NTNU
Norwegian University of
Science and Technology

Multivariate Distributions Through Pair-Copula Construction: Theory and Applications

Markus Nævestad

Master of Science in Physics and Mathematics
Submission date: June 2009
Supervisor: Håvard Rue, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

The purpose of this master thesis is to study the construction of multivariate distributions using pair-copula construction and to investigate its use on financial data.

Assignment given: 19. January 2009
Supervisor: Håvard Rue, MATH

Preface

This thesis is the result of the course *TMA4905 Statistics, Master Thesis*. It represents 20 weeks of work and completes my five year Master of Technology program at the Department of Mathematical Sciences of the Norwegian University of Science and Technology (NTNU) in Trondheim.

The content of this thesis can be read by anyone with basic knowledge of mathematical and statistical concepts. It serves as a good introduction to the concept of copulas and decomposition of multivariate distributions. With this thesis as a basis, a wide range of opportunities open up for the reader, and further work is encouraged.

All implementations are done in the open source statistical software `R`.

I would like to take the opportunity to thank my supervisor Professor Håvard Rue for all his guidance and assistance. His ideas and opinions have been very helpful throughout the project.

Trondheim, June 4, 2009
Markus Nævestad

Abstract

It is often very difficult, particularly in higher dimensions, to find a good multivariate model that describes both marginal behavior and dependence structure of data efficiently. The copula approach to multivariate models has been found to fit this purpose particularly well, and since it is a relatively new concept in statistical modeling, it is under frequent development.

In this thesis we focus on the decomposition of a multivariate model into pairwise copulas rather than the usual multivariate copula approach. We account for the theory behind the decomposition of a multivariate model into pairwise copulas, and apply the theory on both daily and intra day financial returns. The results are compared with the usual multivariate copula approach, and problems applying the theory are accounted for.

The multivariate copula is rejected in favor of the pairwise decomposed model on daily returns with a level of significance less than 1%, while our decomposed models on intra day data does not lead to a rejection of the models with multivariate copulas.

On daily returns a pairwise decomposition with Student copulas is preferable to multivariate copulas, while the decomposed models on intra day data need more development before outperforming multivariate copulas.

Contents

Chapter 1

Introduction

In light of the financial crisis which is upon us, the search in financial analysis is for new and improved models. It is often very difficult, particularly in higher dimensions and in situations where we are dealing with skewed distributions, to find a good multivariate model that describes both marginal behavior and dependence structure of data efficiently. For instance when modeling financial returns, there are few multivariate models that can fit margins with different behaviors in the tails. The copula approach to multivariate models has been found to fit this purpose particularly well.

The Student copula has been found to be generally superior compared with other n-dimensional copulas for financial data such as equities, currencies, commodities and more [1]. However, the Student copula has only one parameter, the number of degrees of freedom, for modeling tail dependence. If we are to consider a multivariate model with pairs of equities with different behavior in the tails, one parameter alone might not capture the structure when increasing the number of equities.

Aas et al. (2007) [2] have developed a framework for constructing a pairwise decomposition of a multivariate model into bivariate copulas alone. In this manner one can keep track of individual parameters regarding the tail dependence. When comparing the goodness-of-fit for different copulas against each other, Berg (2009) [3] has made an overview and comparison of several approaches allowing us to find the most adequate approach for specific cases.

Our thesis is mainly built on the work of Aas et al. and Berg, and to some degree the work on vines introduced by Bedford and Cook (2002) [4]. The numerical section focus on fitting copulas to returns on both indices and specific stocks. We begin this thesis with a short introduction to bivariate copulas in Chapter 2, before we briefly discuss multivariate copulas and the fitting of copulas to data in Chapter 3. We do not account for general theory for multivariate copulas since our focus is mainly on decomposing multivariate models into bivariate copulas.

Chapter 4 and 5 are regarding notation used in the thesis as well as an overview of vines - one of the many aids in building models in the decomposition framework. The edifying of our model, and most of the algorithms used in this thesis, are accounted for in Chapter 6. In Chapter 7, problems arising when implementing the theory and comparing different models are discussed.

Finally, we have performed two thorough numerical experiments, discussed the results of them and outlined further work in Chapter 8 and 9 respectively.

Chapter 2

Bivariate Copulas

Since we in this thesis are regarding the decomposition of multivariate distributions into bivariate copulas, we focus on copulas in two dimensions when giving a brief introduction of general copula theory. We begin with some basic definitions and theorems in Section 2.1 before we look at independence and invariance properties in Section 2.2. The last part of this introduction to copulas accounts for different dependency measures.

2.1 Definitions and Theorems

Definition 2.1.1. ***Copula of* $\mathbf{F}$.** If the random vector $\mathbf{X} = [X_1\ X_2]^{\mathrm{T}}$ has joint distribution function F with continuous marginal distributions F_1 and F_2, then the copula of F (or $\mathbf{X}$) is the distribution function C of $(F_1(X_1), F_2(X_2))$.

This definition is derived from the proof of Sklar's theorem [5], stated later in this section. In [6], a copula is defined as a function C from $\mathbf{I^2}$ to $\mathbf{I}$ which fulfill the following two properties:

1. For every u, v in $\mathbf{I}$,
$$C(u, 0) = 0 = C(0, v)$$
and
$$C(u, 1) = u \text{ and } C(1, v) = v.$$

2. For every u_1, u_2, v_1, v_2 in $\mathbf{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,
$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

Every copula has a lower and upper boundary as the following theorem states [6].

Theorem 2.1.2. *Let C be a copula. Then for every (u, v) in Dom C,*
$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) = M(u, v).$$

These bounds are called the **Fréchet-Hoeffding lower and upper bound** and they are themselves copulas, named the **countermonotonicity copula** and the **comonotonicity copula** respectively. The existence of the partial derivative, as stated in the following theorem, is needed in the decomposition of a multivariate distribution in Section 4.2.

Theorem 2.1.3. *Let C be a copula. For any v in $\mathbf{I}$, the partial derivative $\partial C(u,v)/\partial u$ exists for almost all u (in the sense of Lebesgue measure), and for such v and u,*

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

Similarly, for any u in $\mathbf{I}$, the partial derivate $\partial C(u,v)/\partial v$ exists for almost all v, and for such u and v,

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

Furthermore, the functions $u \mapsto \partial C(u,v)/\partial v$ and $v \mapsto \partial C(u,v)/\partial u$ are defined and non-decreasing almost everywhere on $\mathbf{I}$.

In later sections we need to calculate maximum likelihood measures when fitting copulas to data. For this, we need the **copula density**. In n dimensions it is defined as

$$c_{12\ldots n}(u_1,\ldots,u_n) = \frac{\partial C_{12\ldots n}(u_1,\ldots,u_n)}{\partial u_1 \cdots \partial u_n}. \tag{2.1.1}$$

This density does not exist for all copulas, see [5] et al. for a discussion. For implicit copulas which have no closed form, the density is also be given by [5]

$$c_{12\ldots n}(u_1,\ldots,u_n) = \frac{f_{12\ldots d}(F_1^{-1}(u_1),\ldots,F_d^{-1}(u_n))}{f_1(F^{-1}(u_1))\cdots f_d(F^{-1}(u_n))}, \tag{2.1.2}$$

which is seen by differentiating $C(\cdot)$ in (2.1.4). The following theorem, named after Abe Sklar, is among many considered as the foundation of copula theory.

Theorem 2.1.4. Sklar's theorem. *Let H be a joint distribution function with margins F and G. Then there exists a copula $C : [0,1]^2 \to [0,1]$ such that for all x, y in $\bar{\mathbf{R}} = [-\infty,\infty]$,*

$$H(x,y) = C(F(x),G(y)). \tag{2.1.3}$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on RanF$\times$RanG, where RanF denotes the range of F. Conversely, if C is a copula and F and G are univariate distribution functions, then the function H defined by (2.1.3) is a joint distribution function with margins F and G.

Corollary 2.1.5. *Let H be a joint distribution function with continuous margins F and G, let C be a copula and let $F^{(-1)}$ and $G^{(-1)}$ be quasi-inverses (see [6]) of F and G, respectively. Then for any (u,v) in Dom C,*

$$C(u,v) = H(F^{(-1)}(u), G^{(-1)}(v)). \tag{2.1.4}$$

Equation (2.1.4) provides a useful method for constructing copulas from joint distribution functions together with the inverse transform sampling method.

2.2 Independence and Invariance Property

Two random variables X and Y are said to be independent if their joint distribution function $H(x,y)$ equals the product of their margins, F and G, i.e. $H(x,y) = F(x)G(y)$. To give the corresponding property when dealing with copulas we introduce the **product copula**

$$C(u,v) = uv = \Pi(u,v), \tag{2.2.1}$$

or often just Π.

Theorem 2.2.1. *Let X and Y be continuous random variables (i.e. their distribution functions are continuous). Then X and Y are independent if and only if $C_{XY} = \Pi$.*

$\Pi(u, v)$ is therefore also called the **independence copula**.

Under certain regularity conditions copulas are invariant under transformations, as the following theorem states.

Theorem 2.2.2. *Let X and Y be continuous random variables with copula C_{XY}. If α and β are strictly increasing on RanX and RanY, respectively, then $C_{\alpha(X)\beta(Y)} = C_{XY}$. Thus C_{XY} is invariant under strictly increasing transformations of X and Y.*

2.3 Dependence

One of the most interesting and applicable properties of copulas is their dependency structure. Most practitioners normally use the linear correlation coefficient defined below [7] when they seek a measure of dependence between two variables, for example different financial instruments.

Definition 2.3.1. The linear correlation coefficient between X and Y is

$$\mathrm{Cor}(X, Y) = \rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}. \tag{2.3.1}$$

There are mainly three disadvantages with this measure, see [8] for further discussion of these:

1. X and Y must have finite variances. This is especially a poor property in extreme value analysis where tail dependence is important.

2. Uncorrelatedness implies independence only in the multivariate normal case.

3. It is *not* invariant under non-linear strictly increasing transformations.

As a consequence of this we introduce **Kendall's tau** and **Spearman's rho** [6]:

Definition 2.3.2. Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed (iid) random vectors, each with joint distribution function H. The population version of Kendall's tau, $\tau_{X,Y}$ or just τ, is the probability of concordance minus the probability of disconcordance, i.e.

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]. \tag{2.3.2}$$

Definition 2.3.3. Let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be three independent random vectors with joint distribution function H. The margins of X and Y are F and G. The population version of Spearman's rho, ρ_S, is

$$\rho_S = 3\big(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]\big). \tag{2.3.3}$$

Spearman's rho of X and Y is in fact the linear correlation of $F(X)$ and $G(Y)$, i.e. $\rho_{S_{XY}} = \rho_{F(X)G(Y)}$.

2.3.1 Tail Dependence

The tail dependence of a bivariate distribution is a measure of the dependence in the upper-right- and lower-left-quadrant of the distribution. The definition is divided into two parts, one for upper and one for lower [9].

Definition 2.3.4. Let X and Y be random variables with distribution functions F_1 and F_2. The coefficient of upper tail dependence of X and Y are

$$\lambda_U = \lim_{u \to 1^-} \mathrm{P}[Y > F_2^{-1}(u) | X > F_1^{-1}(u)], \tag{2.3.4}$$

provided a limit $\lambda_U \in [0, 1]$ exists. If $\lambda_U \in (0, 1]$, X and Y are said to be asymptotically dependent in the upper tail; if $\lambda_U = 0$, X and Y are said to be asymptotically independent in the upper tail.

By applying Bayes' rule on (2.3.4), λ_U can be written as a function of copulas, as in the following definition.

Definition 2.3.5. If a bivariate copula C exists such that

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u} \tag{2.3.5}$$

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

The expression in (2.3.5) holds for continuous random variables. The lower tail dependency is defined similarly:

Definition 2.3.6. If a bivariate copula C exists such that

$$\lambda_L = \lim_{u \to 0^+} = \frac{C(u, u)}{u} \tag{2.3.6}$$

exists, then C has lower tail dependence if $\lambda_L \in (0, 1]$, and lower tail independence if $\lambda_L = 0$.

Chapter 3

Multivariate Copulas - Fitting Copulas to Data

Extending the theory of copulas to n dimensions results in difficulties for some copulas. Even though most of the properties discussed in Chapter 2 can be extended to hold in an arbitrary number of dimensions, this is not always the case when discussing specific copulas. In Appendix A we have discussed some of the most known copulas, namely the Gaussian copula, the Student copula and the class of Archimedean copulas with its most known members. While the Gaussian and the Student copula possess the same properties in an arbitrary number of dimensions, the Archimedean copulas do not. It is mainly the generator ϕ defined in (A.1.1) that needs additional restrictions. Since we are mainly considering multivariate distributions through pair-copula construction in this thesis, these restrictions will not be accounted for here, see [6, 5] for further discussion.

In Section 3.1.1 and 3.1.2 we account for the two methods we use; method-of-moments and maximum likelihood (ML).

