λ_L is defined in analogous way.
If $\lambda_U \in (0, 1]$, then X and Y are said to be asymptotically dependent in the upper tail; if $\lambda_U = 0$ then X and Y are said to be asymptotically independent in the upper tail.
Since $P(Y > G^{-1}(u) | X > F^{-1}(u))$ can be written as

$$\frac{1 - P(X \leq F^{-1}(u)) - P(Y \leq G^{-1}(u)) + P(X \leq F^{-1}(u), Y \leq G^{-1}(u))}{1 - P(X \leq F^{-1}(u))},$$

an alternative and equivalent definition can be provided (for continuous random variables), from which it is seen that the concept of tail dependence is indeed a copula property. It is the following:

Definition 4 *If a bivariate copula C such that*

$$\lim_{u \to 1}(1 - 2u + C(u,u))/(1 - u) = \lambda_U$$

exists, then C has an upper tail dependence if $\lambda_U \in (0,1]$, and upper tail independence if $\lambda_U = 0$.

Tail dependence for the bivariate t distribution and for the corresponding T copula with v degrees of freedom, mean μ and correlation ρ are the same. Moreover, because of the symmetry, the upper and lower coefficients are equal, $\lambda_U = \lambda_L$, and given by

$$\lambda_U = \lambda_L = 2\left(1 - t_{v+1}\left(\frac{\sqrt{v+1}\sqrt{1-\rho}}{\sqrt{1+\rho}}\right)\right). \tag{3.2}$$

For details see [22]. Figure 3.3 shows plots of tail dependence for bivariate t distribution with $v =$ degrees of freedom and for bivariate standard normal distribution with different correlations. We can observe that the tail dependence for the normal distribution goes to zero as $u \to 1$.

For random variables with the bivariate Student t distribution or joined by the corresponding T copula we can expect joint extreme movements to occur with positive probability, even when the random variables have small correlation.

Figure 3.4 shows tail dependence coefficient computed from (2.20) for different degrees of freedom for zero correlation. We can see that the tail dependence coefficient is quite high for small degrees of freedom and correlation zero. This also illustrates a fundamental difference between T copula dependence structure and normal copula dependence structure. In the normal case, zero correlation implies independence, while for the t distribution it is essentially the degrees of freedom parameter v that controls the extent of the tail dependence. The tail dependence coefficient increases in ρ and decreases in v.
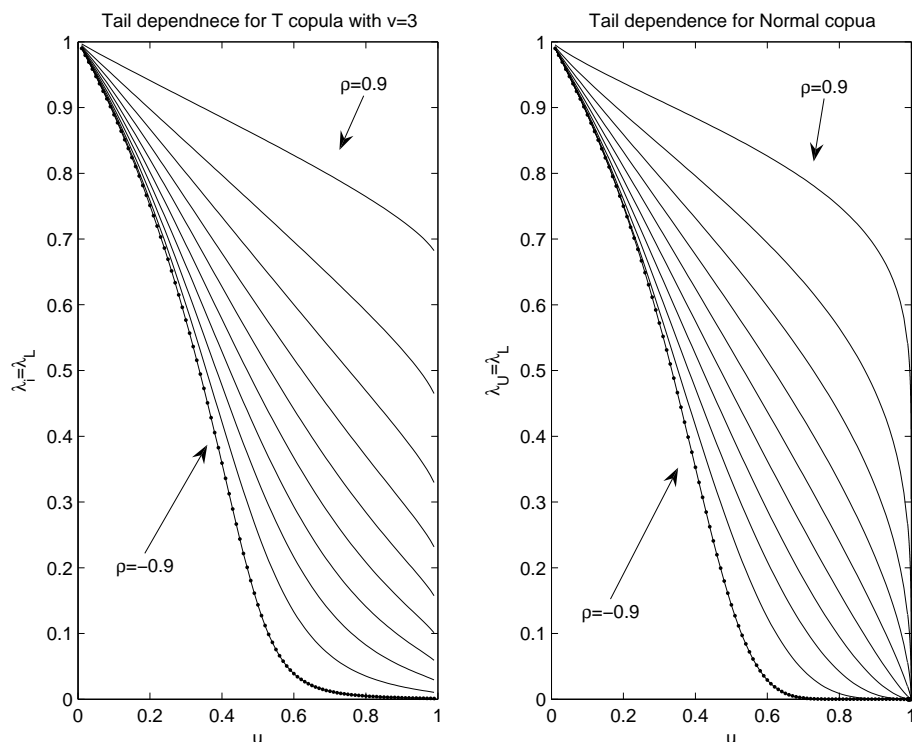
Figure 3.3: Tail dependence for bivariate T copula with v=3 degrees of freedom and for bivariate normal copula with different correlations.

3.2 Sampling the T copula

Sampling the T copula can be accomplished in a simple way. The algorithm is based on the sampling procedure of the multivariate Student t distribution. We present it below.

Algorithm 2

- Find Choleski decomposition A of the correlation matrix R,

- Simulate p independent random variates $z_1, \ldots, z_p$ from $N(0,1)$,

- Simulate a random variate s from χ_v^2 independent of $z_1, \ldots, z_p$,

- Set $\mathbf{y} = A\mathbf{z}$. In this way we obtain a p-variate normal random variable with correlation matrix R.
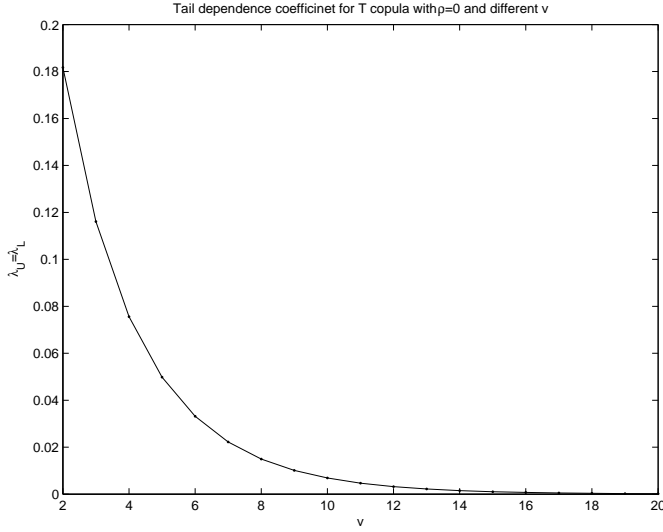
Figure 3.4: Tail dependence coefficient for T copula with $\rho = 0$ and different degrees of freedom parameters v.

- Set $\mathbf{x} = \frac{\sqrt{v}}{\sqrt{s}}\mathbf{y}$. This is the random sample from p-variate t distribution with correlation matrix R and v degrees of freedom. It is based on the representation 3 given in section 2.2.

- Set $u_i = t_v(x_i)$, for $i = 1, \ldots, p$, where t_v denotes the univariate cumulative t distribution function with v degrees of freedom.

- $(u_1, \ldots, u_p)^T \sim C^t_{R,v}$ is a sample from T copula with v degrees of freedom and correlation matrix R.

- In order to use this dependence structure to join some continuous marginal distributions $F_1, \ldots, F_p$, we set

$$(w_1, \ldots, w_p) = (F_1^{-1}(u_1), \ldots, F_p^{-1}(u + p)).$$

Figure 3.5 shows a sample from a bivariate t copula with correlation $\rho = 0.5$ and 3 degrees of freedom. For comparison a sample from normal copula with $\rho = 0.5$, which is simulated in a similar way to the T copula, is plotted. In order to observe the difference between these dependence structures, we plotted also two standard normally distributed random variables joined by both copulas. We can see that the t copula dependence structure exhibits more dependence in upper and lower corners than the normal one.
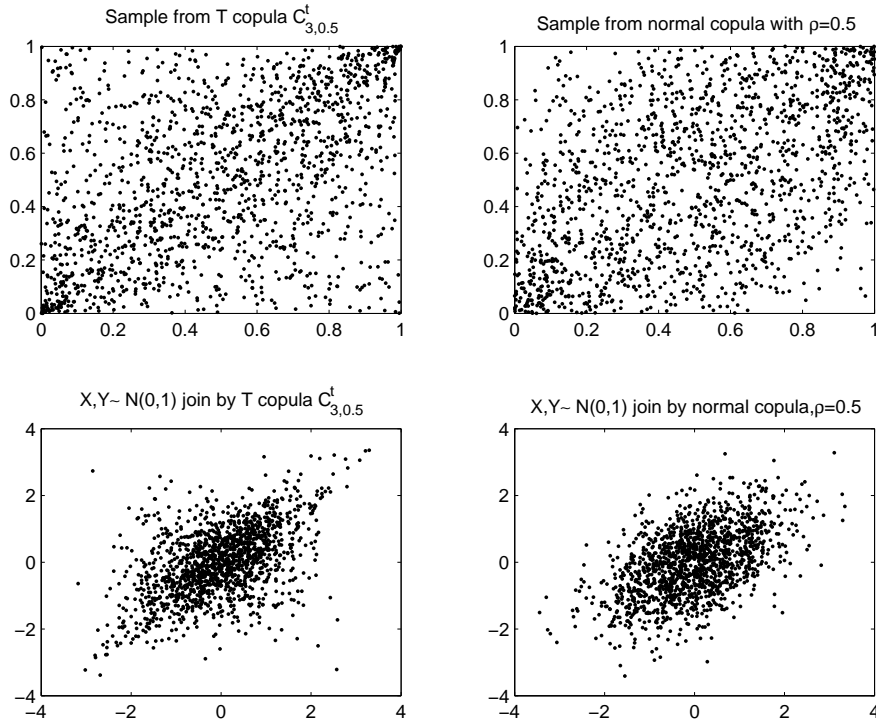
64

Figure 3.5: 2000 samples from T copula $C^t_{v=3,\rho=0.5}$ and from normal copula with linear correlation $\rho = 0.5$ and samples from the distributions obtained by joining two standard normal random variables with these copulas.

3.3 Estimating T copula

We present a pseudo-semi-parametric likelihood method to estimate a T copula, see [26]. The idea is to focus on the dependence structure of the data. Therefore, first we rule out the marginal distributions as these are irrelevant from definition of the copula. The term pseudo- likelihood is used to indicate that the raw data must first be transformed. Further, we use a semi-parametric estimation of the degrees of freedom parameter.

We present here an example of estimating the T copula for the datasets analyzed in section 2.1. We estimate the T copula's parameters from the pseudo-samples obtained by rank transformation of the datasets.