3.1 Fitting Copulas to Data

In this section we will account for two different methods of fitting copulas to data; the method-of-moments using rank correlation and maximum likelihood. We have data vectors $\mathbf{X_1}, \ldots, \mathbf{X_n}$ with identical distribution F, and write $\mathbf{X_t} = (X_{t,1}, \cdots, X_{t,d})^{\mathrm{T}}$ for an individual data vector. We assume the margins of F to be continuous such that Sklar's theorem holds, i.e. we have a unique representation $F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d))$.

As a first approach of fitting a copula to data, we try to limit the possible copulas by considering the nature of the data. For instance; do we have upper or lower tail dependency, or perhaps no (apparent) tail dependency at all. Such characteristics are summarized in Table 3.1[1].

3.1.1 Method-of-Moments Using Rank Correlation

The method-of-moments procedure using rank correlations is an easy way to estimate θ in parametrical copulas because it is not necessary to assume (or estimate) the marginal

[1] Due to the radial symmetry of C, it suffices to consider λ_L to calculate the coefficient of tail dependence λ of C for the Gaussian and Student copula [5].

Table 3.1: Characteristics for typical copulas [5].

Copula	Upper tail dependence	Lower tail dependence
Gumbel	$2-2^{1/\theta}$	0
Clayton	0	$\left\{\begin{array}{ll} 2^{-1/\theta} & \theta>0 \\ 0, & \theta\leq 0 \end{array}\right\}$
Frank	0	0
Gaussian	0	0
Student	$2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}}\right)$	$2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}}\right)$

distributions. The method takes advantage of the fact that many copulas have a one-to-one correspondence between θ and ρ_S and θ and τ. Only Kendall's tau will be discussed here, see [5] for examples with the use of Spearman's rho. The relationship between τ and θ for the Gumbel, Clayton and Frank copulas are shown in Table 3.2, and the standard estimator of Kendall's tau (between two variables X_i and X_j), Kendall's rank correlation, is [5]

$$\hat{\tau}=\binom{n}{2}^{-1}\sum_{1\leq t<s\leq n}\operatorname{sign}((X_{t,i}-X_{s,i})(X_{t,j}-X_{s,j})). \tag{3.1.1}$$

Pairwise estimations of Kendall's tau can be gathered in a matrix R^τ,

$$R^\tau=\binom{n}{2}^{-1}\sum_{1\leq t<s\leq n}\operatorname{sign}(\mathbf{X_t}-\mathbf{X_s})\operatorname{sign}(\mathbf{X_t}-\mathbf{X_s})^{\mathrm{T}}. \tag{3.1.2}$$

Table 3.2: Relations between θ and τ. $D_1(\theta)=\theta^{-1}\int_0^\theta \frac{t}{e^t-1}\,\mathrm{dt}$.

Copula	Kendalls tau, τ	Parameter range
Gumbel, C_θ^{Gu}	$1-1/\theta$	$\theta\geq 1$
Clayton, C_θ^{Cl}	$\theta/(\theta+2)$	$\theta\geq -1$
Frank, C_θ^{Fr}	$1-4\theta^{-1}(1-D_1(\theta))$	$\theta\in\mathbb{R}$

3.1.2 Maximum Likelihood Estimation

In order to fit data to (multivariate) copulas, we can first estimate the margins, $F_1,\ldots,F_d$. Even though our main interest is the copula itself, the estimate of the margins may provide extra information about the data. Hence splitting the modeling into two steps can yield more insight and allow for a more detailed analysis.

If we choose to estimate the margins, we form what is called a pseudo-sample, $\hat{\mathbf{U}}_1,\ldots,\hat{\mathbf{U}}_n$, from the copula [5], where

$$\hat{\mathbf{U}}_t=(\hat{U}_{t,1},\ldots,\hat{U}_{t,d})^{\mathrm{T}}=(\hat{F}_1(X_{t,1}),\ldots,\hat{F}_d(X_{t,d}))^{\mathrm{T}}. \tag{3.1.3}$$

$\hat{F}_i(\cdot)$ can be estimated either parametric, for instance by ML, or non-parametric, using the following estimator

$$\hat{F}_i(x)=\frac{1}{n+1}\sum_{t=1}^{n}I(X_{t,i}\leq x). \tag{3.1.4}$$

We divide by $(n+1)$ instead of n to keep the pseudo-copula data in (3.1.3) in the interior of the unit cube. This is because in some cases the copula density is infinite on the boundaries, thus preventing us from implementing ML [5]. The maximum likelihood estimator (MLE) is defined as the parameter that maximizes the likelihood function. This function is defined in the same manner as with ordinary distribution and density functions, i.e.

$$l(\theta; \hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_n) = \ln L(\theta; \hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_n) = \sum_{t=1}^{n} \ln c_\theta(\hat{\mathbf{U}}_t), \tag{3.1.5}$$

where c_θ is as defined in (2.1.1). We will use the approach with non-parametric estimators for $\hat{F}_i(\cdot)$, known as the pseudo-maximum likelihood [5]. The expressions for the MLE for the Gaussian and Student copula are given in Appendix A.

Chapter 4

Notation and Motivation for Vines

Before we introduce vines in Chapter 5, we introduce some simplifying notation in Section 4.1 to avoid tedious expressions later. In Section 4.2 we show how a multivariate distribution function can be decomposed into pair-copulas and univariate distribution functions. When performing this decomposition, we encounter a new problem which is solved using the h-function described in Section 4.3.

4.1 Notation

Consider a vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$ of random variables with joint distribution function $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$. We write

$$\begin{aligned} f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) &= f_{12\ldots n} \qquad \text{(if necessary } f_{1,2,\ldots,n}\text{)}, \\ f_{X_1,X_2|X_3\ldots,X_n}(x_1, x_2|x_3 \ldots, x_n) &= f_{12|3\ldots n}, \\ C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) &= C_{12\ldots n}, \\ c_{X_1,X_2|X_3}\{F(x_1|x_3), F(x_2|x_3)\} &= c_{12|3}. \end{aligned}$$

Here $c_{12|3}$ is a pair-copula density for the pair of transformed variables $F(x_1|x_3)$ and $F(x_2|x_3)$ and will be discussed in the next sections. Similar simplifying notation will be used where it is found natural, for example with cumulative distribution functions (cdf's).

4.2 Decomposing a Multivariate Distribution Function

We can decompose $f_{12\ldots n}$ in the following (non-unique) way

$$f_{12\ldots n} = f_n \cdot f_{n-1|n} \cdot f_{n-2|n-1,n} \cdots f_{1|2\ldots n}. \tag{4.2.1}$$

Using Definition 2.1.1 of a copula and assuming F to be absolutely continuous with strictly, increasing marginal densities $F_1, \ldots, F_n$, we get

$$\begin{aligned} f_{12\ldots n} &= \frac{\partial F_{12\ldots n}}{\partial x_1 \cdots \partial x_n} \\ &= \frac{\partial C_{12\ldots n}}{\partial x_1 \cdots \partial x_n} \\ &= c_{12\ldots n} \cdot f_1 \cdots f_n. \end{aligned} \tag{4.2.2}$$

The second step is done by using Sklar's theorem and the last step by applying the chain rule. The result in (4.2.2) can be used to represent (4.2.1) with pair-copulas and univariate distribution functions alone. We make use of the following type of factorizations

$$f_{1|2} = \frac{f_{12}}{f_2} = \frac{c_{12} f_1 f_2}{f_2} = c_{12} f_1,$$

$$f_{1|23} = \frac{f_{123}}{f_{23}} = \frac{f_{12|3} f_3}{f_{2|3} f_3} = \frac{c_{12|3} f_{1|3} f_{2|3}}{f_{2|3}} = c_{12|3} c_{13} f_1, \tag{4.2.3}$$

$$f_{1|234} = \frac{f_{1234}}{f_{234}} = \frac{f_{12|34} f_{34}}{f_{2|34} f_{34}} = \frac{c_{12|34} f_{1|34} f_{2|34}}{f_{2|34}} = c_{12|34} c_{13|4} c_{14} f_1. \tag{4.2.4}$$

Note that (4.2.3) and (4.2.4) are not unique in that a change in the conditioning set in step two would give different results, i.e. different pair-copulas in the final results. We see that each term in (4.2.1) can be decomposed by the following iterative procedure [2]

$$f_{x|\mathbf{v}} = c_{xv_j|\mathbf{v_{-j}}}(F_{x|\mathbf{v_{-j}}}, F_{v_j|\mathbf{v_{-j}}}) f_{x|\mathbf{v}_{-j}}, \tag{4.2.5}$$

for a n-dimensional vector $\mathbf{v}$. Here v_j is a component of $\mathbf{v}$, and $\mathbf{v}_{-j}$ is the $\mathbf{v}$-vector without component j. Decomposing a distribution function with four variables could then be done as follows:

$$\begin{aligned} f_{1234} &= f_1 \cdot f_{2|1} \cdot f_{3|12} \cdot f_{4|123} \\ &= f_1 \cdot c_{12} f_2 \cdot c_{23|1} c_{13} f_3 \cdot c_{34|12} c_{24|1} c_{14} f_4 \\ &= c_{34|12} c_{23|1} c_{24|1} c_{12} c_{13} c_{14} \prod_{i=1}^{4} f_i. \end{aligned} \tag{4.2.6}$$

This decomposition was especially chosen so that it would coincide with parts of the vine in Figure 5.1.2 discussed in Chapter 5.

There are 3 different pair-copula decompositions for a three-dimensional distribution function, there are 24 different for the four-dimensional case and as many as 240 for the five-dimensional case [2]. With this is mind, a method helping us find the "best way" to decompose a distribution function would be desirable. That is, we want the decomposition that describes and preserves the (in advance) known information about the dependence structure among the variables as good as possible. This is where the concept of vines is a good aid. Vines will be treated in the next chapter.

4.3 The h-function

In (4.2.5) we need an expression for $c_{xv_j|\mathbf{v}_{-j}}$'s arguments, $F_{x|\mathbf{v}_{-j}}$ and $F_{v_j|\mathbf{v}_{-j}}$. In [10] the following relation is derived (under certain regularity conditions)

$$F_{x|\mathbf{v}} = \frac{\partial C_{xv_j|\mathbf{v}_{-j}}}{\partial F_{v_j|\mathbf{v}_{-j}}}. \tag{4.3.1}$$

Parts of the derivation of (4.3.1) are shown in Appendix B. If $\mathbf{v}$ is univariate, i.e. $\mathbf{v} = v$, we get

$$F_{x|v} = \frac{\partial C_{xv}}{\partial F_v}, \tag{4.3.2}$$

and when x and v are uniform, we define this as the ***h*-function** [2].

$$h(x, v, \Theta) = F_{x|v} = \frac{\partial C_{xv}(x, v, \Theta)}{\partial v}. \tag{4.3.3}$$

In (4.3.3), Θ is the set of parameters for the current copula, and the second parameter of $h(\cdot)$ is the conditioning variable. We define $h^{-1}(x, v, \Theta) = F^{-1}_{x|v}$ as the inverse of $h(x, v, \Theta)$ with respect to x. The most common and applicable copulas, and their h-functions, are discussed in Appendix A.

Chapter 5

Vines

In order to model multivariate data through pair-copula construction in a satisfactory fashion, we introduce the concept of **vines**. Vines is a relatively new concept used to represent multivariate distributions introduced by Bedford and Cooke [4]. It has some resemblance with Markov trees used in Bayesian inference in that it is a hierarchic way of representing the dependence structure among random variables. Vines serve as an aid to take advantage of information known in advance (about the dependency structure), before we divide a multivariate distribution function into bivariate copulas and univariate distribution functions.

In this chapter we give a short introduction to **D-vines** and **canonical vines**. The use of these will be thoroughly accounted for when applying the theory in Chapter 8.

5.1 D-vines and Canonical Vines

A vine is a sequence of trees where the edges in tree T_i are the nodes in tree T_{i+1}. Each vine represents one way of decomposing a multivariate distribution. Two of the most common vines are the D-vine and the canonical vine illustrated for five variables in Figure 5.1.1 and Figure 5.1.2. We will return to these two vines, and subsets of them in examples later. In [4] they choose a slightly different way to illustrate the vines graphic. We choose to follow [2] as it is more self-explanatory.