Recall that having a random sample $\mathbf{X} = \{X_i\}_{i=1}^n$ that contains observations $X_i = (X_{i1}, \ldots, X_{i,p})$. Each is a vector of data for each time period $i \in \{1, \ldots, n\}$. Since copulas are defined on the unit hypercube, we first need

to transform the data. We use the empirical marginal transformation for this purpose. For each $x \in R$ define

$$\overline{F}_j(x) := \frac{1}{n}\Sigma_{i=1}^n I_{\{X_{ij} \leq x\}},$$

for $j = 1, \ldots, p$. Next, define

$$\overline{U}_i = \left(\overline{F}_1(X_{i1}), \ldots, \overline{F}_p(X_{ip})\right),$$

for $i = 1, \ldots, n$. $\overline{U}_n = \{\overline{U}_i\}_{i=1}^n$ is the pseudo-sample. We slightly altered the definition of the empirical marginal transformation, namely $n/(n+1)\overline{F}$, to avoid edge effects that occur as some of the variables tend to one, which may result in unboundedness of the log-likelihood function. For large samples sizes the $\overline{F}$ converges to F uniformly on the real-line, almost surely. From the central limit theorem we also have that $\overline{F}$ is asymptotically normal and centered around the true distribution F.

In order to find the parameters, we need to maximize the pseudo-likelihood function. It is constructed using the density $c_{v,R}^t$ from (3.1) of the T copula

$$L_n(\theta) = II_{i=1}^n c_{R,v}^t \left((\overline{U}_i; \theta)\right),$$

where $\theta = (v, R)$ is the set of parameters.

The maximum likelihood function should be maximized simultaneously with respect to both v and Σ. This procedure is quite involved. Instead it is suggested to use the rank correlation estimator, Kendall's τ, in order to obtain the correlation matrix R. Based on this estimator, we can maximize the pseudo-likelihood function with respect to degrees of freedom parameter v.

In section 2.5 we presented the Kendall's τ and we showed also how to estimate this parameter from the data. Exploiting the relation between Kendall's τ and linear correlation for elliptical distributions, we obtained an estimator for the linear correlation. Since the rank correlation is invariant under strictly increasing transformations of the marginals we anticipate that the empirical marginal transformation ought to work well for large enough financial time-series. The estimate $\overline{\tau}$ does not use the information on the degrees of freedom parameter v, what justifies applying semi-parametric method. The general procedure is summarized in the following algorithm.

Algorithm 3

- Transform the data, $\mathbf{X_n}$, to the pseudo-sample, U_n, using the empirical marginal transformation.

- Estimate the correlation matrix $\overline{R}$ using the relation $\overline{R_{ij}} = \sin\left(\frac{\pi}{2}\overline{\tau_{ij}}\right)$ with Kendall's τ, which can be estimated from the data.

- Maximize the pseudo-log-likelihood function in order to find $\overline{v}$,

$$\overline{v} = argmax_{v \in (2,\infty]} \left\{ \Sigma_{i=1}^{n} log\left(c_{R,v}^{t}\left(\overline{U}_i; v, \overline{R}\right)\right) \right\}.$$

We can see that this procedure is quite appealing and easy to implement. It would be desired to have a tool to verify if the estimated dependence structure of a dataset is appropriately modeled by a chosen copula. For this purpose we could use the goodness of fit test. The literature provides a few tests, with no suggestion of which is the best. We are not going to treat this issue here, but Appendix A provides a description of the goodness of fit test for copulas based on the Rosenblatt transformation. This test often appears in the literature and we should be aware of its advantages as well as of its drawbacks, which are shortly pointed out.

3.4   Case Study

The methodology described above will be applied now to estimate the parameters of the T copula for the datasets from section 2.1. Those are the returns of the foreign exchange rates of the German mark, X_3, Canadian dollar, X_1, and Swiss franc, X_2, vs. the American dollar. A bivariate T copula will be estimated for each pair of data and a three-variate one for all of them together. Figure 3.6 shows the scatter plots of pairs of data and their rank transformation are presented in Figure 3.6. We can observe small correlations between pairs (X_1, X_2) and (X_1, X_3). The correlation between variables X_2 and X_3 seems to be high.

Applying the pseudo-semi-parametric log likelihood to estimate T copula for each pair of exchange rates and for all of them results in the following table.
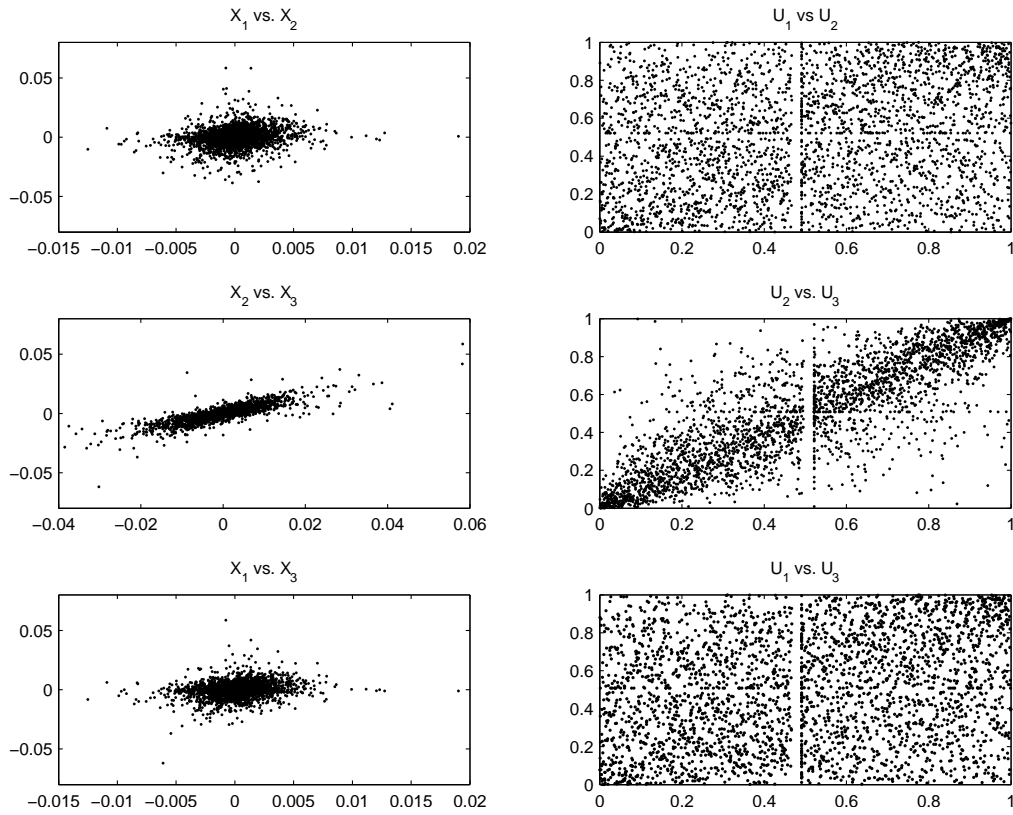
Figure 3.6: Plots of pairs of the foreign exchange returns and plots of pairs of corresponding pseudo-samples.

Table 3.1	c_{12}	c_{13}	c_{23}	c_{123}		
$\overline{\tau}$	0.1533	0.1506	0.6835	$\begin{pmatrix} 1 & 0.1533 & 0.1506 \\ 0.1533 & 1 & 0.6835 \\ 0.1506 & 0.6835 & 1 \end{pmatrix}$		
$\overline{\rho}$	0.2384	0.2344	0.8789	$\begin{pmatrix} 1 & 0.2384 & 0.2344 \\ 0.2384 & 1 & 0.8789 \\ 0.2344 & 0.8789 & 1 \end{pmatrix}$		
$\overline{v}$	14	14.6	4.4	8.2		
MaxLogLike	84.8	87.27	2116.3	2166.6		

Copula joining variables U_i, U_j is denoted here as c_{ij}.
Table 3.1 contains estimated Kendall's $\overline{\tau}$, corresponding linear correlation, $\overline{\rho}$, estimated degrees of freedom parameter $\overline{v}$ and maximal value of the pseudo-

log-likelihood.

In case of the three-variate T copula c_{123} and copula c_{23} the maximum value of the pseudo-likelihood is very high. The reason will become clear when we look at Figures 3.7 and 3.8.
They show the values of density functions of all copulas with estimated parameters for which the pseudo-log likelihood function reached its maximum for the pseudo-samples. Figures 3.7 and 3.8 present also corresponding logarithms of these values. Table 3.2 contains means of the densities values and means of the logarithms of these values. When we multiple them by the sample size, which is 2909, we get the maximum pseudo-log likelihood values, (see table 3.1.)

Table 3.2	c_{12}	c_{13}	c_{23}	c_{123}
mean of the density values	1.0657	1.0681	4.2115	4.7415
mean of the logarithm(density values)	0.0292	0.03	0.7275	0.7448

As we can see, the copula density values for some points of the uniform marginals can be very high. We marked some of them on the pictures. It turns out that the points for which these densities are large lie in the right upper and left lower corner of the uniform cubes, (hypercube for c_{123}) what can be observed on the picture 3.9 Those high values are the cause of the large pseudo-log likelihood function.

We built a three dimensional copula for foreign exchange data with $v = 8.2$ degrees of freedom. However, bivariate copulas estimated for each pair of data have different v. It means that the tail dependence of each pair will not be preserved correctly when we model the data with three dimensional T copula. In general, that would be a problem with higher dimensions. A way of overcoming this inconvenience is to decompose a multivariate density function using copulas and apply dependence model called vines as it was proposed in [27]. Next chapter presents pair-copula decomposition and provides a short description of vines.