Each edge in a vine represents a pair-copula corresponding to the label on the edge, e.g. the first edge in T_3 in Figure 5.1.1 corresponds to the copula $c_{14|23}$ (recall that this means $c_{14|23}\{F(x_1|x_2,x_3), F(x_4|x_2,x_3)\}$). The multivariate density (of all the variables in T_1, i.e f_{12345}) could then be decomposed into the product of all pair-copulas in the tree and the marginal densities of all the variables. The two following formulas provide general expressions for the decomposition of a multivariate density using a D-vine and a canonical vine [2].

$$\text{D-vine:} \qquad f_{12\ldots n} = \prod_{k=1}^{n} f_k \underbrace{\prod_{j=1}^{n-1}}_{trees} \underbrace{\prod_{i=1}^{n-j}}_{edges} c_{i,i+j|i+1,\ldots,i+j-1}. \tag{5.1.1}$$

$$\text{Canonical vine:} \qquad f_{12\ldots n} = \prod_{k=1}^{n} f_k \underbrace{\prod_{j=1}^{n-1}}_{trees} \underbrace{\prod_{i=1}^{n-j}}_{edges} c_{j,j+i|1,\ldots,j-i}. \tag{5.1.2}$$

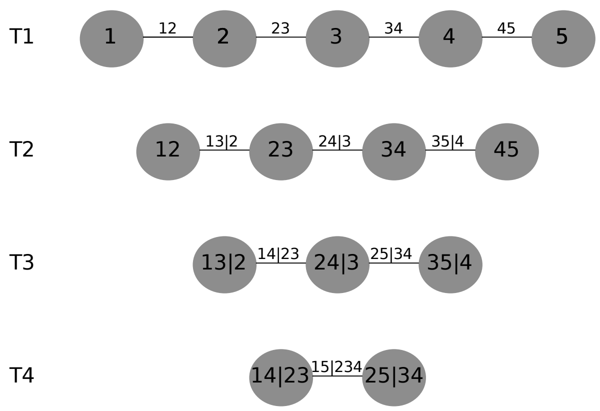


Figure 5.1.1: A D-vine on 5 variables.

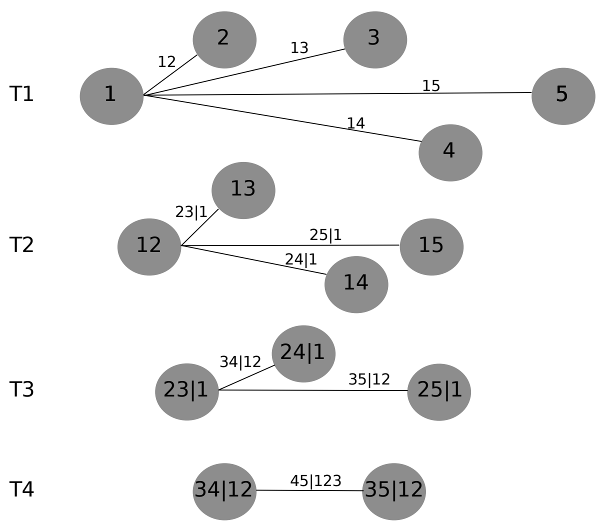


Figure 5.1.2: A canonical vine on 5 variables. Note: If we remove variable 5, we get the decomposition described in (4.2.6).

Both (5.1.1) and (5.1.2) are subsets of what [4] denotes as **regular vines**, a much more extensive class of vines. We will only consider D-vines and canonical vines. As Figure 5.1.1 and Figure 5.1.2 illustrate, they represent different dependency structures. While the canonical vine should be used when there is one variable that (apparently) serves as a main variable, i.e. it interacts with all of the other variables, a D-vine is more appropriate when there is no such variable present. In both cases T_1 is organized in the way that represents the presumed conditional structure in the most natural way. If we use the decompositions in Figure 5.1.1 and 5.1.2, the five-dimensional density f_{12345} can be expressed as:

D-vine:

$$f_{12345} = \underbrace{c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{45}}_{E_1} \cdot \underbrace{c_{13|2} \cdot c_{24|3} \cdot c_{34|4}}_{E_2} \cdot \underbrace{c_{14|23} \cdot c_{25|34}}_{E_3} \cdot \underbrace{c_{15|234}}_{E_4} \cdot \prod_{i=1}^{5} f_i. \tag{5.1.3}$$

Canonical vine:

$$f_{12345} = \underbrace{c_{12} \cdot c_{13} \cdot c_{14} \cdot c_{15}}_{E_1} \cdot \underbrace{c_{23|1} \cdot c_{25|1} \cdot c_{24|1}}_{E_2} \cdot \underbrace{c_{34|12} \cdot c_{35|12}}_{E_3} \cdot \underbrace{c_{45|123}}_{E_4} \cdot \prod_{i=1}^{5} f_i. \tag{5.1.4}$$

Here E_i is the set of edges in T_i, which again are nodes in T_{i+1}. As is seen when looking at (5.1.4) and Figure 5.1.2, when we handle a canonical vine, we can freely choose which variable we want to use as the root node in each tree (up to tree T_{n-2}, where $n = \#$ variables). To decide the root node in T_i, $i > 1$, we can either generate observations form T_{i-1} and find out which copula to use in the root node in T_i, or we can calculate the likelihood for all the remaining decompositions, and choose the structure thereafter. The latter option is better the lower we are in the tree. If we on the other hand use a D-vine, the whole decomposition is determined when we fixate T_1. In both cases there exists $n!/2$ unique vines [2].

Chapter 6

Building a Model

The procedure for sampling and performing inference on vines is not straight-forward. The use of the h-function quickly becomes intricate when increasing the number of dimensions, which is seen when only considering a three-dimensional distribution function in Section 6.3.

In Section 6.1 we briefly discuss how to sample from a vine, before we in Section 6.2 show how a model can be simplified by assuming variables to be conditionally independent. The last two sections of this chapter are devoted to inference and verification of our implementation of the algorithms described here.

The pseudo-codes presented in this section are all collected from [2].

6.1 Sampling From the Vines

We will assume that the margins of the distributions we analyze is uniform, i.e. $U[0,1]$. This does not restrict our possibilities in any way; we could either perform a probability integral transform (PIT) of the variables as discussed in [2], or in some other way transform our variables to ensure that they are $U[0,1]$. The procedure for sampling $U[0,1]$-variables is identical for both the canonical vine and the D-vine.

Algorithm 1 Sampling n dependent $U[0,1]$ variables from a canonical vine or D-vine.

1: Sample $w_1, \dots, w_n$ independent $U[0,1]$
2: $x_1 = w_1$
3: **for** $i = 2, \dots, n$ **do**
4: $\quad x_i = F^{-1}(w_i | x_1, \dots, x_{i-1})$
5: **end for**
6: **return** $x_1, \dots, x_n$

In line 4 in Algorithm 1 we must calculate $F^{-1}(w_i|x_1,\dots,x_{i-1})$. Here we use the h-function defined in (4.3.3), but with different choices of the conditioning variable, v_j, in (4.3.1) for the two types of vines.

$$\text{D-vine, } v_j = x_1\text{:} \quad F(x_j|x_1,\dots,x_{j-1}) = \frac{\partial C_{j,1|2,\dots,j-1}}{\partial F_{1|2,\dots,j-1}}. \tag{6.1.1}$$

$$\text{Canonical vine, } v_j = x_{j-1}\text{:} \quad F(x_j|x_1,\dots,x_{j-1}) = \frac{\partial C_{j,j-1|1,\dots,j-2}}{\partial F_{j-1|1,\dots,j-2}}. \tag{6.1.2}$$

Due to this difference, the algorithms describing how to sample from the two vines are different. To get a picture of how the h-functions are used recursively when calculating the conditional distributions, see Section 6.3 and Appendix C.

Algorithm 2 Generates a sample $x_1, \dots, x_n$ from a D-vine. Here $v_{i,j} = F(x_i|x_1, \dots, x_{j-1})$ and $\Theta_{j,i}$ is the copula-parameter corresponding to the copula $c_{i,i+j|i+1,\dots,i+j-1}$.

Sample $w_1, \dots, w_n$ independent $U[0,1]$
$x_1 = v_{1,1} = w_1$
$x_2 = v_{2,1} = h^{-1}(w_2, v_{1,1}, \Theta_{1,1})$
$v_{2,2} = h(v_{1,1}, v_{2,1}, \Theta_{1,1})$
for $i = 3, \dots, n$ **do**
    $v_{i,1} = w_i$
    for $k = i-1, i-2, \dots, 2$ **do**
        $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,2k-2}, \Theta_{k,i-k})$
    end for
    $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,1}, \Theta_{1,i-1})$
    $x_i = v_{i,1}$
    if $i == n$ **then**
        Stop
    end if
    $v_{i,2} = h(v_{i-1,1}, v_{i,1}, \Theta_{1,i-1})$
    $v_{i,3} = h(v_{i,1}, v_{i-1,1}, \Theta_{1,i-1})$
    if $i > 3$ **then**
        for $j = 2, \dots, i-2$ **do**
            $v_{i,2j} = h(v_{i-1,2j-2}, v_{i,2j-1}, \Theta_{j,i-j})$
            $v_{i,2j+1} = h(v_{i,2j-1}, v_{i-1,2j-2}, \Theta_{j,i-j})$
        end for
    end if
    $v_{i,2i-2} = h(v_{i-1,2i-4}, v_{i,2i-3}, \Theta_{i-1,1})$
end for

6.2 Simplifying the Model; Assuming Conditional Independence

Since we can choose the tree-structure arbitrary, it is beneficial to include variables that are independent, or conditionally independent if they are in tree $T_i,\ i > 1$.

Definition 6.2.1. Two random variables X and Y are **conditionally independent** given an event Z if they are independent in their conditional probability distribution given Z, i.e.

$$P(X \cap Y|Z) = P(X|Z)P(Y|Z), \tag{6.2.1}$$

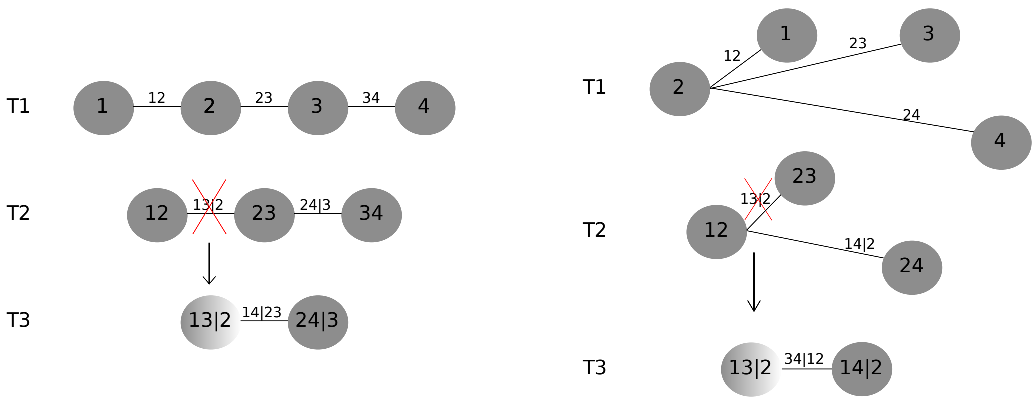
or equivalently

$$P(X|Y \cap Z) = P(X|Z). \tag{6.2.2}$$

We write $X \perp Y|Z$ when X is conditionally independent of Y given Z. If we were to make a model for four variables, and we know that $X_1 \perp X_3|X_2$, it would be natural to use one of the vines in Figure 6.2.1. Note that we in Figure 6.2.1 (b) have used X_2 as

Algorithm 3 Generates a sample $x_1, \dots, x_n$ from a canonical vine. Here $v_{i,j} = F(x_i|x_1, \dots, x_{j-1})$ and $\Theta_{j,i}$ is the copula-parameter corresponding to the copula $c_{j,j+i|1,\dots,j-1}$.

Sample $w_1, \dots, w_n$ independent $U[0,1]$
$x_1 = v_{1,1} = w_1$
for $i = 2, \dots, n$ **do**
  $v_{i,1} = w_i$
  for $k = i-1, i-2, \dots, 1$ **do**
    $v_{i,1} = h^{-1}(v_{i,1}, v_{k,k}, \Theta_{k,i-k})$
  end for
  $x_i = v_{i,1}$
  if $i == n$ **then**
    Stop
  end if
  for $j = 1, \dots, i-1$ **do**
    $v_{i,j+1} = h(v_{i,j}, v_{j,j}, \Theta_{j,i-j})$
  end for
end for



(a) D-vine (b) Canonical vine

Figure 6.2.1: A D-vine (a) and a canonical vine (b) on 4 variables, with $X_1 \perp X_3|X_2$.

the root node in order to be able to use the conditional independence in T_2. We get the following expressions for f_{1234}:

D-vine:

$$f_{1234} = c_{12} \cdot c_{23} \cdot c_{34} \cdot \underbrace{c_{13|2}}_{=1} \cdot c_{24|3} \cdot c_{14|23} \cdot \prod_{i=1}^{4} \underbrace{f_i}_{=1} = c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{24|3} \cdot c_{14|23},$$

Canonical vine:

$$f_{1234} = c_{12} \cdot c_{23} \cdot c_{24} \cdot \underbrace{c_{13|2}}_{=1} \cdot c_{14|2} \cdot c_{34|12} \cdot \prod_{i=1}^{4} \underbrace{f_i}_{=1} = c_{12} \cdot c_{23} \cdot c_{24} \cdot c_{14|2} \cdot c_{34|12}.$$

Here, $f_i = 1$, since we have assumed $X_i \sim U[0,1]$. Since $X_1|X_2$ is independent of $X_3|X_2$, the copula $C_{13|2}(u,v) = uv$, and $c_{13|2} = \frac{\partial}{\partial u \partial v} uv = 1$. In order for us to find variables having this property, we need to analyze the data thoroughly before setting up the structure of the vine , i.e. decomposing the multivariate distribution. It is important to emphasize that the nodes representing conditional independent variables is not removed from the tree, they are simply ignored (set equal to 1) when multiplying together all the pair-copulas. This is visualized by shading the node 13|2 in Figure 6.2.1.