Figure 3.7: Plots of the copula densities values for the pseudo-samples and values of the corresponding logarithms.
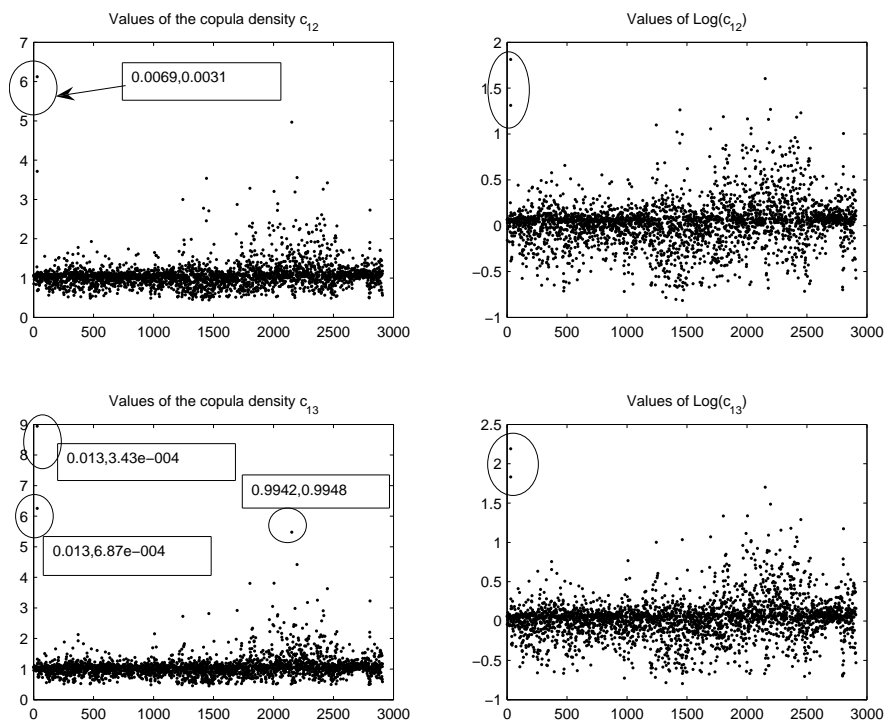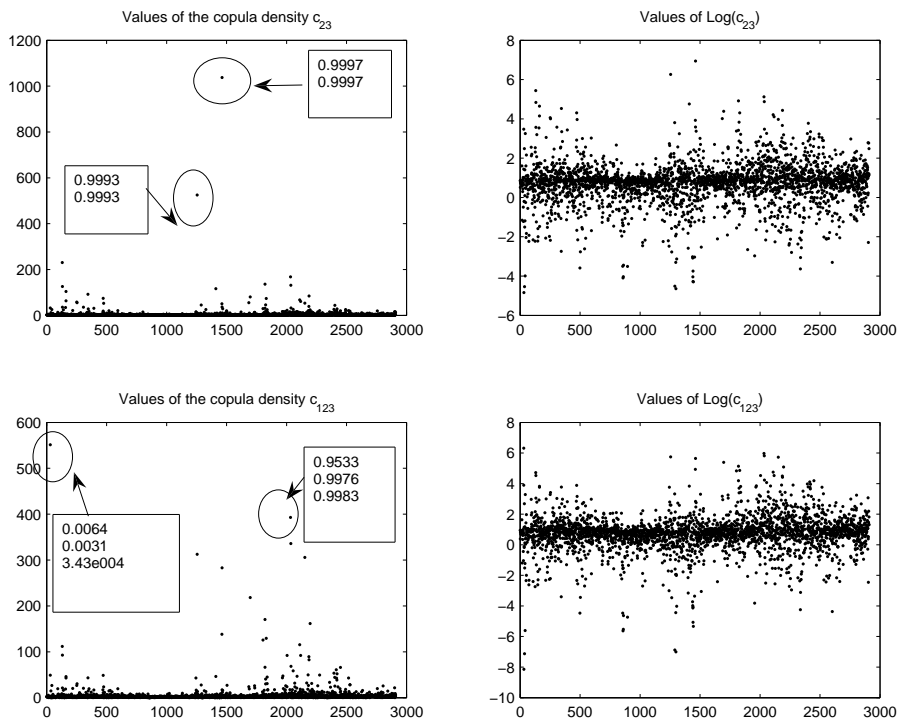
Figure 3.8: Plots of the copula densities values for the pseudo-samples and values of the corresponding logarithms.
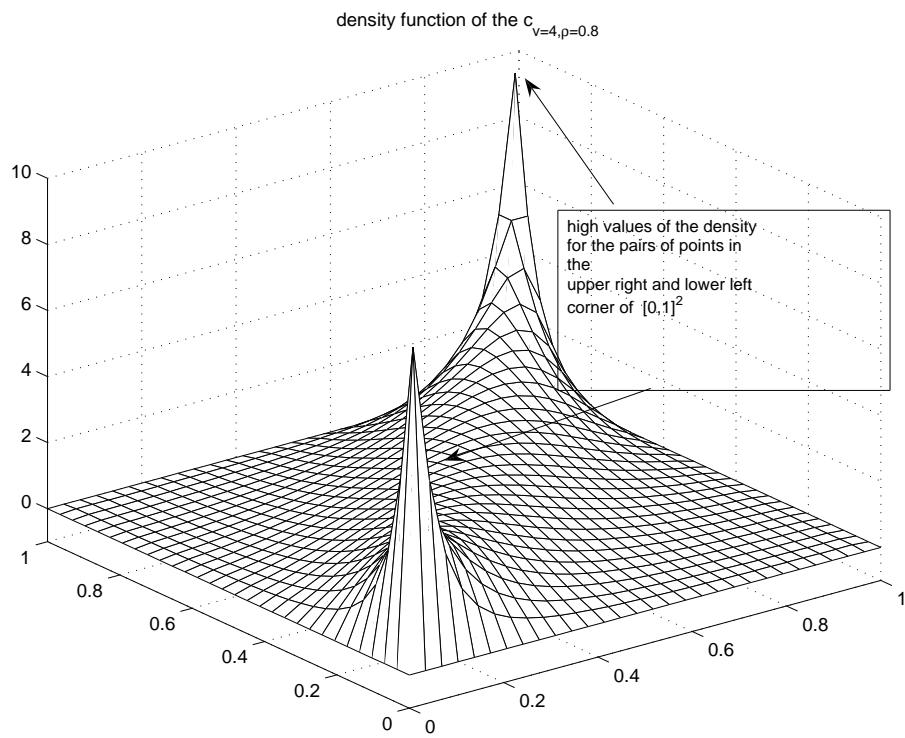
density function of the $c_{v=4, \rho=0.8}$

high values of the density for the pairs of points in the upper right and lower left corner of $[0,1]^2$

Figure 3.9: Density function of the T copula with $v = 4$ and correlation $\rho = 0.8$.

Chapter 4

Vines

Vine is a graphical dependence model. It was introduced in Cooke in 1997 and studied in Bedford and Cooke in 2002, see [28].

A multivariate density can be decomposed into a product of conditional densities. We apply vines in order to organize a pair-copula decomposition of a multivariate distribution. There are many possible decompositions and a computational effort of estimation and simulation can be diminished by applying vines. This procedure enables to incorporate more dependencies between pairs of variables than the multivariate T copula itself.

We follow work of Kjersti Aas, [27].

4.1 Pair-Copula Decomposition

Suppose we have a vector $(X_1, \ldots, X_p)$ of random variables with a joint density function $f(x_1, \ldots, x_p)$. It can be factorized in the following way:

$$f(x_1, \ldots, x_p) = f_p(x_p) f(x_{p-1}|x_p) \ldots f(x_1|x_2, \ldots, x_p). \qquad (4.1)$$

This decomposition is unique up to re-labeling of the variables. It describes a dependence between variables and their marginal behavior.

From chapter 3, we know that copula separates the dependence structure from the margins. From Corollary 3.1, we also know that we can express the p-dimensional copula in terms of the joint p-variate cumulative distribution function F and the quantile functions of the margins:

$$C(u_1, \ldots, u_p) = F(F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p)).$$

For an absolutely continuous F with strictly increasing, continuous marginal densities $f_1, \ldots, f_p$, the multivariate density function can be written in terms of the uniquely defined multivariate copula density function:

$$f(x_1, \ldots, x_p) = c_{1,\ldots,p}(F_1(x_1), \ldots, F_p(x_p))f_1(x_1), \ldots f_p(x_p).$$

It follows from the equation 3.1. For the bivariate case it simplifies to

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2))f_1(x_1)f_1(x_2),$$

where c_{12} is a pair copula density for the pair of transformed variables $F_1(X_1)$ and $F_2(X_2)$. The conditional density is of the form:

$$f(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2))f_1(x_1),$$

for the same copula c_{12}. For three random variables X_1, X_2 and X_3 we can write

$$f(x_1|x_2, x_3) = \frac{f(x_1, x_2|x_3)}{f(x_2|x_3)} = \frac{c_{12|3}(F(x_1|x_3), F(x_2|x_3))f(x_1|x_3)f(x_2|x_3)}{f(x_2|x_3)}$$

$$= c_{12|3}(F(x_1|x_3), F(x_2|x_3))c_{13}(F_1(x_1), F_3(x_3))f_1(x_1). \tag{4.2}$$

for the appropriate $c_{12|3}$ copula applied to the transformed variables $F(X_1|X_3)$ and $F(X_2|X_3)$. An alternative decomposition is

$$f(x_1|x_2, x_3) = c_{13|2}(F(x_1|x_2), F(x_3|x_2))c_{12}(F_1(x_1), F_2(x_2))f_1(x_1),$$

where $c_{13|2}$ is different from $c_{12|3}$. Proceeding in this manner leads to the pair-copula decomposition of the density in 4.1. Each term in equation 4.1 can be rewritten into an appropriate pair-copula multiplied by a conditional marginal density. The general formula is given by

$$f(x|\mathbf{v}) = c_{xv_j|\mathbf{v}_{-\mathbf{j}}}(F(x|\mathbf{v}_{-\mathbf{j}}), F(v_j|\mathbf{v}_{-\mathbf{j}})),$$

for a d-dimensional vector $\mathbf{v}$. Here v_j is an arbitrary chosen element of $\mathbf{v}$ and $\mathbf{v} - \mathbf{j}$ denotes $v-$vector, excluding component j.

For a three-variate case, the joint density can be decomposed as follows:

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)c_{13}(F_1(x_1), F_3(x_3))c_{23}(F_2(x_2), F_3(x_3)) \tag{4.3}$$

$$\times c_{13|2}(F(x_1|x_2), F(x_3|x_2)).$$

Under appropriate regularity conditions, a multivariate density can be expressed in terms of a product of pair-copulas acting on several different conditional probability distributions. The construction is iterative. Given a specific factorization, there are many different re-parameterizations.

The pair-copula decomposition involves marginal conditional distributions of the form $F(x|\mathbf{v})$. They can be written as

$$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}-j}(F(x|\mathbf{v}-j), F(v_j|\mathbf{v}-j))}{\partial F(v_j|\mathbf{v}-j)}, \qquad (4.4)$$

where $C_{i,j|\mathbf{k}}$ is a bivariate copula distribution function. When v is univariate, we have that

$$h(x, v, \theta) = F(x|v) = \frac{\partial C_{v,x}(F(x), F(v))}{\partial F(v)}, \qquad (4.5)$$

where the second parameter of $h(\cdot)$ function always corresponds to the conditioning variable and Θ denotes the set of parameters for the copula of the joint distribution function of x and v. Further, $h^{-1}(u, v, \theta)$ denotes the inverse of the $h-$function with respect to the first variable u, see [27] and references there.

4.2   Vines

The concept of vines will be introduced in this section. It will be applied to organize pair-copula decompositions of the multivariate density. There are many such decompositions and their number grows rapidly with the dimension. Regular vines, canonical and D-vine, will be used to organize them. For details see [29]. Each model gives a special way of decomposing density.

A vine on N variables is a nested set of trees, where the edges of tree j are the nodes of tree $j + 1$, for $j = 1, \ldots, N - 2$ and each tree has the maximum number of edges.

A *regular vine* on N variables is a vine in which an edge in tree $j + 1$ connects two edges in tree j, which share the same node. To each edge of a vine, the following sets of variables are prescribed:

- *Constraint set* of an edge contains variables that can be reached from this edge.

- *Conditioning set* of an edge e is formed by the intersection of the constraint sets of edges in the previous tree joined by edge e.

- *Conditioned set* of an edge e is a symmetric difference of the constraint sets of the edges in the previous tree joined by edge e.

A regular vine. *V is a vine on n elements if*

1. *$V = (T_1, \ldots, T_{n-1})$.*

2. *T_1 is a connected tree with nodes $N_1 = (1, \ldots, n)$ and edges E_1; for $i = 2, \ldots, n-1$ T_i is a connected tree with nodes $N_i = E_{i-1}$.*
   *and v is a **regular vine** on n elements if additionally*

3. **(proximity)** *For $i = 2, \ldots, n-1$ if $\{a, b\} \in E_i$, then $\#a \Delta b = 2$, where Δ denotes the symmetric difference. In other words, if a and b are nodes of T_i connected by an edge in T_i, where $a = \{a_1, a_2\}$, $b = \{b_1, b_2\}$, then exactly one of the a_i equals one of the b_i.*

Definition 4.2 *A regular vine is called a*

- **D-vine** *if each node in T_1 has degree as most 2.*

- **Canonical or C-vine** *if each tree T_i has a unique node of degree $n-i$. The node with maximal degree in T_1 is the root.*

Above definitions are taken from [29].
Figures below show examples of regular vines, canonical and D-vine. To each edge, sets of variables are assigned. Those on the left of the sign | are the *conditioned* set and on the right - *conditioning* set. Both of them are the *constraint* set.
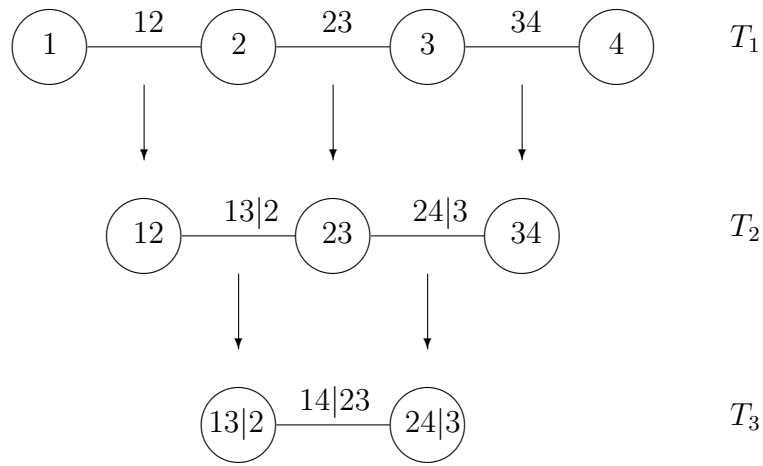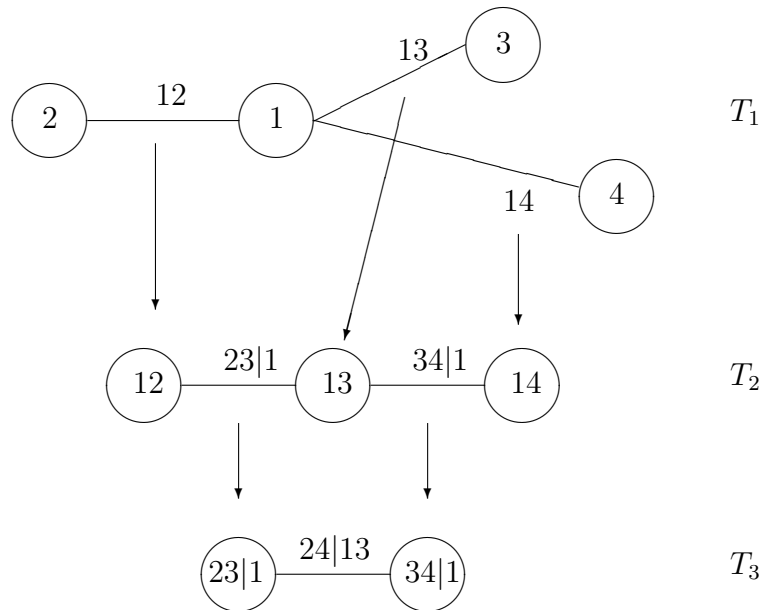
Figure 16. D-vine.

Figure 17. Canonical vine.

Regular vines can be specified by copulas. It means that to each edge we can assign a copula density with appropriate subscript. For instance, edge 14|23

can correspond to a copula density $c_{14|23}$. Formally, a copula-vine specification is defined as follows:

Definition 4.1 *Copula-vine specification (F, V, B) is a copula-vine specification if*

1. *$F = (F_1, \ldots, F_N)$ is a vector of distribution functions for random vector $X_1, \ldots, X_N$ such that $X_i \neq X_j$ for $j \neq i$.*

2. *V is a regular vine on n elements.*

3. *$B = \{C_{j,k}|e(j, k) \in \bigcup_{i=1}^{N-1} E_i\}$, where $e(j, k)$ is the unique edge with conditioned set $\{j, k\}$, and $C_{j,k}$ is a copula for $\{X_j, X_k\}$ conditional on the conditioning set for the edge $e(j, k)$.*

The whole decomposition is then defined by $n(n-1)/2$ edges and the marginal densities of each variable. Looking at the pictures and from definitions we see that nodes in tree T_j are necessary to determine labels of the edges in tree T_{j+1}. Two edges in T_j are joined by an edge in T_{j+1} only if these edges in T_j share a common node.

Density of a p-dimensional distribution written in terms of a regular vine can be found in Cooke and Bedford,[28]. It can be specialized to a D-vine and canonical vine, see [27]. Density $f(x_1), \ldots, x_p)$ corresponding to a D-vine may be written as

$$\prod_{k=1}^{p} f_k(x_k) \prod_{j=1}^{p-1} \prod_{i=1}^{p-j} c_{i,i+j|i+1,\ldots,i+j-1}(F(x_i|x_{i+1}, \ldots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1})),$$

(4.6)

where index j identifies the trees and i the edges in each tree.
The p-dimensional density corresponding to a canonical vine is given by

$$\prod_{k=1}^{p} f_k(x_k) \prod_{j=1}^{p-1} \prod_{i=1}^{p-j} c_{j,j+i|1,\ldots,j-1}(F(x_j|x_1, \ldots, x_{j-1}), F(x_{j+1}|x_1, \ldots, x_{j-1})).$$

(4.7)

D-vine and canonical vines have different structures. The choice between them depends on whether or not there is one variable that governs interactions in the dataset. If there is one, then a canonical vine should be chosen to model the data, see Figure 17.

Applying equations 4.4 and 4.5 to a three-variate case, we can get three

78

different decompositions. What is more, in this particular case a D-vine and canonical vine decompositions of the density are the same. For instance the density can take the following form:

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)c_{13}(F_1(x_1), F_3(x_3))c_{23}(F_2(x_2), F_3(x_3))$$

$$\times c_{13|2}(F(x_1|x_2), F(x_3|x_2)). \qquad .$$

It can be recognized as the decomposition given by the equation 4.3.

For four variables there are dozen of different D-vine decompositions and a dozen of different canonical decompositions. None of them are equal. There are no other possible regular vine decompositions. In five dimensional case, there is 60 different D-vines and 60 different canonical vine. None of these D-vines is the same as any of the canonical vines decompositions. What is more, there are 120 other regular vines, so in total we have 240 possible decompositions. We can see from these examples, that indeed the number of possible decompositions grows fast with dimension of the density function.

It is possible to specify vine on p elements in such a way that the resulting p-variate distribution will be Gaussian. In this case we deal with a **normal vine.** Namely, correlations and partial correlations are prescribed to each edge. Then we use the fact that partial and conditional correlations for normal distribution are equal, so we can exploit the recursive formula, given below -see [29], in order to complete correlation matrix:

$$\rho_{12;3,...,p} = \frac{\rho_{12;3,...,p-1} - \rho_{1p;3,...,p-1} \times \rho_{2p;3,...,p-1}}{\sqrt{1 - \rho_{1p;3,...,p-1}^2}\sqrt{1 - \rho_{2p;3,...,p-1}^2}}. \qquad (4.8)$$

We can also specify normal vine using rank correlations ρ_r, which can be easily estimated from the data. In order to obtain a correlation matrix we apply the implicit relation between linear correlation ρ and Spearman's ρ_r :

$$\rho = 2\sin\left(\frac{\pi}{6}\rho_r\right).$$

For example, for a three variate case showed on the Figure 18, we calculate a positive definite correlation matrix R using the above relations:
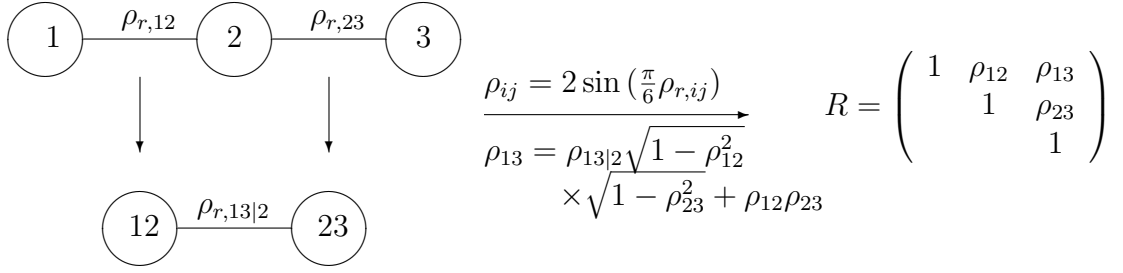
Figure 18. Specification of a normal vine.

Here $\rho_{r,ij}$ and ρ_{ij} are rank correlations and linear correlations respectively. Another way of specifying normal vine is assigning bivariate Gaussian copulas to each edge with appropriate parameters. Details about specifying vines can be found in [29] and references there.