6.3 Sampling From a Three-Dimensional Vine

In this section we derive how to use the h-function recursively when using Algorithm 1 for three-dimensional vines. For a three-dimensional distribution, all three decompositions are both a D-vine and a canonical vine [2]. These are shown in figure 6.3.1.

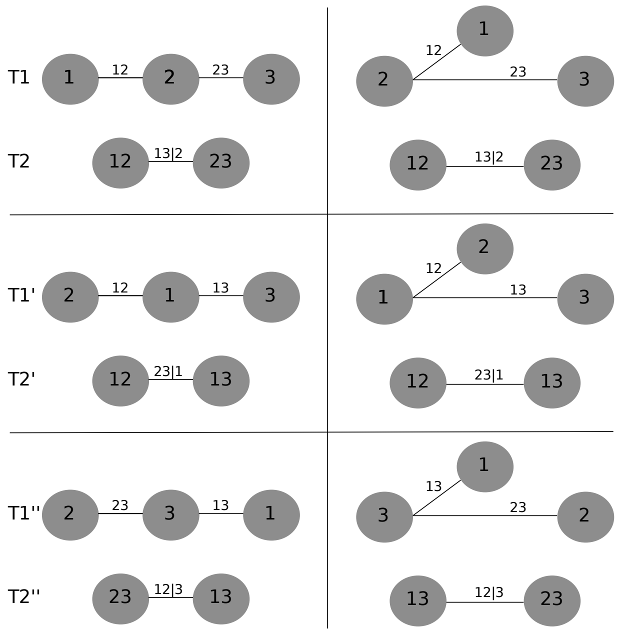


Figure 6.3.1: All three D-vines (left) and canonical vines (right) on 3 variables.

The first two samples are trivial:

$$
\begin{aligned}
x_1 &= w_1, \\
x_2 &= F^{-1}(w_2|x_1) = h^{-1}(w_2, x_1, \Theta_{11}).
\end{aligned}
$$

Recall that the parameter Θ_{ji} is different for the two vines. The indices of Θ are merely aids in the algorithms; they do not benefit the reader when interpreting the h-functions. It is easier to look at the indices of the two other variables in $h(\cdot)$ when deciding which copula Θ corresponds to. For x_3 it gets a bit more intricate. If we choose v_j in (4.3.1) to be x_2, we have

$$
\begin{aligned}
F(x_3|x_1, x_2) &= \frac{\partial C_{x_3 x_2|x_1}(F_{x_3|x_1}, F_{x_2|x_1})}{\partial F_{x_2|x_1}} = \frac{\partial C_{x_3 x_2|x_1}(h(x_3, x_1, \Theta_{12}), h(x_2, x_1, \Theta_{11}))}{\partial h(x_2, x_1, \Theta_{11})} \\
&= h\left[h(x_3, x_1, \Theta_{12}), h(x_2, x_1, \Theta_{11}), \Theta_{21}\right]. \qquad (6.3.1)
\end{aligned}
$$

In (6.3.1) the indices of the Θ's represent the two center vines in Figure 6.3.1, i.e. Θ_{12}, Θ_{11} and Θ_{21} are parameters of c_{13}, c_{12}, and $c_{23|1}$ respectively. This is obvious when looking at the other two arguments of $h(\cdot)$ and the indices of the x's. Recall that $h^{-1}(x, v, \Theta)$ is with respect to x. We want to set $x_i = F^{-1}(w_i|x_1, \ldots, x_{i-1})$, which is equivalent to $w_i = F(x_i|x_1, \ldots, x_{i-1})$. We have

$$
\begin{aligned}
& w_3 = F(x_3|x_1, x_2) = h\left[h(x_3, x_1, \Theta_{12}), h(x_2, x_1, \Theta_{11}), \Theta_{21}\right] \\
\Rightarrow\ & h(x_3, x_1, \Theta_{12}) = h^{-1}(w_3, h(x_2, x_1, \Theta_{11}), \Theta_{21}) \\
\Rightarrow\ & x_3 = h^{-1}\left[h^{-1}(w_3, h(x_2, x_1, \Theta_{11}), \Theta_{21}), x_1, \Theta_{12}\right].
\end{aligned}
$$

The returned values, (x_1, x_2, x_3), would be a sample from both of the two vines in the center of Figure 6.3.1. Examples of expressions for x_4 and x_5, when dealing with a four- and/or five-dimensional vine, are given in Appendix C.

6.4 Inference

In this section we will analyze data on the form $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,T})$, $i = 1, \ldots, n$, where n is the number of variables observed at T points. As mentioned in Section 6.1, we perform the analysis on variables that are $U[0, 1]$. We also assume that the observations at different points are independent over time.

In Chapter 8 we will look at financial returns which certainly are not independent over time. This is easily solved by analyzing the residuals of a (for example) GARCH(1,1) model.

In [2] it is emphasized that when performing inference on real data sets, the margins are unknown, and we have to use approximations of these. The data which are being analyzed are then only approximately independent and uniform. As a consequence, when we maximize the likelihood, only a pseudo-likelihood estimate is achieved. When we later on discuss the likelihood under such circumstances, we are referring to the pseudo-likelihood. The log-likelihoods are given by [2]

D-vine:

$$\sum_{j=1}^{n-1}\sum_{i=1}^{n-j}\sum_{t=1}^{T}\log\left\{c_{j,j+1|1,\dots,j-1}\big(F(x_{j,t}|x_{1,t},\dots,x_{j-1,t}),F(x_{j+i,t}|x_{1,t},\dots,x_{j-1,t})\big)\right\}, \quad (6.4.1)$$

Canonical vine:

$$\sum_{j=1}^{n-1}\sum_{i=1}^{n-j}\sum_{t=1}^{T}\log\left\{c_{i,i+j|i+1,\dots,i+j-1}\big(F(x_{i,t}|x_{i+1,t},\dots,x_{i+j-1,t}),F(x_{i+j,t}|x_{i+1,t},\dots,x_{i+j-1,t})\big)\right\}. \quad (6.4.2)$$

The corresponding algorithms are as given in Algorithm 4 and 5 respectively.

Algorithm 4 Likelihood evaluation for a D-vine.

log-likelihood $= 0$
for $i = 1,\dots,n$ **do**
    $v_{0,i} = x_i$
end for
for $i = 1,\dots,n-1$ **do**
    log-likelihood = log-likelihood + $L(\mathbf{v}_{0,i},\mathbf{v}_{0,i+1},\Theta_{1,i})$
end for
$\mathbf{v}_{1,1} = h(\mathbf{v}_{0,1},\mathbf{v}_{0,2},\Theta_{1,1})$
for $k = 1,\dots,n-3$ **do**
    $\mathbf{v}_{1,2k} = h(\mathbf{v}_{0,k+2},\mathbf{v}_{0,k+1},\Theta_{1,k+1})$
    $\mathbf{v}_{1,2k+1} = h(\mathbf{v}_{0,k+1},\mathbf{v}_{0,k+2},\Theta_{1,k+1})$
end for
$\mathbf{v}_{1,2n-4} = h(\mathbf{v}_{0,n},\mathbf{v}_{0,n-1},\Theta_{1,n-1})$
for $j = 2,\dots,n-1$ **do**
    for $i = 1,\dots,n-j$ **do**
        log-likelihood = log-likelihood + $L(\mathbf{v}_{j-1,2i-1},\mathbf{v}_{j-1,2i},\Theta_{j,i})$
    end for
    if $j == n-1$ **then**
        Stop
    end if
    $\mathbf{v}_{j,1} = h(\mathbf{v}_{j-1,1},\mathbf{v}_{j-1,2},\Theta_{j,1})$
    if $n > 4$ **then**
        for $i = 1,\dots,n-j-2$ **do**
            $\mathbf{v}_{j,2i} = h(\mathbf{v}_{j-1,2i+2},\mathbf{v}_{j-1,2i+1},\Theta_{j,i+1})$
            $\mathbf{v}_{j,2i+1} = h(\mathbf{v}_{j-1,2i+1},\mathbf{v}_{j-1,2i+2},\Theta_{j,i+1})$
        end for
    end if
    $\mathbf{v}_{j,2n-2j-2} = h(\mathbf{v}_{j-1,2n-2j},\mathbf{v}_{j-1,2n-2j-1},\Theta_{j,n-j})$
end for

Algorithm 5 Likelihood evaluation for a canonical vine.

log-likelihood = 0
for $i = 1, \ldots, n$ **do**
    $v_{0,i} = x_i$
end for
for $j = 1, \ldots, n-1$ **do**
    for $i = 1, \ldots, n-j$ **do**
        log-likelihood = log-likelihood + $L(\mathbf{v}_{j-1,1}, \mathbf{v}_{j-1,i+1}, \Theta_{j,i})$
    end for
    if j==n-1 **then**
        Stop
    end if
    for i =1, ..., n-j **do**
        $\mathbf{v}_{j,i} = h(\mathbf{v}_{j-1,i+1}, \mathbf{v}_{j-1,1}, \Theta_{j,i})$
    end for
end for

6.5 Verifying Code

To confirm that our implementation of Algorithm 2 through 5 is correct, we performed the following procedure for both a canonical- and a D-vine.

1. Choose a set of parameters for the vine, $\mathbf{\Theta} = \{\Theta_{11}, \Theta_{12}, \ldots, \Theta_{(n-1),1}\}$, and the corresponding copulas, i.e. types of copulas, $C = \{C_{11}, C_{12}, \ldots, C_{(n-1),1}\}$[1].

2. Generate observations $x_{1,i}, \ldots, x_{n,i},\ i = 1, \ldots, T$ (T large) from the vine, using Algorithm 2 (or 3), and calculate the empirical cdf's in (3.1.4).

3. Find $\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}} l(\mathbf{\Theta})$ using Algorithm 4 (or 5) with $\hat{F}(x_{i,j})$'s as input together with `nlminb()` in `R`.

4. Compare $\hat{\mathbf{\Theta}}$ to $\mathbf{\Theta}$ to assure that the real value of $\mathbf{\Theta}$ approximately achieves the highest log-likelihood value.

Note that in step 3, we must also insert the copulas we used in step 1; The copula types together with the data are both inputs in Algorithm 4 (and 5). A procedure for finding an initial value of $\hat{\mathbf{\Theta}}$ is desirable, and in the next section, the procedure proposed by [2] is accounted for.

6.5.1 Initializing $\hat{\mathbf{\Theta}}$

The following procedure will be used to initialize the values of $\hat{\mathbf{\Theta}}$.

1. Estimate the parameters in tree 1, $\Theta_{11}, \cdots, \Theta_{1(n-1)}$, from the empirical distribution of the observed data, $\{\big(\hat{F}(x_{11}), \ldots, \hat{F}(x_{1T})\big), \ldots, \big(\hat{F}(x_{n1}), \ldots, \hat{F}(x_{nT})\big)\}$. Here n is the number of variables in the model, and T the number of data from each variable.

[1] Recall that the indices on Θ_{ij} indicates which copula we have, and that the indexing is different for canonical vines and D-vines, see Section 6.1. For simplicity we have chosen C_{ij} to be the copula with parameter Θ_{ij} in this section.

2. Calculate observations from tree 2 using the estimated parameters from step 1 together with h-functions and the data, i.e. produce x_{ij}^{T2}, $i = 1, \ldots, n-1$, $j = 1, \ldots, T$.

3. Estimate the parameters in tree 2, $\Theta_{21}, \cdots \Theta_{2(n-2)}$, from the data calculated in step 2.

4. Calculate observations from tree 3 using the estimated parameters from step 3 together with h-functions and x_{ij}^{T2}.

5. Estimate the parameters in tree 3, and continue this way until the whole tree is covered.

Whenever we estimate parameters, we first plot the data and look for typical characteristics, such as the ones in Table 3.1. The initial estimation is performed by built-in functions in `R`.

6.5.2 Numerical Examples of the Verification

Example 1: A Canonical- and D-Vine on Five Variables

We generated $T = \{100, 1000, 5000, 10000\}$ observations respectively from a canonical vine and a D-vine with copulas and parameters as illustrated in Figure 6.5.1. The data were

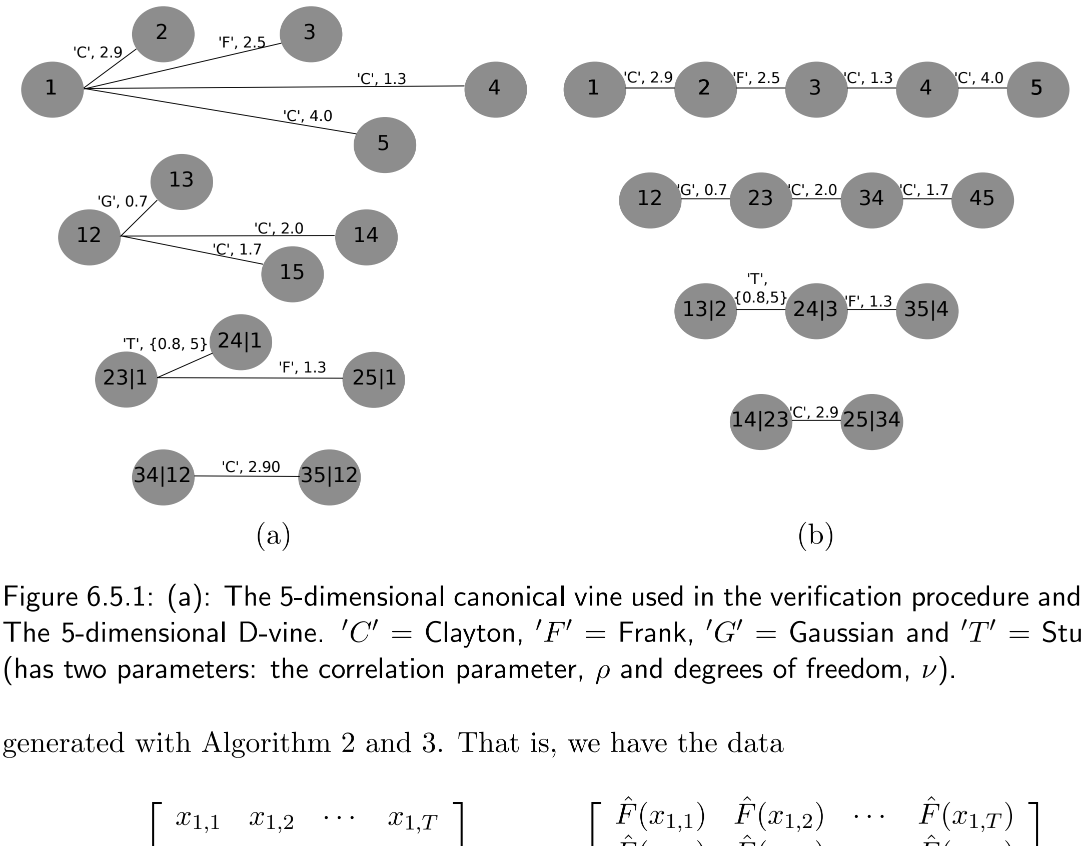


Figure 6.5.1: (a): The 5-dimensional canonical vine used in the verification procedure and (b): The 5-dimensional D-vine. $'C' =$ Clayton, $'F' =$ Frank, $'G' =$ Gaussian and $'T' =$ Student (has two parameters: the correlation parameter, ρ and degrees of freedom, ν).

generated with Algorithm 2 and 3. That is, we have the data

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,T} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,T} \\ \cdot & \cdot & \cdots & \cdot \\ x_{5,1} & x_{5,2} & \cdots & x_{5,T} \end{bmatrix} \Rightarrow \hat{\mathbf{U}} = \begin{bmatrix} \hat{F}(x_{1,1}) & \hat{F}(x_{1,2}) & \cdots & \hat{F}(x_{1,T}) \\ \hat{F}(x_{2,1}) & \hat{F}(x_{2,2}) & \cdots & \hat{F}(x_{2,T}) \\ \cdot & \cdot & \cdots & \cdot \\ \hat{F}(x_{5,1}) & \hat{F}(x_{5,2}) & \cdots & \hat{F}(x_{5,T}) \end{bmatrix}.$$

We used the built-in minimization function `nlminb()` [11] in `R` to find the set of parameters that maximizes[2] the log-likelihood values achieved by Algorithm 4 and 5, i.e.

$$\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}} l(\mathbf{\Theta}), \tag{6.5.1}$$

where $l(\mathbf{\Theta})$ is either the expression in (6.4.1) or (6.4.2). The initial values were obtained by following the procedure in Section 6.5.1 on the whole data set (i.e. using 10.000 observations). The results are listed in Table 6.1 and 6.2.

Table 6.1: Estimated parameters and log-likelihood values for the canonical vine in Figure 8.1.4 (a).

Parameter.	Start	100	1000	5000	10000	Real value
θ_{12}	2.99	1.97	2.76	2.84	2.88	2.90
θ_{13}	2.49	1.87	2.12	2.33	2.46	2.50
θ_{14}	1.35	0.88	1.18	1.26	1.32	1.30
θ_{15}	4.11	3.07	3.86	3.95	4.02	4.00
$\theta_{23\|1}$	0.70	0.67	0.69	0.71	0.70	0.70
$\theta_{24\|1}$	1.97	1.92	1.90	2.02	2.01	2.00
$\theta_{25\|1}$	1.67	1.77	1.63	1.68	1.70	1.70
$\rho_{34\|12}$	0.80	0.77	0.78	0.80	0.81	0.80
$\nu_{34\|12}$	5.52	6.03	5.59	4.60	5.21	5.00
$\theta_{35\|12}$	1.32	1.82	1.30	1.33	1.39	1.30
$\theta_{45\|123}$	2.79	1.74	2.67	2.70	2.74	2.90
$l(\hat{\mathbf{\Theta}})$	36689.79	324.64	3835.84	20072.07	40559.48	39745.50

Table 6.2: Estimated parameters and log-likelihood values for the D-vine in Figure 8.1.4 (b).

Parameter.	Start	100	1000	5000	10000	Real value
θ_{12}	2.88	3.23	2.46	2.98	2.84	2.90
θ_{23}	2.46	2.41	2.00	2.45	2.50	2.50
θ_{34}	1.32	1.40	1.21	1.26	1.34	1.30
θ_{45}	4.02	4.23	3.96	3.95	4.05	4.00
$\theta_{13\|2}$	0.70	0.72	0.72	0.69	0.70	0.70
$\theta_{24\|3}$	2.01	2.08	1.95	2.06	1.93	2.00
$\theta_{35\|4}$	1.70	1.07	1.56	1.76	1.66	1.70
$\rho_{14\|23}$	0.81	0.77	0.80	0.79	0.80	0.80
$\nu_{14\|23}$	5.21	300.00	6.13	5.46	4.58	5.00
$\theta_{25\|34}$	1.39	1.10	1.10	1.31	1.30	1.30
$\theta_{15\|234}$	2.74	1.62	2.13	2.63	2.88	2.90
$l(\hat{\mathbf{\Theta}})$	39799.67	370.41	3786.04	20200.54	40636.02	39911.08

It is clear that the higher the number of observations, the closer $\hat{\mathbf{\Theta}}$ gets to $\mathbf{\Theta}$. This is also seen from the error-plot in Figure 6.5.2. As T increases, $\hat{\mathbf{\Theta}}$ clearly converges to $\mathbf{\Theta}$, and hence verifies our code.

[2] In our implementation we return the negative value of the log-likelihood.

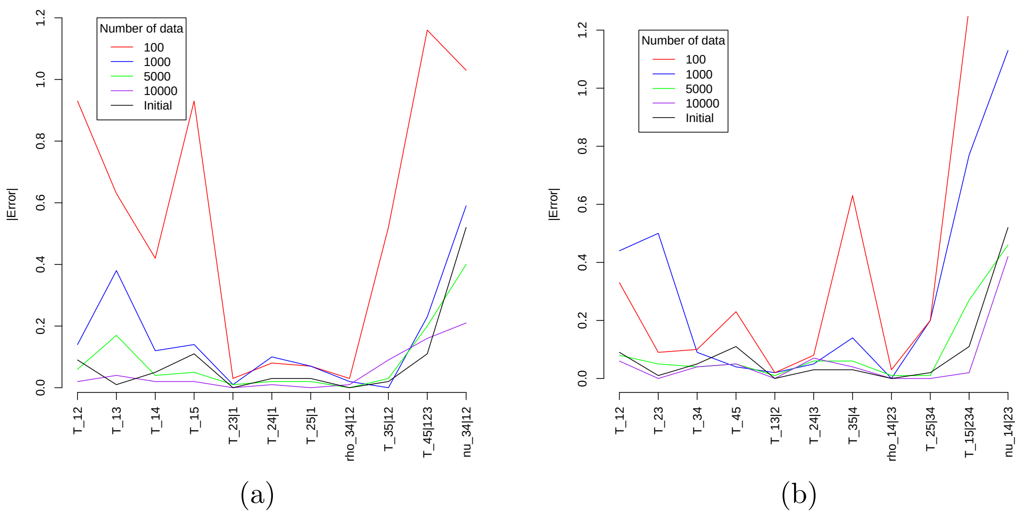


Figure 6.5.2: (a): $|\mathbf{\Theta} - \hat{\mathbf{\Theta}}|$ for the parameters in the canonical vine.(b): $|\mathbf{\Theta} - \hat{\mathbf{\Theta}}|$ for the parameters in the D-vine. Note that $\nu_{14|23}$ is left out for the case with 100 data. This is since it reached the maximum allowable number, and contains little information about the fit.

It is interesting to notice that the initial value is actually closer to the correct value for some parameters, even in the case with 10.000 data. This enlightens the fact that when we optimize the log-likelihood value of the vine, we do not get the local maximum, i.e. Θ_{ij} is not necessary the parameter(s) maximizing the likelihood for c_{ij}, but we get the global maximum. By analyzing the error plots in Figure 6.5.2, we also see that the parameter far down in the vines are harder to estimate accurately.

Example 2: A Decomposition Consisting of Student Copulas

In this example we verify our code for a decomposition using a D-vine consisting of student copulas exclusively. This case is treated separately since it has two parameters for each copula and is therefore more time consuming and somewhat difficult to estimate. The Student distribution/copula is also a type of model/decomposition that is often used in financial models, which we will look into in Chapter 8. Again, we estimate parameters for data sets with different sizes to get a sense of how fast $\hat{\mathbf{\Theta}}$ converges to $\mathbf{\Theta}$. Our initial values were $\rho_{initial} = (\rho_{ij} - 0.1)$ and $\nu_{initial} = \{3.5, 9, 10, 190, 125, 15\}$, where

$$\mathbf{\Theta} = \begin{bmatrix} \Theta_{12} = \{-0.25, 4\} & \Theta_{23} = \{0.47, 10\} & \Theta_{34} = \{-0.17, 12\} \\ \Theta_{13|2} = \{-0.11, 200\} & \Theta_{24|3} = \{0.02, 130\} & \\ \Theta_{14|23} = \{0.29, 16\} & & \end{bmatrix}, \quad \Theta_{ij} = \{\rho_{ij}, \nu_{ij}\}.$$

When we said that it is somewhat difficult to estimate the parameters in the Student copula, we were referring to the degrees of freedom (df). As the number of df increases, the dependence between the variables decline [2]. We also know that as the number of df increases, the Student distribution converges to the standard normal distribution [7]. Hence, for high values of the df, it would be natural to assume that the difference between different models would be somewhat insignificant. For example, the difference between choosing ν to, say, 100 rather than 300, would not make a great impact in the fitting procedure. The process of maximizing the log-likelihood value with respect to the df also has a huge impact on the number of iterations performed in `nlminb()`. We therefore tried to penalize the log-likelihood value when ν was large, in order to check if this would make $\hat{\mathbf{\Theta}}$ converge any faster. We chose to introduce the penalizing term

$$\text{likelihood} = \text{likelihood} \cdot e^{-\frac{1}{2}\max(\nu-15,0)^2},$$

or equivalently

$$\text{log-likelihood} = \text{log-likelihood} - \frac{1}{2}\max(\nu - 15, 0)^2,$$

whenever computing the contribution to the log-likelihood from a Student copula in the vine[3]. The results with and without this term are summarized in Table 6.3.

Table 6.3: Parameters returned when optimizing the log-likelihood with respect to $\mathbf{\Theta}$. Initial values: $\rho_{initial} = (\rho_{ij} - 0.1)$ and $\nu_{initial} = \{3.5, 9, 10, 190, 125, 15\}$.

	$\nu > 15$ penalized			$\nu \in (2, 300]$		
Parameter	100	1500	5000	100	1500	5000
ρ_{12}	-0.37	-0.20	-0.27	-0.37	-0.20	-0.27
ρ_{23}	0.46	0.48	0.46	0.46	0.48	0.46
ρ_{34}	-0.31	-0.15	-0.17	-0.32	-0.15	-0.17
$\rho_{13\|2}$	-0.27	-0.13	-0.12	-0.27	-0.13	-0.12
$\rho_{24\|3}$	-0.19	0.01	0.02	-0.19	0.02	0.02
$\rho_{14\|23}$	-0.06	0.30	0.31	0.05	0.30	0.31
ν_{12}	2.35	3.26	4.20	2.38	3.26	4.29
ν_{23}	2.62	7.93	9.36	2.64	7.97	9.54
ν_{34}	15.04	5.59	9.82	300.00	5.42	10.41
$\nu_{13\|2}$	10.77	15.09	15.86	11.29	24.63	77.29
$\nu_{24\|3}$	15.03	15.25	16.03	300.00	283.42	300.00
$\nu_{14\|23}$	3.65	15.14	11.42	3.38	29.92	11.08
$l(\hat{\mathbf{\Theta}})$	34.21	263.86	1311.04	35.15	266.88	1326.48

Comparing the results in Table 6.3, it is clear that in both procedures, $\hat{\rho}_{ij}$ converges to ρ_{ij}, while $\hat{\nu}$ seems to be more difficult to fit accurate. This is also seen from the error plot in Figure 6.5.3 where only the method without the penalizing term is plotted. We experienced that the procedure penalizing the likelihood when $\nu > 15$ actually performs far more iterations than the one without the penalizing term. It would therefore seem natural to reject this approach, and let ν vary in a larger interval, say $\nu \in (2, 300]$ (we must have $\nu > 2$ due to the expression of variance for a Student distributed variable). Unless otherwise stated, we use the constraint $\nu \in (2, 300]$.