Analogous procedure can be applied in order to obtain a **T vine**. That would be a vine, such that the marginal distributions would be Student t distributed with the same degrees of freedom parameter v. It is essential that the marginals have the same v, in order to obtain a canonical multivariate t distribution. Instead a Spearman's ρ_r we would assign Kendall's τ for the edges and exploit the implicit relation with linear correlation:

$$\rho = \sin\left(\frac{\pi}{2}\tau\right).$$

Further in order to complete the correlation matrix, we use the fact that partial correlation equals conditional correlation. For the three-variate we have the following example:
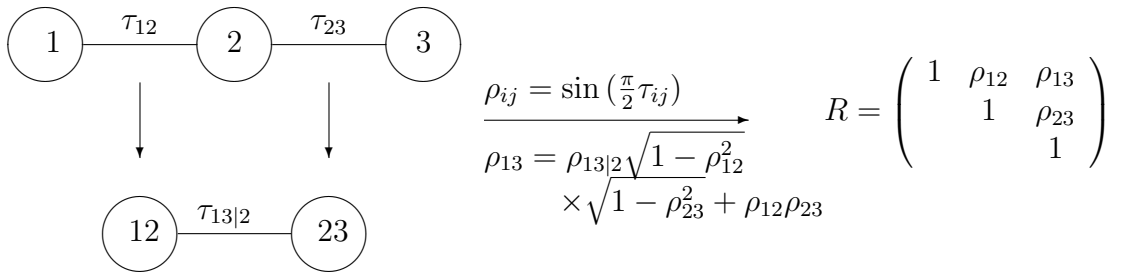

Figure 19. Specification of a T vine.

A T vine can also be obtained by assigning a bivariate T copula to each edge. It is important that the degrees of freedom parameter v is the same for each t copula in tree T_1. Otherwise, resulting distribution will not be canonical t distribution, see chapter 2.

Normal vine and T vine can be applied to realize variables $X_1, \ldots, X_p$ with arbitrary continuous distribution functions and with stipulated conditional rank correlations. This procedure is described in [29].

Conditional independence. Assuming conditional independence may reduce the number of levels of the pair-copula decomposition. For instance, if we decompose a three variate density function in (4.2) and assume that variables X_1 and X_3 are independent given variable X_2, then we have that

$$c_{13|2}(F(x_1|x_2), F(x_3|x_2)) = 1.$$

It follows from the fact that we can write copula $c_{13|2}$ in terms of conditional densities, see equation (4.2).

$$c_{13|2}(F(x_1|x_2), F(x_3|x_2)) = \frac{f(x_1|x_2, x_3)}{f(x_1|x_2)} = \frac{f(x_1|x_2)}{f(x_1|x_2)} = 1.$$

Then given that X_2, X_1 and X_3 are conditionally independent, we will not get any additional information about X_1 given X_2 and X_3. Using this assumption the pair-copula decomposition 4.3 simplifies to

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)c_{13}(F_1(x_1), F_3(x_3))c_{23}(F_2(x_2), F_3(x_3)).$$

In general, for any vector of variables $\mathbf{V}$ we say that two random variables X, Y are conditionally independent given $\mathbf{V}$ if and only if

$$c_{x,y|\mathbf{v}}(F(x|\mathbf{v}), F(y|\mathbf{v})) = 1.$$

Assuming conditional independence of the variables simplifies model. Therefore, initial factorization of the density should be made in such a way that takes it into account conditional independence assumptions.

4.2.1 Vine Simulation

General procedure for sampling D-vine and canonical vine is the same. It was discussed in [29] and [28].

Let assume for simplicity, that the margins of the distribution of the interest are uniform. Then the algorithm for sampling p dependent uniform $U(0,1)$ variables from a D-vine and canonical vine is the following:

Algorithm 4

- Sample $w_1, \ldots, w_p$ independent uniform on $[0, 1]$.

- Set

$$x_1 = w_1,$$
$$x_2 = F^{-1}(w_2|x_1),$$
$$x_3 = F^{-1}(w_3|x_1, x_2),$$
$$\ldots = \ldots$$
$$x_p = F^{-1}(w_p|x_1, x_2, \ldots, x_{p-1}).$$

In order to determine the conditional distribution $F(x_j|x_1, x_2, \ldots, x_{j-1})$ for each j, definition of the conditional copula is applied. It is used recursively. Depending on the choice between D-vine and canonical vine, we use an appropriate conditioning variable v_j. For the canonical vine we take:

$$F(x_j|x_1, x_2, \ldots, x_{j-1}) = \frac{\partial C_{j,j-1|1,\ldots,j-2}(F(x_j|x_1, x_2, \ldots, x_{j-2}), F(x_{j-1}|x_1, x_2, \ldots, x_{j-2})}{\partial F(x_{j-1}|x_1, x_2, \ldots, x_{j-2}}},$$

and for the D-vine we use:

$$F(x_j|x_1, x_2, \ldots, x_{j-1}) = \frac{\partial C_{j,1|2,\ldots,j-1}(F(x_j|x_2, \ldots, x_{j-1}), F(x_1|x_2, \ldots, x_{j-1})}{\partial F(x_1|x_2, \ldots, x_{j-1}}}.$$

The algorithms for sampling the canonical and D-vine can be found in [27]. Here we only provide sampling algorithm for a three dimensional case. It is the same for both ,the canonical and D-vine. It will be used later on.

Algorithm 5

- Sample w_1, w_2, w_3 independent uniform on $[0, 1]$,

- Set $x_1 = w_1$,

- We have that $F(x_2|x_1) = h(x_2, x_1, \theta_{11})$, and we set

$$x_2 = h^{-1}(w_2, x_1, \theta_{11}),$$

- Further we have $F(x_3|x_1, x_2) = h(h(x_3, x_1, \theta_{12}), h(x_2, x_1, \theta_{11}), \theta_{21})$ and we set
$$x_3 = h_{-1}(h^{-1}(w_3, h(x_2, x_1, \theta_{11}), \theta_{21}, x_1, \theta_{12}).$$

The set of parameters $\theta_{i,j}$ of the copulas corresponds to the copula assigned to the edge j of the tree i.

4.2.2 Estimating Vine

Estimating vines proposed in [27] is based on the pseudo-likelihood method. This method was presented in section 3.3. However, estimating vines differs from estimating copulas in that a cascade of pair-copulas is estimated instead of one multivariate copula. Asymptotic properties of the method are not known yet. A general procedure is introduced here and a three variate example is given as well.

Suppose we observe p variables at T time points. Since in practice, we do not know empirical distributions of the variables, it is proceeded as in section 3.2. It means that we transform dataset by normalizing ranks of the data. The log-likelihood for the canonical vine is then given by

$$\sum_{j=1}^{p-1}\sum_{i=1}^{p-j}\sum_{t=1}^{T}\log(c_{j,j+i|1,\ldots,j-1}(F(x_{j,t}|x_{1,t},\ldots,x_{j-1,t})), F(x_{j+i,t}|x_{1,t},\ldots,x_{j-1,t}))).$$

The log-likelihood for the D-vine is given by

$$\sum_{j=1}^{p-1}\sum_{i=1}^{p-j}\sum_{t=1}^{T}\log(c_{j,j+i|1+1,\ldots,i+j-1}(F(x_{i,t}|x_{i+1,t},\ldots,x_{i+j-1,t})), F(x_{i+j,t}|x_{i+1,t},\ldots,x_{i+j-1,t}))).$$

First sum in the equations goes over the trees, second sum goes over the nodes in each tree and the last sum goes over T observations of the variables.

For each copula in this equation there is a set of parameters to be determined. It depends on the copula applied. For the T copula the set of parameters consists of the degrees of freedom parameter and the correlation matrix R. In this particular case we apply semi-parametric log-likelihood function in order to estimate a vine decomposition. Algorithms for estimating canonical and a D-vine are provided in [27].

The inference procedure for a specific pair-copula decomposition is as follows:

1. Estimate the set of parameters of the copula in tree T_1 from the original data.

   - The data are transformed in order to obtain a pseudo-sample and then the pseudo log-likelihood method is applied to each copula.

2. Compute observations for tree T_2 using the copula parameters from tree T_1.

   - A conditional distribution functions $F(x|v)$ are computed using function $h(\cdot)$.

3. Estimate the set of parameters of the copula in tree T_2 using the observations generated in 2.

4. Compute the observations for tree T_3 using the copula parameters from tree T_2 and the h-function.

5. Estimate the parameters of the copula in tree T_3 using the observations generated at 4.

6. Repeat the procedure for the remaining trees.

The estimation of this form is easier to perform than maximizing the log-likelihood with respect to all parameters involved, since we deal only with bivariate copulas recursively.

As an example we present an estimating procedure for a three-variate model. The pseudo log-likelihood reduces to:

$$\sum_{t=T}^{T} (\log c_{12}(u_{1,t}, u_{2,t}), \theta_{11} + \log c_{23}(u_{2,t}, u_{3,t}), \theta_{12} + \log c_{13|2}(w_{1,t}, w_{2,t}), \theta_{21}),$$

where

$$w_{1,t} = F(u_{1,t}|u_{2,t}) = \frac{\partial C_{12}(u_{1,t}, u_{2,t})}{\partial u_{2,t}} = h(u_1, u_2, \theta_{11}),$$

and

$$w_{2,t} = F(u_{2,t}|u_{3,t}) = \frac{\partial C_{23}(u_{2,t}, u_{3,t})}{\partial u_{3,t}} = h(u_2, u_3, \theta_{12}).$$

The parameters to be estimated are $\boldsymbol{\Theta} = (\theta_{11}, \theta_{12}, \theta_{21})$, where as before $\theta_{i,j}$ is the set of parameters corresponding to the copula density $c_{i,j+i|i+1,\ldots,i+j_1}(\cdot;\cdot)$. Figure 20 presents the procedure graphically:

1. Estimate c_{12} and c_{23} in tree 1.

2. Generate observations for tree 2.
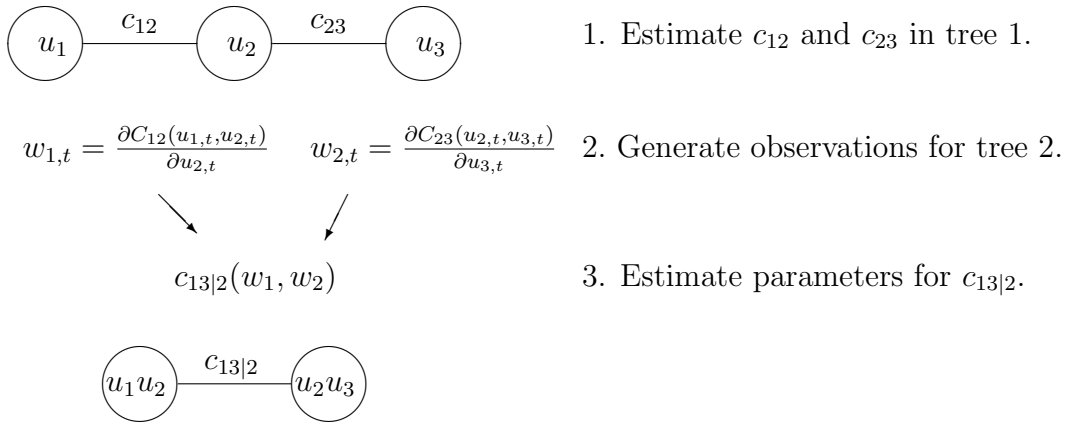
3. Estimate parameters for $c_{13|2}$.

Figure 20. Estimation procedure of the three variable vine model.