[3]We penalize the log-likelihood function with subtraction since nlminb() minimizes the log-likelihood.

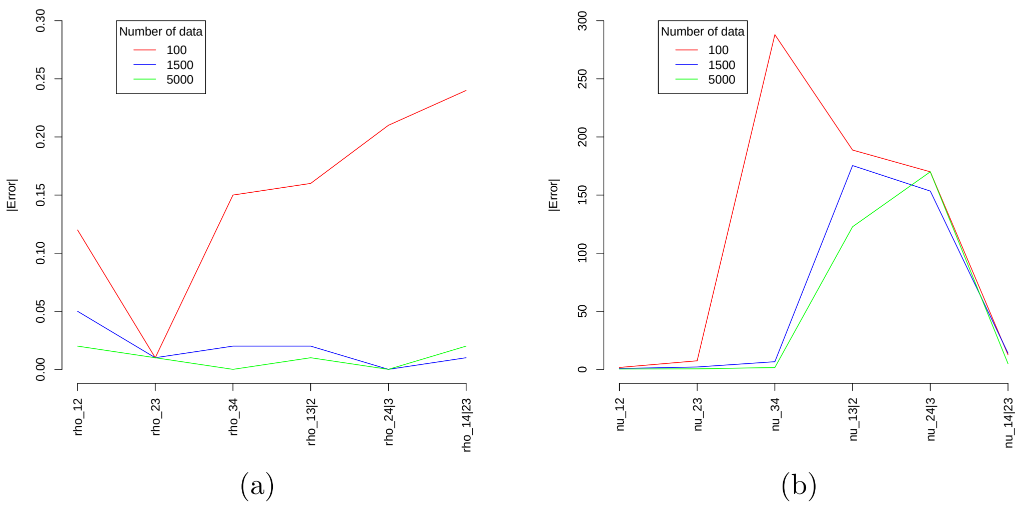


Figure 6.5.3: Error plot for ν and ρ without penalizing large ν's. (a): $|\rho_{ij} - \hat{\rho}_{ij}|$.(b): $|\nu_{ij} - \hat{\nu}_{ij}|$.

Chapter 7

Complications and Practical Problems Applying the Theory

When applying the theory discussed in the previous chapters, we run into some more or less complicated problems. In this chapter we discuss how to cope with them. We consider the cases in the order we run into them when using real data sets in Chapter 8; first we look at how we decide which vine we use and what structure it has in Section 7.1. The rest of the chapter is devoted to goodness-of-fit analysis.

7.1 Deciding the Structure of the Vine

The first thing we have to decide is whether to use a canonical vine or a D-vine and how to order the variables. The easiest way would of course be to test all of the $n!$ possible structures ($n!/2$ for each type) and choose the one with the highest likelihood. In higher dimensions this becomes extremely costly, and an aid to help us decide the structure in tree 1 is desirable. First, we check if there is a root variable. This can be done by calculating the correlation coefficient, Kendall's tau or Spearman's rho for all of the $\binom{n}{2}$ pairs of variables, and see if one of the variables dominates. This can be checked by using a table like Table 7.1 (with the desired dependence measure), and choose the variables with highest absolute value in such a manner that all variables are used. If one of the variables is used in all the chosen pairs, this is the root node in tree 1. This procedure also sets up the structure for the D-vine, as described below.

Table 7.1: Example of how to choose which vine to use using dependency measures.

	Case 1			Case 2		
	X_1	X_2	X_3	X_1	X_2	X_3
X_4	0.3	**0.76**	0.54	**0.80**	**0.76**	0.67
X_1	-	**0.70**	**0.67**	-	**0.70**	0.54
X_2		-	0.45		-	0.45

In case 1 in Table 7.1, we would prefer a vine with the copulas C_{24}, C_{12} and C_{13} in tree 1. This is a D-vine with the structure $X_4 \leftrightarrow X_2 \leftrightarrow X_1 \leftrightarrow X_3$ in tree 1. In case 2 however, we cannot choose the three pairs with highest correlation. This arrangement would leave out X_3. In stead, C_{12} is substituted with C_{34}, making us choose a canonical vine with

X_4 as root node in tree 1. This method works fine when investigating copulas with only one parameter, but when dealing with for instance financial data, the copulas tend to be Student copulas. In this case we have two parameters describing the dependency in the copula; the degrees of freedom and the correlation parameter. We will discuss this case when we run into it in Section 8.1.

7.2 Goodness-of-Fit

As we will see when applying the theory covered in this paper on real data in Chapter 8, it is not always easy to decide which type of copula to use. Even though practitioners tend to use Student copulas on financial returns, it is not necessarily the case for us when we decompose the model into pair-copulas. As described in Section 6.4, we calculate the log-likelihood for the whole vine when fitting parameters to the data. However, we run into a problem when we want to compare the log-likelihood values achieved by vines consisting of different types of copulas. For instance if we have a vine consisting merely of Student copulas and we change one of them into a Clayton copula. The Clayton and Student copulas are not nested in each other, and we cannot compare their log-likelihood values directly [2].

A method for comparing the fit of copulas to data, i.e. Goodness-of-Fit (g-o-f) procedures, are needed in order to decide which copulas to use in the vine. In this section we look closer on the hypothesis

$$\mathrm{H}_0 : \ C \in \mathcal{C} = \{C_\theta; \theta \in \Theta\} \text{ vs. } \mathrm{H}_1 : \ C \notin \mathcal{C} = \{C_\theta; \theta \in \Theta\}, \tag{7.2.1}$$

where Θ is the parameter space. In [3], (7.2.1) is analyzed in depth with nine different approaches and many different choices of $\mathcal{C}$. Each approach is examined with different null hypotheses and true copulas. Hence one can find the one most suitable approach for specific cases. The recommendations given there are followed and used in this thesis.

7.2.1 G-o-F Approaches

We focus on approaches that is found to reject Student and Clayton copulas most often when the data does not arise form these copulas, i.e. the null hypothesis is false. We focus on these types of approaches since the data we shall analyze in Chapter 8 is found to fit either Student or Clayton copulas, but the distinction between them is not easily seen from plots alone.

Approach 1: Dimension Reduction

This approach is what is called a dimension reduction approach. That is, the multivariate problem is reduced to a univariate problem, which is preferred from a numeric perspective. This is however not a very important issue in this thesis as we always have bivariate copulas. This approach uses the empirical copula defined as

$$\hat{C}(\mathbf{u}) = \frac{1}{n+1} \sum_{j=1}^{n} I(Z_{j1} \leq u_1, \ldots, Z_{jd} \leq u_d), \tag{7.2.2}$$

where $\mathbf{z}_j = (z_{j1}, \ldots, z_{jd}) = \left(\frac{R_{j1}}{n+1}, \ldots, \frac{R_{jd}}{n+1}\right)$. R_{ji} is the rank of x_{ji} amongst $(x_{1i}, \ldots, x_{ni})$, i.e. it is the same as using the empirical cdf defined in (3.1.4). The test statistic is based on measuring the distance between the empirical- and null hypothesis distribution functions, i.e.

$$\hat{T}_1 = n \int_{[0,1]^d} \left\{ \hat{C}(\mathbf{z}) - C_{\hat{\theta}}(\mathbf{z}) \right\}^2 \, \mathrm{d}\hat{C}(\mathbf{z}) = \sum_{j=1}^{n} \left\{ \hat{C}(\mathbf{z}_j) - C_{\hat{\theta}}(\mathbf{z}_j) \right\}^2 . \tag{7.2.3}$$

Here, $\hat{C}(\mathbf{z})$ is compared to our estimate $C_{\hat{\theta}}(\mathbf{z})$ of C_θ. T_1 is a Cramé-von Mises (CvM) statistic, i.e. it is a form of minimum distance estimation [3]. This approach is recommended in [3] when testing

$$\mathrm{H}_0: \; C \in \mathcal{C} = \{C_\theta^{Student}; \theta \in \Theta\} \text{ vs. } \mathrm{H}_1: \; C \notin \mathcal{C} = \{C_\theta^{Student}; \theta \in \Theta\}, \tag{7.2.4}$$

but also gives good results when testing the hypothesis in (7.2.5).

Approach 2: Unbiased Estimators of Θ

This approach is only used for testing the possibility of the data arising from a Clayton copula, i.e.

$$\mathrm{H}_0: \; C \in \mathcal{C} = \{C_\theta^{Clayton}; \theta \in \Theta\} \text{ vs. } \mathrm{H}_1: \; C \notin \mathcal{C} = \{C_\theta^{Clayton}; \theta \in \Theta\}. \tag{7.2.5}$$

It is based on two unbiased estimators of θ, the dependence parameter in the Clayton copula, see Appendix A. We have θ_τ, based on the relationship between Kendall's tau and θ described in Table 3.2, and θ_W, a weighted rank-based estimator.

$$\hat{\theta}_\tau = \frac{2\hat{\tau}}{1-\hat{\tau}}, \tag{7.2.6}$$

$$\hat{\theta}_W = \frac{\sum_{i<j} \Delta_{ij}/W_{ij}}{\sum_{i<j}(1-\Delta_{ij})/W_{ij}}, \tag{7.2.7}$$

where $\hat{\tau} = -1 + 4\sum_{i<j} \Delta_{ij}/(n(n-1))$, $\Delta_{ij} = I\{(Z_{i1} - Z_{j1})(Z_{i2} - Z_{j2})\}$ and $W_{ij} = \sum_{k=1}^{n} I\{Z_{k1} \le \max(Z_{i1}, Z_{j1}), Z_{k2} \le \max(Z_{i2}, Z_{j2})\}$. In two dimensions, the test statistic is

$$\hat{T}_2 = (\hat{\theta}_\tau - \hat{\theta}_W)^2. \tag{7.2.8}$$

7.2.2 Testing Procedure

To obtain P-value estimates for the tests of hypotheses, we use a parametric bootstrap method adopted from [3]. Pseudo code for a total of nine approaches, including the two we use, can be found in Appendix C in [12]. We have verified our implementation of approach 1 and 2 in Appendix D.

Approach 1

1 Convert the data into normalized ranks, $\mathbf{z}_1, \ldots, \mathbf{z}_n$, i.e. use (3.1.4) on the data.

2 Estimate the parameters θ with a consistent estimator $\hat{\theta} = \hat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

3 Compute $\hat{C}(\mathbf{z})$ according to (7.2.2).

4 If there is an analytical expression for C_θ, compute $\hat{T}_1$ by (7.2.3) with $\hat{C}(\mathbf{z})$ and $C_{\hat{\theta}}(\mathbf{z})$. Go to step 5.
It there is no analytical expression for C_θ, choose $N_b \ge n$ and carry out a double bootstrap:

(i) Generate a random sample $(\mathbf{x}_1^*, \ldots, \mathbf{x}_{N_b}^*)$ from the null hypothesis copula $C_{\hat{\theta}}$ and compute the corresponding pseudo-sample $(\mathbf{z}_1^*, \ldots, \mathbf{z}_{N_b}^*)$ using the ranks as described in step 1.

(ii) Approximate $C_{\hat{\theta}}$ by $C_{\hat{\theta}}^*(\mathbf{u}) = \frac{1}{N_b+1}\sum_{l=1}^{N_b} I(\mathbf{z}_l^* \leq \mathbf{u}),\ \mathbf{u} \in [0,1]^d$.

(iii) Approximate (7.2.3) by $\hat{T}_1 = \sum_{j=1}^{n}\left\{\hat{C}(\mathbf{z}_j) - C_{\hat{\theta}}^*(\mathbf{z}_j)\right\}^2$.

5 For some large integer K repeat the following for $k = 1, \ldots, K$ (parametric bootstrap).

(a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\hat{\theta}}$ and compute the corresponding pseudo-sample $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ using the ranks as described in step 1.

(b) Estimate the parameters θ^0, with a consistent estimator $\hat{\theta}_k^0 = \hat{\mathcal{V}}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$. We use MLE, but others can be used, see [3] for a discussion.

(c) Let $\hat{C}_k^0(\mathbf{u}) = \frac{1}{n+1}\sum_{j=1}^{n} I(\mathbf{z}_{j,k}^0 \leq \mathbf{u}),\ \mathbf{u} \in [0,1]^d$.

(d) If there is an analytical expression for C_θ, let $\hat{T}_{1,k}^0 = \sum_{j=1}^{n}\left\{\hat{C}_k^0(\mathbf{z}_{j,k}^0) - C_{\hat{\theta}_k^0}(\mathbf{z}_{j,k}^0)\right\}^2$. Go to step 6.
If there is no analytical expression for C_θ, choose $N_b \geq n$ and proceed:

(i) Generate a random sample $(\mathbf{x}_{1,k}^{0*}, \ldots, \mathbf{x}_{N_b,k}^{0*})$ from the null hypothesis copula $C_{\hat{\theta}_k^0}$ and compute the corresponding pseudo-sample $(\mathbf{z}_{1,k}^{0*}, \ldots, \mathbf{z}_{N_b,k}^{0*})$ using the ranks as described in step 1.

(ii) Approximate $C_{\hat{\theta}_k^0}$ by $C_{\hat{\theta}_k^0}^{0*}(\mathbf{u}) = \frac{1}{N_b+1}\sum_{l=1}^{N_b} I(\mathbf{z}_{l,k}^{0*} \leq \mathbf{u}),\ \mathbf{u} \in [0,1]^d$.

(iii) Approximate (7.2.3) by $\hat{T}_{1,k}^* = \sum_{j=1}^{n}\left\{\hat{C}_k^0(\mathbf{z}_{j,k}^0) - C_{\hat{\theta}_k^0}^{0*}(\mathbf{z}_{j,k}^0)\right\}^2$.

6 An approximate P-value for approach 1 is then given by $\hat{p} = \frac{1}{K+1}\sum_{k=1}^{K} I(\hat{T}_{1,k}^0 > \hat{T}_1)$.

Approach 2

1 Convert the data into normalized ranks, $\mathbf{z}_1, \ldots, \mathbf{z}_n$, i.e. use (3.1.4) on the data.

2 Estimate the parameters θ with a consistent estimator $\hat{\theta} = \hat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

3 Estimate $\hat{\theta}_\tau$ and $\hat{\theta}_W$ according to (7.2.6) and (7.2.7).

4 Compute $\hat{T}_2$ according to (7.2.8).

5 For some large integer K repeat the following for $k = 1, \ldots, K$.

(a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\hat{\theta}}$ and compute the corresponding pseudo-sample $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ using the ranks as described in step 1.

(b) Estimate $\hat{\theta}_{\tau,k}^0$ and $\hat{\theta}_{W,k}^0$ according to (7.2.6) and (7.2.7).

(c) Compute $\hat{T}_{2,k}^0$ according to (7.2.8).

6 An approximate P-value for approach 2 is then given by $\hat{p} = \frac{1}{K+1}\sum_{k=1}^{K} I(\hat{T}_{2,k}^0 > \hat{T}_2)$.

Chapter 8

Numerical Experiments

In this chapter we analyze real data sets using the techniques described in the previous chapters. In Section 8.1 we look at four indices on world wide stocks, while we in Section 8.2 look at financial returns on intra day level.

8.1 Fitting Bivariate Copulas to Financial Returns

We chose to examine the following four indices:

- The Oslo Stock Exchange Benchmark Index (denoted O), an investable index, which comprises the most traded shares listed on the Oslo Stock Exchange.
- S&P 500 index (denoted S), a value weighted index of the prices of 500 large cap common stocks actively traded in the United States.
- The FTSE 100 Index (denoted F), also called the "footsie", is a share index of the 100 most highly capitalized UK companies listed on the London Stock Exchange.
- Dow Jones EURO STOXX 50 (denoted E), a stock index of Eurozone stocks designed by STOXX Ltd, a joint venture of Deutsche Boerse AG, Dow Jones & Company and the SWX Group. According to STOXX, its goal is "to provide a blue-chip representation of Super sector leaders in the Eurozone [13]."

We looked at the period from 06.06.2002 to 06.06.2006, that is we have 934 points[1]. Figure 8.1.1 shows the log-returns for each pair of asset. The only pair that stand out is Euro 50 and Footsie, which seem to have a linear relationship. The further analysis is performed on the empirical distribution function (3.1.4) for the standardized residuals of a GARCH(1,1)-model on the log-returns. These are plotted in Figure 8.1.2, and it is clear that all pairs have both lower and upper tail dependence.

We have plotted the autocorrelation of the residuals in Figure 8.1.3, and we can see that there is little, or none, autocorrelation left. As a first approach to break down the model describing all four indices using only pair-copulas, we fitted a bivariate Student copula to each of the pairs. Here we used the built-in R-function `fitCopula()`, with maximum likelihood, to achieve the best fit. The df and ρ for each pair are listed in Table 8.1 and 8.2.

[1]Due to different public holidays in the different regions, some quotes do not exist for the all indices. Whenever one index is "missing" a quote, we remove the date in question for all four indices.

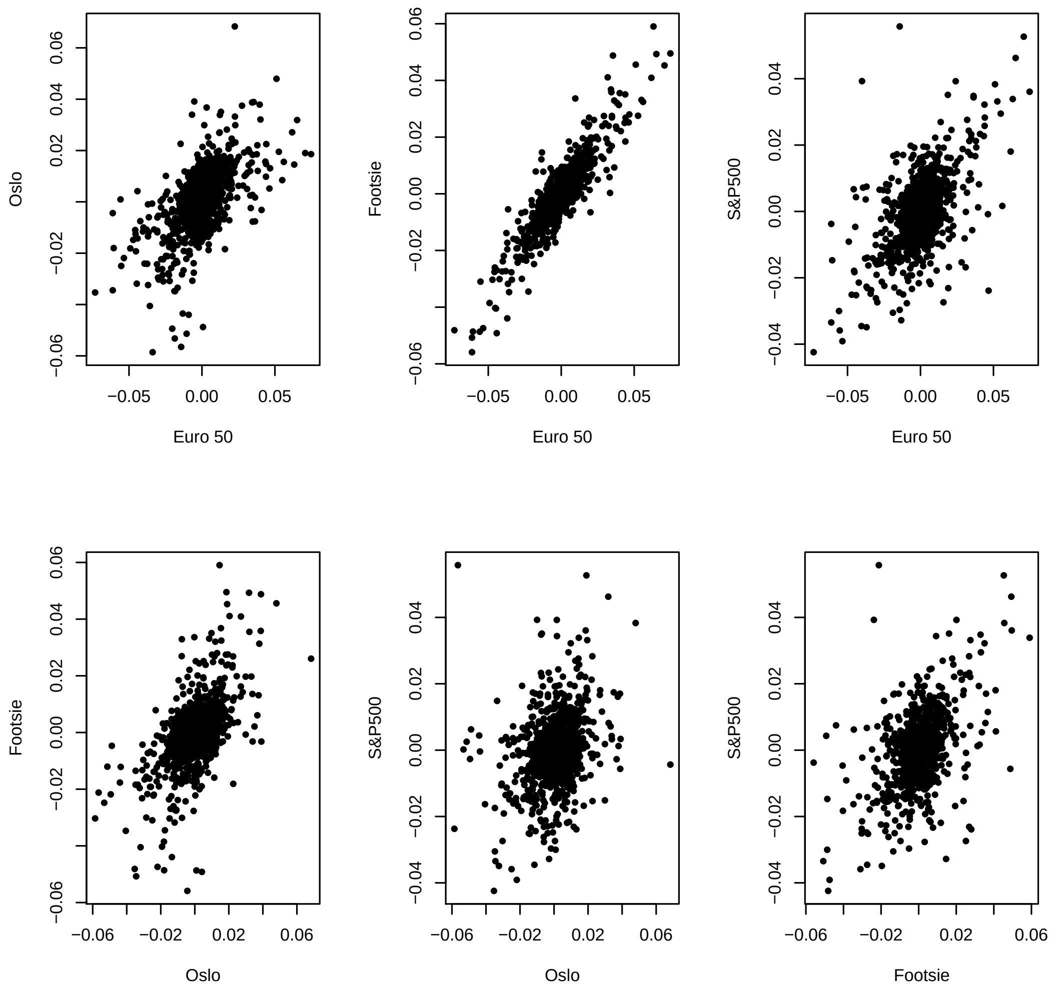


Figure 8.1.1: Log-returns for all the pairs of indices in the period 06.06.2002-06.06.2006.

Table 8.1: The numbers of degrees of freedom for each pair of Student copula. The three lowest values are highlighted: These are the pairs used in tree 1 in the D-vine we fitted the data.

	Oslo (O)	Footsie (F) (D)	S&P500 (S)
Euro 50 (E)	17.88	10.33	**4.38**
Oslo	-	**8.16**	10.84
Footsie		-	**5.76**

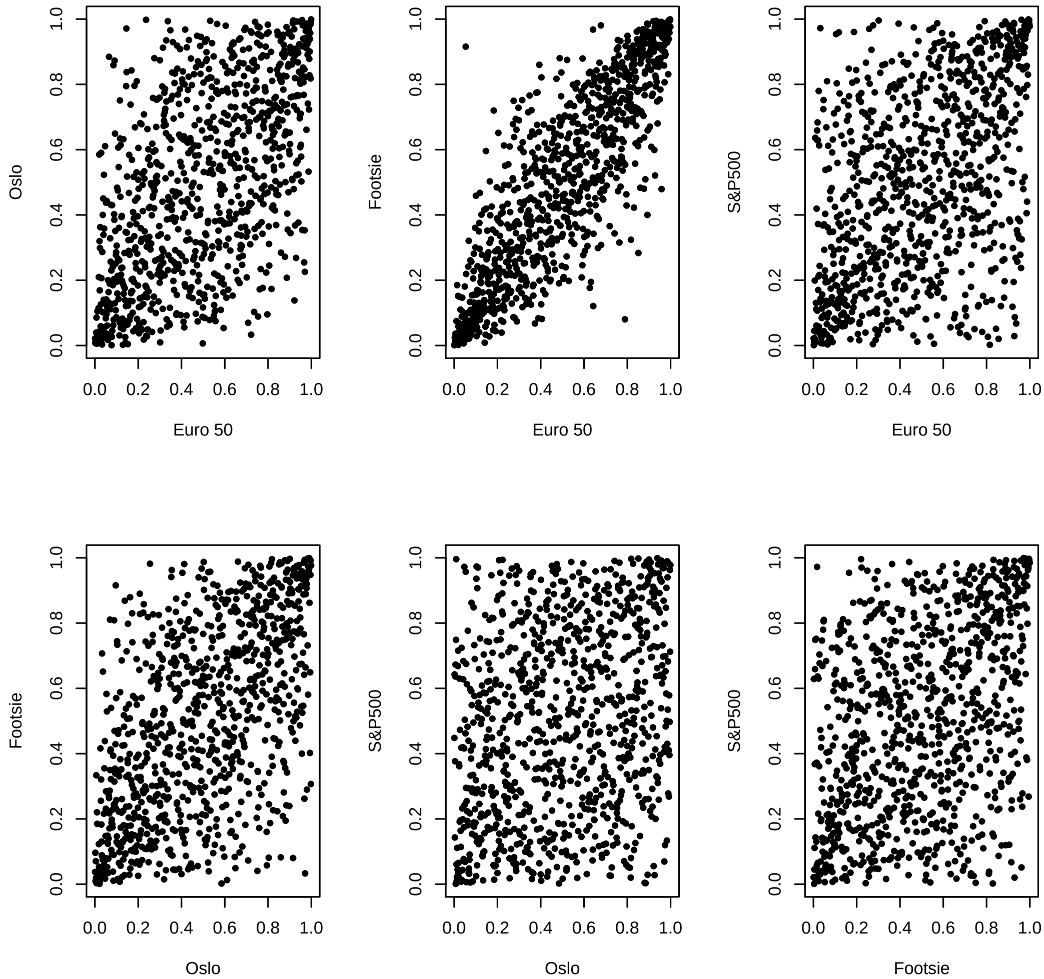
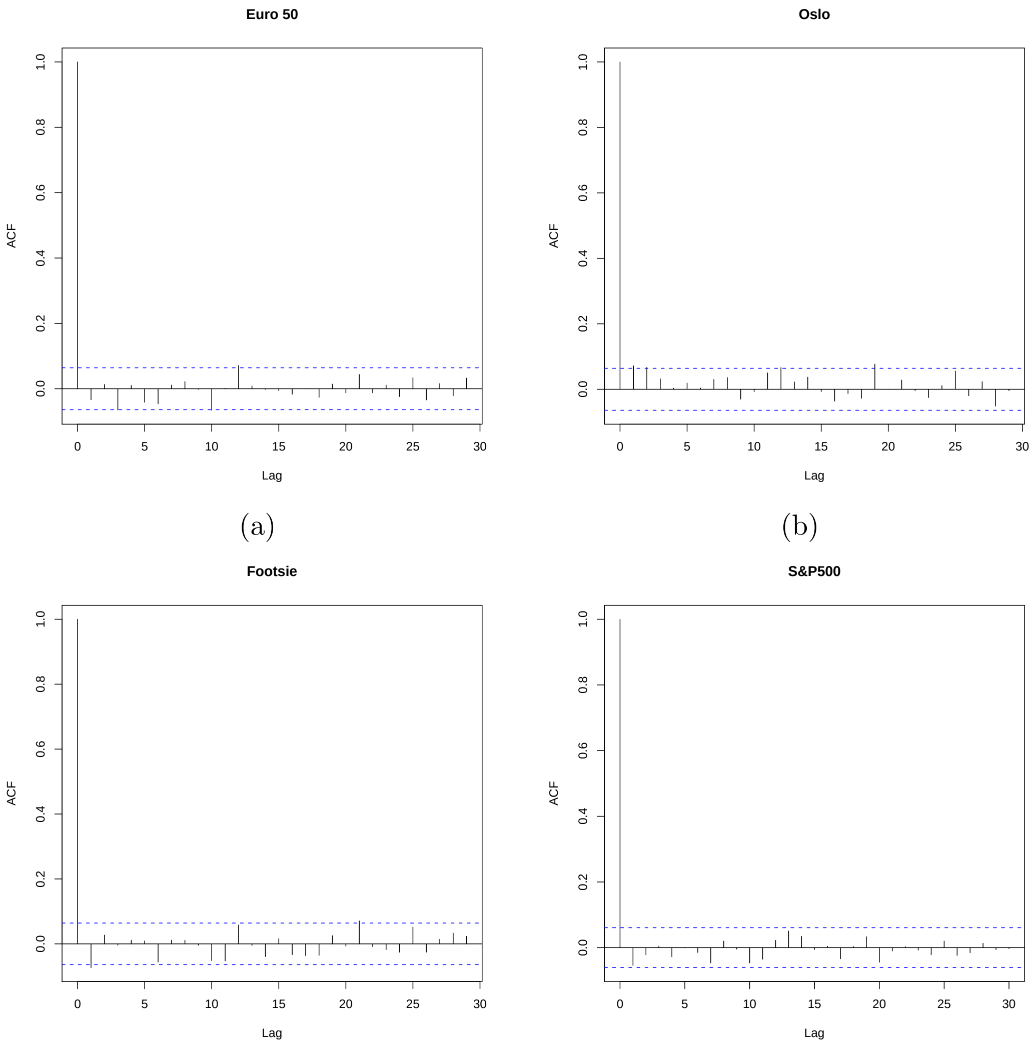