The estimating procedure described in this section is meant for a specific pairs of copulas. It should be stressed, however, that in order to select an appropriate model for a dataset, an appropriate factorization should be chosen. The type of copulas applied is also very important.

Recalling that the number of possible density decompositions grows rapidly with the dimension of the dataset, makes it hard to compare all possibilities. Only for small dimensions like 3 or 4 we can estimate parameters for all of them and pick the best one. For higher dimensions it is suggested to determine which bivariate relationships are the most important and let them determine the choice of the decomposition to be used. For instance, in a canonical vine there is one variable for which we determine the relations with the rest of variables. However, in a D-vine structure we can select more

freely which pairs to model.

The type of copulas applied is very important to model the dependences correctly. What is more, the pair-copulas do not have to belong to the same family. The resulting multivariate distribution will be valid if we choose for each pair of variables the parametric copula that best fit data. A model obtained in such a way will be even more accurate.

An important issue is to verify whether a dependence structure of data is modeled correctly. There are attempts made in order to construct a goodness of fit test for vines. Some details about that can be found in [27]. We do not study this issue here.

Comparison of the models fitted to the dataset can be accomplished by computing an AIC criterion. The concept of the AIC is explained below.

Akaike Information Criterion,AIC. Model selection for the dataset is based on the AIC criterion. It represents an information-theoretic selection based on the Kullback-Leibler information loss approach for choosing a model. The Kullback-Leibler number can be interpreted as the information lost when model g, probability distribution, is used to approximate f, full reality. It is defined for continuous functions as the integral

$$I(f,g) = \int f(x) \log \left(\frac{f(x)}{f(x,\theta)} \right) dx,$$

where θ is a set of parameters for model g. The best model looses the least information relative to other models. This is equivalent to minimizing $I(f,g)$ over g.

The criterion cannot be used directly in the models selection, because it requires knowledge of full reality and the parameter θ in the approximating models g_i. There is substantial uncertainty in the parameter estimation from the data. Models based on estimated parameters represent a major distinction from the case where model parameters are known. Therefore a change in the model selecting criterion is required. Expected estimated K-L is minimized instead of minimizing known K-L information.

K-L information can be expressed as

$$I(f,g) = \int f(x) \log(f(x)) dx - \int f(x) \log(g(x,\theta)) dx$$

or

$$I(f,g) = E_f[\log(f(x))] - E_f[\log(g(x,\theta))],$$

86

where the expectations are taken with respect to the truth. The quantity $E_f[\log(f(x))]$ is a constant, C, across the models. Hence, we can write

$$I(f, g) = C - E_f[\log(g(x, \theta))],$$

where

$$C = \int f(x) \log(f(x)) dx$$

does not depend on the data or the model. Thus, only relative expected K-L information, $E_f[\log(g(x, \theta))]$, needs to be estimated for each model in the set.

Akaike showed that the critical issue for getting a rigorous model selection criterion based on K-L information was to estimate

$$E_y E_x[\log(g(x, \overline{\theta}(y)))].$$

The inner part is just $E_f[\log(g(x|\theta))]$ with θ replaced by the maximum likelihood estimator of θ based on the assumed model g and data y. It is convenient to think about x and y as independent random samples from the same distribution. Both expectation are taken with respect to truth f.

Akaike found a formal relation between K-L information and likelihood theory. According to his findings a biased estimate of $E_x[\log(g(x, \overline{\theta}(y)))]$ is the maximized log-likelihood value and the bias is approximately equal to K, the number of estimable parameters in the approximating model, g. Therefore, an approximately unbiased estimator of $E_x[\log(g(x, \overline{\theta}(y)))]$ for large samples and "good" models, is $\log(L(\overline{\theta}|data)) - K$. This is equivalent to

$$\log(L(\overline{\theta}|data)) - K = C - \overline{E}_{\overline{\theta}}[I(f, \overline{g})],$$

where $\overline{g} = g(\cdot|\overline{\theta})$. In this manner estimation and model selection are combined under a unified optimization framework. Akaike found an estimator of expected, relative K-L information based on the maximized log-likelihood function, corrected for asymptotic bias,

$$relative \ \overline{E}(K - L) = \log(L(\overline{\theta}|data)) - K,$$

where K is the asymptotic bias correction term. This result was multiplied by -2 and this became *Akaike information criterion,* (AIC),

$$AIC = -2\log(L(\overline{\theta}|data)) + 2K.$$

The preferred model is the one with the lowest AIC value. The AIC methodology attempts to find the model that best explains the data with a minimum of free parameters. For details it is referred to [30].

The individual values AIC are not interpretable as they contain arbitrary constants and are affected by the sample size. Therefore it is suggested to rascal AIC to

$$\Delta_i = AIC_i - AIC_{min},$$

where AIC_{min} is the minimum of the different values AIC_i. The best model has $\Delta_i = 0$ and the rest of the models have positive values. The constant representing $E_f[\log(f(x))]$ is eliminated from these Δ_i values. Consequently Δ_i is interpreted as the information loss due to using model g_i instead the best model g_{min}. Therefore a meaningful interpretation can be made without the unknown scaling constants and sample size issue that enter into AIC values.

The Δ_i allow to compare and rank the hypothesis or models. The larger the Δ_i, the less plausible is fitted model i with respect to the model with minimum AIC value. It is sometimes important to know the second best model in the candidate set. The Δ_i can be a measure of its standing with respect to the best model. Simple rules allowing to interpret relative merits of models in the set are the following:

- If $\Delta_i \leq 2$, then model has substantial support,

- If $4 \leq \Delta_i \leq 7$, then model has considerably less support,

- If $\Delta_i \geq 10$, then model has essentially no support.

These are rough guidelines, but the importance of a $\Delta_i = 10$ should not be questioned, even if two values of the AIC are very big or very small. The reason is that Δ_i are free of large scaling constants. The Δ_i are interpretable as the strength of evidence.

Before we compare models fitted to the dataset using AIC and Δ_i, a simple test is performed.
A 3000 sample from a three variate vine is taken, using procedure described in section 4.3. This vine is specified by T copulas with parameters given on the Figure 18. Sample size is close to the size of the foreign exchange dataset. In this way, the results of the experiment should be plausible for the case study.
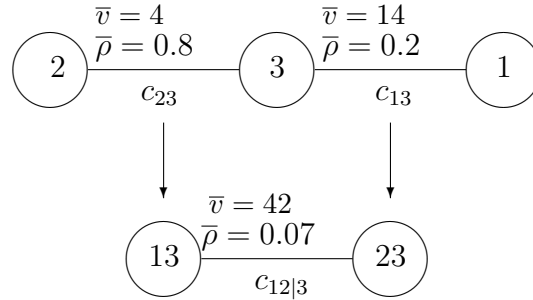
Figure 18. Sampling Vine.

We estimate a vine for this sample as well as a three variate T copula. A T copula can be interpreted as the special case of the vine model. For each estimated model AIC is computed. We repeat it 100 times and take the average over AIC criterion. For an estimated vine average AIC is -3325.7 and for the estimated three variate T copula AIC criterion equals -3315.5. Clearly, The AIC criterion indicates that vine model is more appropriate for this sample than the T copula. The value of $\Delta_{Tcopula}$ is 10.2. According to the guidelines, there is no strong evidence for the T copula model for the sample.

Similar procedure is performed for a sample from a T copula with 8 degrees of freedom and correlation matrix $R = \begin{pmatrix} 1 & 0.2 & 0.2 \\ & 1 & 0.8 \\ & & 1 \end{pmatrix}$. Again a vine and a T copula were estimated for this sample and the AIC computed 100 times. Average values of the AIC criterion for the vine is -3386.3 and for the T copula is -3394.1. Criterion indicates that T copula is more plausible for this data than the vine. The $\Delta_i = 7.8$ meaning that estimated vine has less support than the T copula. That what was expected.

We can conclude that the AIC criterion can be applied in order to find the best model for the foreign exchange dataset. It is used to compare different vines as well.

4.3 Case Study

In the previous chapter we fitted a bivariate T copulas for each pair of foreign exchange rates datasets. A three dimensional copula was fitted to the data as well. We noticed, that the three-variate copula was not capable of capturing all the dependences between pairs of the data. Applying the methodology described in this chapter results in a model that can capture the pair-wise relations between variables. It will be compared with a three dimensional copula estimated in chapter 3.

Three-variate density function can be decomposed in three distinct ways. These correspond to both a canonical and a D-vine decomposition, which are equal in this case. Estimation procedure will be performed for all of them. A AIC criterion will be applied to decide, which model is the best for the dataset. Bivariate T copulas will be fitted on each level of the estimation. First, it need to be decided how to order variables in tree 1. In a three-variate case we will compare all three possibilities. For higher dimensions it is suggested to proceed a bit differently. A bivariate T copulas should be fitted to each pair of data, obtaining estimated degrees of freedom parameter for each pair. Having these information, the variables in the first tree should be ordered such that the corresponding bivariate T copulas are in increasing order with respect to the estimated degrees of freedom parameter, v. A low number of v indicates strong dependence. The smaller v the more significant tail dependence.

Table 1 contains values of the degrees of freedom parameters estimated in Chapter 3.4.

Table 4.1	c_{12}	c_{13}	c_{23}
$\overline{\rho}$	0.2384	0.2344	0.8789
$\overline{v}$	14	14.6	4.4

Clearly the strongest dependence is between variables 2 and 3 and further between variables 1 and 2. Following suggested ordering, first decomposition is presented on the Figure 21. On the edges of the vine the estimated parameters for the corresponding T copula are written. Maximal value of the log-likelihood with respect to the set of $K = 6$ parameters of the vine is given beside the figure. Additionally the AIC criterion is given, in order to compare this vine with two other vines, which are presented on Figure 22 and 23.

$\overline{v} = 4.4$
$\overline{\rho} = 0.8789$
c_{23}

$\overline{v} = 14$
$\overline{\rho} = 0.2384$
c_{12}

$\overline{v} = 42.2$
$\overline{\rho} = 0.053$
$c_{13|2}$

log-likelihood= 2210.2

AIC= -4408.5

$\Delta_1 = 0$

Figure 21. Vine 1.

$\overline{v} = 14.6$
$\overline{\rho} = 0.2344$
c_{13}

$\overline{v} = 4.4$
$\overline{\rho} = 0.8789$
c_{23}

$\overline{v} = 68.4$
$\overline{\rho} = -0.028$
$c_{12|3}$

log-likelihood= 2205.3

AIC= -4398.6

$\Delta_2 = 9.9$
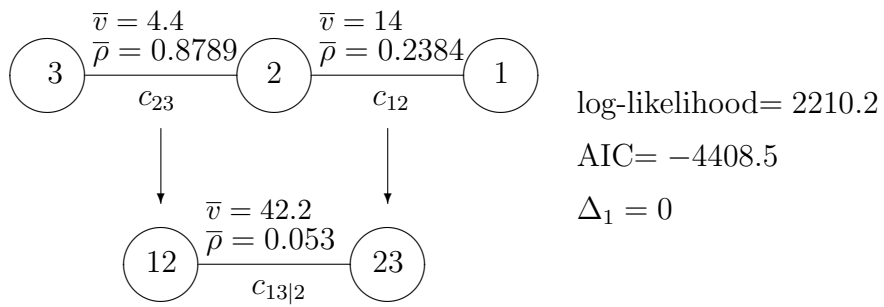
Figure 22. Vine 2.

Figure 23. Vine 3.

$\overline{v} = 14$
$\overline{\rho} = 0.2384$
c_{12}

$\overline{v} = 14.6$
$\overline{\rho} = 0.2344$
c_{13}

$\overline{v} = 4.4$
$\overline{\rho} = 0.8724$
$c_{23|1}$

log-likelihood= 2201.1

AIC= -4390.2

$\Delta_3 = 18.3$
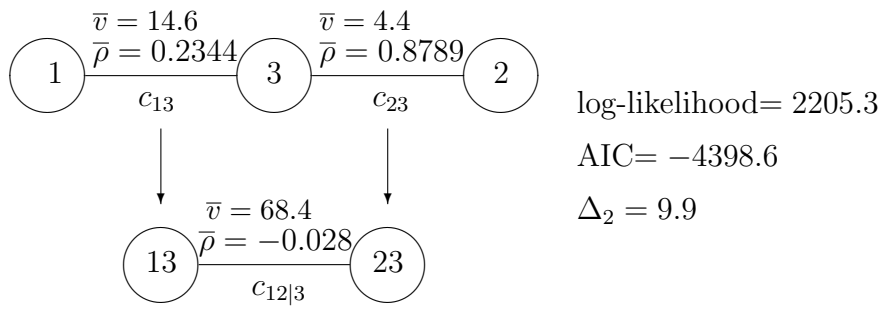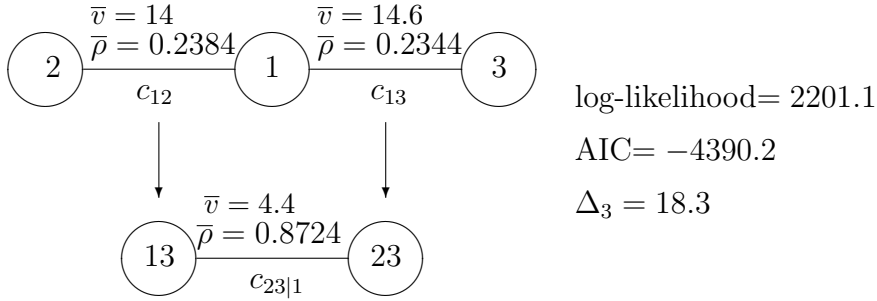
According to the AIC criterion, the best model for the dataset is vine 1. The second best model is vine 2 and the third one is vine 3, with the least support.

A three variate T copula estimated for the dataset has 8.2 degrees of freedom and the correlation matrix $R = \begin{pmatrix} 1 & 0.2384 & 0.2344 \\ & 1 & 0.8789 \\ & & 1 \end{pmatrix}$. The max log-likelihood with respect to the set of $K = 4$ parameters is 2166.6. The AIC criterion equals -4341.2 and $\Delta_{Tcopula} = 67.3$. It means that the T copula model for the dataset has essentially the least support among the set of models.

What is more, taking a sample from a vine 1 given on the Figure 21 and estimating bivariate copulas for each pair of data we have the following result:

Table 4.2	$\mathbf{c_{23}}$	$\mathbf{c_{13}}$	$\mathbf{c_{12}}$
$\overline{\rho}$	0.82	0.18	0.24
$\overline{v}$	4.8	23.6	17.2

It means that vine structure preserves bivariate relations between variables of the dataset. We can see that parameters of the copula c_{12}, that was not modeled directly are preserved quite well. That is an advantage of the vine structure over the T copula.

Consequently, the tail dependence between variables is better estimated using vines. This quantity is of importance for practical applications, when a decisions are made looking at the probability of the extreme co-movements.

Table 4.3 presents tail dependence coefficients computed for vine 1 and marginal copulas of the estimated T copula. Clearly, they are different for both structures. In this case, T copula overestimate this coefficients for c_{12} and c_{13}, and underestimate it for c_{23}.

Table 4.3	**Vine1**	**Tcopula**
c_{12}	0.0083	0.0401
c_{13}	0.0069	0.0407
c_{23}	0.577	0.45

Chapter 5

Conclusions

The Student t distribution has useful properties that make it an alternative for the normal distribution. It can describe extreme events. The additional parameter of degrees of freedom governs the heaviness of the tails. The form of the density is more complicated in contrast to the normal distribution and canonical Student t distribution does not posses the independence property. There exists, however, a construction that has this feature. It was shown in section 2.4. The T distribution is applied in statistics, risk management and to model financial data. In this work an example of foreign exchange dataset was shown in section 2.5.

T copula that is based on the Student t distribution is easy to derive and simulate. It is an elliptical copula and therefore have properties that follow from membership in this family. The T copula can be applied to model the dependence structure of the financial data. It captures the extreme co-movements of the datasets, because of the nontrivial tail dependence. Decomposing multivariate distribution using T copulas and vines results in a model that captures complicated dependence structures. This hierarchical structure, where copulas are simple building blocks preserves the pair-wise interactions between variables. Moreover, cascade inference is based on the bivariate copulas and therefore is easy to perform.

The possibility of constructing vine based model using different types of copulas should be explored further, as it results in a good approximation of the dependence structure. The algorithms for finding the best decompositions could be developed. The other criteria for comparing models should be investigated as well as the goodness of fit tests for vines.

Chapter 6

Appendix A

A Goodness of Fit Test for Copulas

This appendix is based on the paper of J.Dobric and Friedriech Schmid, see [31]. The authors investigated goodness of fit test for copulas based on Rosenblatt transformation, which is an example of the probability integral transformation: *Let X be a real-valued random variable defined on a probability space (Ω, F, P) . Let $F(x) = P(\omega : X(\omega) \leq x), x \in (-\infty, \infty)$ define the cumulative distribution function (CDF). Let $U(0,1)$ denote a random variable that is uniformly distributed on (0,1). The probability integral transform theorem is the following.*

Theorem A. *If X has CDF $F(.)$ which is continuous, then the random variable $Y = F(X)$ has the distribution of $U(0,1)$.*

$$P(Y \leq y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y.$$

The following theorem is related to the probability integral transform theorem but applies to general CDFs.

Theorem B. *Let F be a CDF. If $F^{-1} : (0,1) \rightarrow (-\infty, \infty)$ is defined by $F^{-1}(y) = \inf\{x : F(x) \geq y\}, 0 \leq 1$ and U has the distribution function $U(0,1)$, then $X = F^{-1}(U)$ has CDF F.*

Rosenblatt transformation Let X and Y denote two random variables with joint distribution function $F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$ for $(x,y) \in R^2$ and the marginal distribution functions $F_X(x) = P(X \leq x)$ and $F_Y(y) =$

$P(Y \leq y)$ for $x, y \in R$. Assume that F_X and F_Y are continuous functions. Therefore, by Sklar's theorem, there exists a unique copula $C : [0,1]^2 \to [0,1]$ such that

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)).$$

Let $z = (z_1, z_2) = T(x, y)$, where T is the Rossenblatt's transformation given by

$$T(x) = z_1 = P(X \leq x) = F_X(x),$$

$$T(y) = z_2 = P(Y \leq y | X = x) = F_{Y|X}(y|x).$$

Then the random variables $Z_1 = T(X), Z_2 = T(Y)$ are uniformly distributed on [0,1] and independent, what follows from:

$$P(Z_1 \leq z_1, Z_2 \leq z_2) = \int_{\{Z|Z-1\leq z_1\}} \int_{\{Z|Z_2\leq z_2\}} d_y F_{Y|X}(y|x) d_x F_X(x)$$

$$= \int_0^{z_1} \int_0^{z_2} dz_1 dz_2 = z_1 z_2 = P(Z_1 \leq z_1) P(Z_2 \leq z_2),$$

where $z_1, z_2 \in [0, 1]$, by their definition and vector $Z = (Z_1, Z_2)$ is uniformly distributed on $[0, 1]^2$.

Suppose now that C is a copula such that

$$F_{X,Y}(x, y) = C(F_X(x), F_G(y)).$$

Let $C(u, v)$ denotes the joint distribution function of the variables $U = F_X(X)$ and $V = F_Y(Y)$. The conditional distribution of V given U is given by

$$
\begin{aligned}
C(v|u) &= P(V \leq v | U = u) \\
&= \lim_{\Delta u \to 0} P(V \leq v | u \leq U \leq u + \Delta u) \\
&= \lim_{\Delta u \to 0} \frac{C(u + \Delta u, v) - C(u, v)}{\Delta u} \\
&= \frac{\partial}{\partial u} C(u, v).
\end{aligned}
$$

Hence, we can write variables Z_1, Z_2 using conditional distribution C,

$$Z_1 = C(F_X(X), 1) = F_X(X) = U,$$

$$Z_2 = C(F_Y(Y)|F_X(X)) = C(V|U).$$

96

If $(F(F_X(X)), F_Y(Y))$ has distribution C, then $\Phi^{-1}(Z_1)$ and $\Phi^{-1}(Z_2)$ are independent and normally distributed. Consequently,

$$S(X, Y) = [\Phi^{-1}(Z_1)]^2 + [\Phi^{-1}(Z_2)]^2$$

has chi-square distribution with two degrees of freedom. Now, if $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a random sample from (X, Y), then c is a random sample from chi-square distributed variable.

Goodness of fit test for copulas Above preliminaries are used to perform a test for the null hypothesis of interest:

$$H_0 : (X, Y) \text{ has copula } C(u, v),$$

where the marginal distributions are known. In that case the values of $S(X_1, Y_1), \ldots, S(X_n, Y_n)$ can be computed and can be used to test the auxiliary null hypothesis

$$H_0^* : S(X, Y).2cm\text{is } \chi_2^2 \text{ distributed.}$$

As H_0 implies H_0^* we reject H_0 is H_0^* is rejected.

Dobric and Schmid's results. The Dobric and Schmid suggested testing this hypothesis with the Kolmogorow Smirnov test, the Cramer von Mises test or the Anderson Darling test. They checked performance of the test using Anderson Darling, which is a modification of Kolmogorow Smirnov test. They wanted to deny the assumption of Beayman that the test "will not be significantly affected by the use of the empirical distribution function".

To do so, they focused on the first kind error probability α and on the power of the test. Dobrovic and Schmid used Monte Carlo method to check whether the true error probability of the test corresponds to prescribed level. They considered two cases:

1. The marginal distributions are known and parameters of copula is known as well.

2. The marginal distribution are not known and replaced by corresponding empirical distribution functions; the parameters for a chosen copula are estimated form the data. (This case is relevant for applications.)

Investigation of these two cases allowed them to make comparison and to check correctness of their programming in Matlab. What they obtained was almost perfect agreement of the prescribed and true error probability of the first kind in case one. However, in the second case, replacement of the true marginal distribution functions by corresponding empirical ones caused that true level of the test was .00 regardless of the prescribed level. It means that the test does not work correctly. They concluded that the reason of such situation was using empirical marginal distributions. Test's ability to reject the null hypothesis when it is actually false- its power, was investigated as well. According to simulations which were made, there was an effect of replacing marginal distributions by their empirical counterparts. The authors showed that the goodness of fit test based on the Rosenblatt transformation works well in the standard case, when parameters and marginal distribution functions are known. They presented improved test for the second case. Critical values of Anderson Darling statistic were determined by bootstrapping[1] because of their strong dependence on parameter p, which was unknown in the second case. Having this, the null hypothesis was rejected if Anderson Darling computed from the original observations $(x_i, y_i), i = 1, \ldots, n$ was larger that critical value.

---

[1] The bootstrap is a method of Monte Carlo simulation where no parametric assumptions are made about the underlying population that generated the random sample. Instead, we use sample as an estimate of the population. This estimate is called the empirical distribution $\overline{F}$, where x_i has the same likelihood of being selected in a new sample taken from $\overline{F}$. When we use $\overline{F}$ as our pseudo-population, then we re-sample with replacement from the original sample $(x_1, \ldots, x_n)$. We denote the new sample obtained in this manner by $x^* = (x_1^*, \ldots, x_n^*)$. Since we are sampling with replacement from the original sample, there is a possibility that some points x_i will appear more than once in x^* or maybe not at all. We are looking at the univariate situation, but the bootstrap concept can also be applied in the multivariate case. In many situations, the analyst is interested in estimating some parameter θ by calculating a statistic from the random sample.

Basic Bootstrap Procedure

- Given a random sample $x = (x_1, \ldots, x_n)$, calculate $\overline{\theta}$.
- Sample with replacement from the original sample to get $x^{*b} = (X_1^{*b}, \ldots, x_n^{*b})$.
- calculate the same statistic using the bootstrap sample in step 2 to get $\overline{\theta^{*b}}$.
- Repeat steps 2 and 3 B times.
- Use this estimate of the distribution of $\overline{\theta}$, (bootstrap replicates), to obtain desired characteristic.

They showed that their bootstrap version of the test keeps the prescribed values for sufficiently well and further, that the improved test has ability to detect correct null hypothesis.

The algorithm of the bootstrapped version of the test for the normal copula is given below.

Algorithm A

1. Product moment correlation $\overline{\rho}$ for a sample $(x_1, y_1), \ldots, (xn, y_n)$ is estimated by applying the equation

$$\overline{\rho} = 2\sin(\rho_r \pi/6)$$

   to the Spearman's ρ_r computed from the data.

2. 1000 i.i.d. observations $(x_1^*, y_i^*), \ldots, (x_n^*, y_n^*)$ from Normal Copula corresponding to $\overline{\rho}$ are generated.

3. Correlation $\overline{\rho^*}$ is estimated and $\overline{S^*}(x_1^*, y_1^*), \ldots, \overline{S^*}(x_n^*, y_n^*)$ is calculated in order to get the value of the test statistic, Kolmogorov-Smirnov or ANderson-Darling.

4. Steps 2 and 3 are repeated B times and new critical value of the test is determined as the $(1 - \alpha)$ quantile of the test statistics.

Depending on the test, we reject or accept null hypothesis with respect to the new critical value.

There was also third case mentioned in [31]. Namely, when the distributions of X and Y are modeled in a parametric way, $F_X(x) = F_X(x, \alpha)$ and $F_Y(y) = F_Y(y, \beta)$, where α and β are vectors of parameters which can be estimated from $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. Estimation of the copula parameter set θ and goodness of fit testing for the copula is based on $F_X(x, \alpha)$ and $F_Y(y, \beta)$. The encountered problems are the following:

- Parametric model for the distributions of X and Y may introduce misspecification. Some preliminary goodness of fit testing for the margins is necessary and it will have additional impact on the properties of fit test for the copula.

- The effect of pretesting is very difficult to investigate, because rejection of the parametric model is a possible consequence of pretesting and will often occur in practice, even if the parametric model is correctly specified.

Dobric and Schmidt did not find the way of improving the test to solve these problems. They believe that bootstrap method will work.

Bibliography

[1] W.T.Shaw,K.T.A.Lee, Apr 2007. *Copula Methods vs Canonical Multivariate Distributions: the multivariate Student T Distribution with general degrees of freedom.* Working paper;

[2] M.Abramovitz, I.A.Stegun, 1965, *Hanbook of Matematical Functions with Formulas, Graphs and Matematical Tables* New York: Dover;

[3] W.T.Shaw, 2006,*Sampling Sudent's T distribution - use of the inverse cumulative distribution function* Journal of Computational Finance Vol 9, Issue 4.pp 37-73;

[4] Devroye,L., 1986,*Non-uniform Random Variate Generation*, Springer. Available online from http://cg.scs.carleton.ca/ luc/rnbookindex.html;

[5] J.E.Gentle,2002, *Elements of Computational Statistics*, New York: Springer;

[6] S.I. Resnick,207, *Heavy-Tail Phenomena. Probabilistic and Statistical Modeling*, Springer Series in Operations Research and Financial Engineering;

[7] O.V.Pictet,M.M.Dacorogna,U.A.Muller, *Hill, Bootstrap and Jackknife estimators for Heavy Tails* in *A Practical Guide to Heavy Tails. Statistical Techniques and Applications.*,by R.J.Adler,R.E.Feldman,M.S.Taqqu, Boston: Brickhauser 1998;

[8] P.Embrechts,C.Kuppelberg,M.Thomas,1997, *Modeling extremal events for insurance and finance.* Springer, Berlin;

[9] B.M.Hill, 1975, *A Simple General Approach to Inference about the Tail of a Distribution.* The Annals of Statistics, Vol 3, No.5, pp 1163-1174;

[10] P.Hall, 1990, *Using the boostrap to estimate mean square error ans select smoothing parameter in nonparametric problem.* Journal of Multivarirate Analysis, Vol 32, pp 177-203;

[11] P.Boothe,D. Glassman, 1987, *The statistical distribution of echange rates,empirical evidence and economic implications*, Journal of International Economics,Vol 22, pp 297-319;

[12] U.A.Muller,M.M.Dacorogna,O.V.Pictet, *Heavy Tails in High-Frequency Financial Data* in *A practical guide to heavy tails. Statistical Techniques and Applications.*, by R.J.Adler,R.E.Feldman,M.S.Taqqu, Boston: Brickhauser, 1998;

[13] R.C.Blattberg,N.J.Gonedes, 1974,*A comparison of stable and student distributions as statistical models for stock prices*, Journal of Business,Vol 47,pp 244-280;

[14] S.Kotz,S.Nadarajah, 2004, *Multivarite t Distributions and Their Applications*, Cambridge University Press;

[15] B.M.Golam Kibria, A.H. Joarder, Oct 2006, *On the Multivariate T-Distribution and some of its Applications*, King Fahd University of Pertoleum and Minerals, Department of Mathematical Science;

[16] W.J.Krzanowski, 1988, *Principles of Multivariate Analysis. A User's Perpective.*, Clarendon: Press Oxford;

[17] K.T.Fang,S.Kotz and K.W.Ng,1990, *Symmetric Multivariate and Related Distributions.* Chapman&Hall, New York;

[18] S.Cambanis,S.Hung&G.Simons,1981, *On the theory of elliptically contoured distributions.* Journal of Multivariate Analysis,Vol 11,pp 368-385;

[19] F.Lindskog,A.McNeil& U.Schock, June 26, 2001, *Kendall's Tau for Elliptical Distributions.*, Remark 2 added August 7, 2002 Discussion of credit risks added Nov. 28, 2002;

[20] K.Baba,R.Shibata&M.Sibuya,2004, *Partial Correlation and Conditional Correlation as measures of Conditional Independence.* Aust.N.Z.J. Stat. Vol 46(4), pp 657-664;

[21] Kotz,S.,2005, *Sampling distributions associated with the multivariate t distribution.* Statistica Neerlandica, Vol 59, nr 2, pp 214-234;

[22] P.Embrechts,F.Lindskog & A.McNeil, Sep 10,2001, *Modeling Dependence with Copulas and Applications to Risk Management.* In *Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, Chapter 8, pp 329-384;

[23] Lindskog,F., December 11, 2002, preprint,*Linear Correlation Estimation*, http://www.math.ethz.ch/ lindskog;

[24] Nelsen,R.B., 1999, *Introduction to Copulas*, Springer;

[25] Demarta,S.,McNeil,J.,2004, *THe t Copula and Related Copulas*, International Statistical Review, Vol 73(1), pp 111-129;

[26] Mashal,R.,Zeevi,A.,2002,*Beyond Correlaion: Extreme Co-movements Between Financial Assets.*, Technical report, Columbia University;

[27] Aas,K.,et al.,*Pair-Copula constructions of multiple dependence.* Insurance Mathematics and Economics (2007), doi:10.1016/j.insmatheco.2007.02.001;

[28] Bedford,T.,Cooke,R.,2002,*Vines-a new graphical model for dependent random variables.* The Annals of Statistics Vol 30, No 4, pp 1031-1068;

[29] Cooke,R.,Kurowicka,D., 2006, *Uncertainty Analysis with High Dimensional Dependence Modeling.* Wiley series in Probability and Statistics;

[30] Burnham,K.,P.,Anderson,D.,R.,2004,*Multi-model Inference. Understanding AIC and BIC in Model Selection.* Sociological Methods & Research, Vol 33, No 2, 261-304.

[31] 1Dobric,J.,Schmid,F., 2007, *A Goodness of Fit Test for Copulas based on Rosenblatt's Transformation* Computational Statistics&Data Analysis, Vol 51, Issue 9, pp 4633-4642. (May 2007)