Figure 8.1.2: The empirical distribution function of the residuals of the GARCH(1,1) model on the log-returns for all the pairs of indices in the period 06.06.2002-06.06.2006.

Table 8.2: The correlation coefficient, ρ, for each pair of Student copula. The three highest values (making a pairwise decomposition possible) are highlighted: These are the pairs used in tree 1 in the canonical vine we fitted the data.

	Oslo (O)	Footsie (F)	S&P500 (S)
Euro 50 (E)	**0.60**	**0.86**	**0.49**
Oslo	-	0.59	0.29
Footsie		-	0.43



(a) (b) (c) (d)

Figure 8.1.3: The autocorrelation for the four indices. The significance level for testing if the autocorrelations are zero is set to 0.05 (the dotted line).

.

We know that a low number of df indicates strong dependence between the variables [2], so if we were to choose the three pairs with lowest number of df, we would fit a D-vine with $E \leftrightarrow S \leftrightarrow F \leftrightarrow O$ as the ordering in tree 1. If we on the other hand select the three pairs with largest value of ρ, we would fit a canonical vine with Euro 50 as the kernel variable in tree 1. These choices are highlighted in Table 8.1 and 8.2. We note that even though Euro 50 and Footsie had by far the highest linear correlation, they are only linked in the canonical vine, i.e. when we used ρ to decide the structure. Since there is nothing ruling in favor of neither vine, we fit both a canonical and a D-vine and compare how they preserve the ν's and ρ's in Table 8.1 and 8.2. Using EF as the kernel in tree 2 gave a slightly higher log-likelihood value than the other two possibilities, i.e. ES and EO, when fitting a canonical vine. The structure of the two vines is shown in Figure 8.1.4, and the final parameters are listed together with the initial values in Table 8.3.

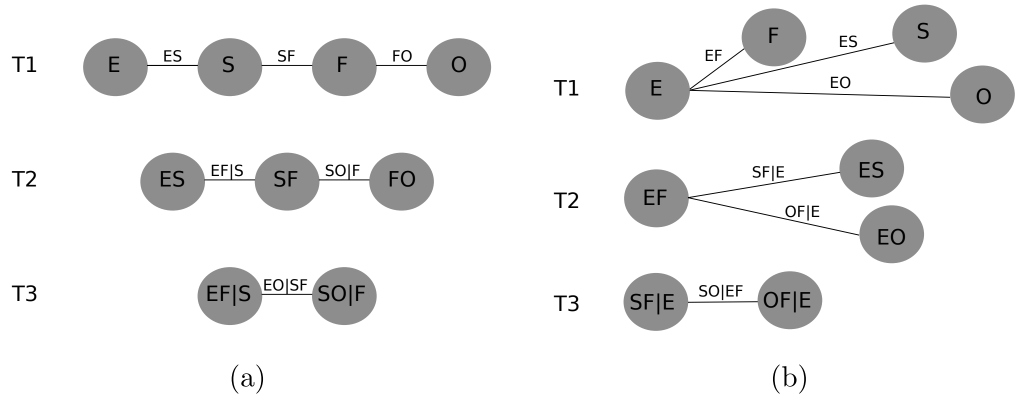


Figure 8.1.4: (a): The structure of the D-vine fitted to the data. (b): The structure of the canonical vine fitted to the data.

8.1.1 Verification of the Models

To check how well the two models captured the relationship between the indices, we drew 10^5 observations from the fitted vines, and fitted a bivariate Student copula to each of the six possible pairs as we did in the first step in Section 8.1. We then checked how well the df and ρ's were preserved. The results are listed in Table 8.4 and 8.5. By comparing the simulated values in Table 8.5 with the values obtained from the data set in Table 8.2, it is evident that both vines preserves both the ordering and the values of ρ remarkably well even between the pairs we did not directly model. The difference between the vines arises when we compare the simulated df in Table 8.4 with those in Table 8.1. The canonical vine has almost kept the ordering and to some degree the magnitude, while the D-vine is not quite that accurate. This, together with the log-likelihood value achieved in the previous section, speaks in favor of the canonical vine. However, for illustrative purposes, we choose to continue our analysis on both vines.

8.1.2 Expanding and Confining the Vines

In this section we check if any of the two vines can be expanded or confined, i.e. if we can replace some of the copulas or if can we assume some of the variables to be conditionally independent as discussed in Section 6.2. In order to check this, we first plot the data used in

Table 8.3: Estimated parameters and log-likelihood values for the two vines in Figure 8.1.4. The initial values were achieved by following the procedure in Section 6.5.1.

D-vine			Canonical vine		
Parameter	Initial value	Final value $\nu \in (2, 300]$	Parameter	Initial value	Final value $\nu \in (2, 300]$
ρ_{ES}	0.49	0.50	ρ_{EF}	0.86	0.86
ρ_{SF}	0.43	0.44	ρ_{ES}	0.49	0.49
ρ_{FO}	0.59	0.59	ρ_{EO}	0.60	0.60
$\rho_{EF\|S}$	0.82	0.82	$\rho_{SF\|E}$	0.01	0.01
$\rho_{SO\|F}$	0.04	0.03	$\rho_{OF\|E}$	0.20	0.20
$\rho_{EO\|SF}$	0.21	0.21	$\rho_{SO\|EF}$	-0.03	-0.03
ν_{ES}	4.38	6.51	ν_{EF}	10.33	9.62
ν_{SF}	5.76	9.17	ν_{ES}	4.38	4.30
ν_{FO}	8.16	9.24	ν_{EO}	17.88	16.87
$\nu_{EF\|S}$	9.15	10.45	$\nu_{SF\|E}$	77.30	77.25
$\nu_{SO\|F}$	19.19	17.71	$\nu_{OF\|E}$	9.04	9.18
$\nu_{EO\|SF}$	17.37	17.30	$\nu_{SO\|EF}$	15.98	16.25
$l(\hat{\mathbf{\Theta}})$	1010.27	1013.07	$l(\hat{\mathbf{\Theta}})$	1019.06	1019.24

Table 8.4: The number of degrees of freedom for each pair of Student copula fitted the simulated data from the vines with parameters as in Table 8.3

	D-vine			Canonical vine		
	O	F	S	O	F	S
E	10.73	8.97	6.31	16.16	9.35	4.33
O	-	10.10	12.23	-	11.39	9.01
F		-	9.15		-	6.06

Table 8.5: The correlation coefficient, ρ, for each pair of Student copula fitted the simulated data from the vines with parameters as in Table 8.3

	D-vine			Canonical vine		
	O	F	S	O	F	S
E	0.60	0.86	0.50	0.60	0.86	0.49
O	-	0.59	0.27	-	0.60	0.28
F		-	0.44		-	0.42

the initializing procedure in the previous section, and analyze them. Note: This is usually something we would have done before fitting Student copulas to all the pair-copulas. We did not do this because it is often normal practice to use (multivariate) Student copulas when fitting copulas to financial related data [14, 15]. We start by looking at the D-vine.

D-vine

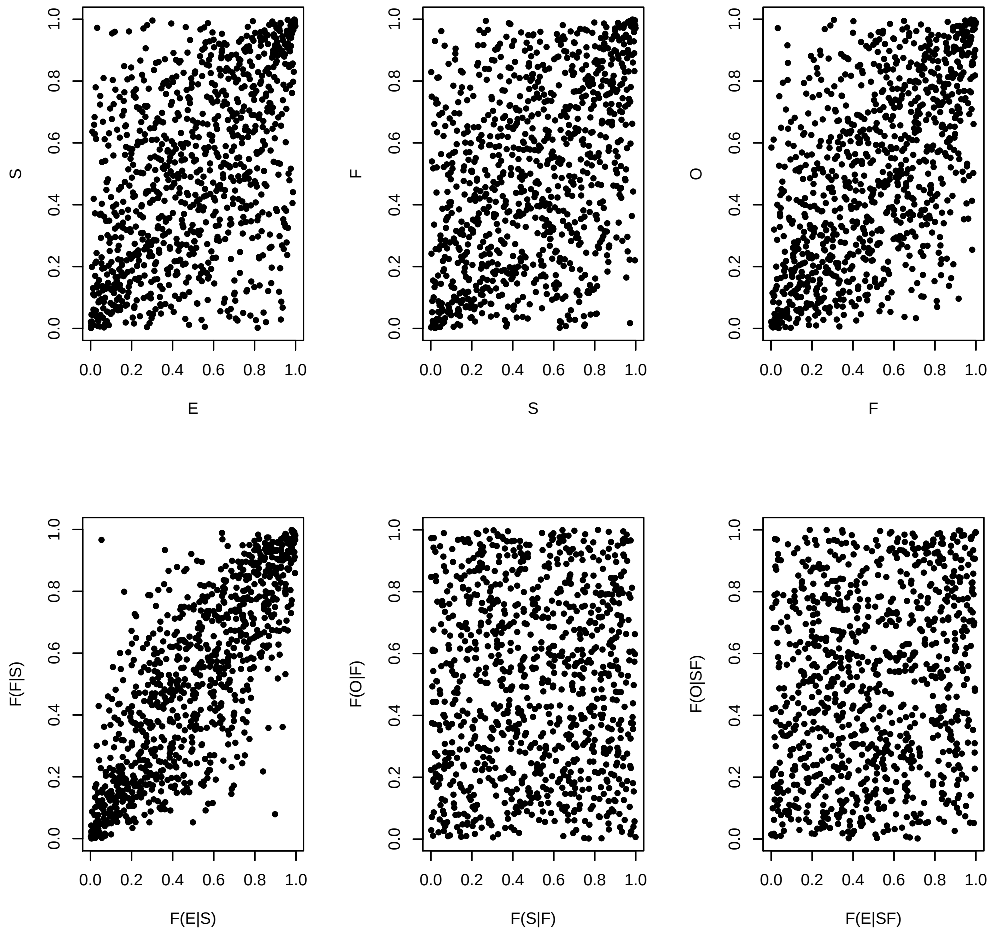


Figure 8.1.5: The data used to find the initial value for the D-vine. The data is achieved by following the procedure in Section 6.5.1.

Considering Figure 8.1.5, it seems natural to assume a Student copula for the first four scatter plots since they all clearly have both lower and upper tail dependence. The last two are more diffuse, and will be discussed closer here. In the fifth plot, i.e. $F(S|F)$ vs. $F(O|F)$, there does not seem to be a distinct pattern. Either that or we only have lower tail-dependence; there seem to be a clustering of observations in the square $[0, 0.2] \times [0, 0.18]$, while there is no obvious pattern in the upper tail. Looking at Table 3.1, this indicates a Clayton copula. In the sixth plot, λ_L and λ_U are clearly much lower than in the first four plots, if at all greater than 0. This may indicate that E and O are conditional independent of SF. The two copulas $C_{SO|F}$ and $C_{EO|SF}$ were also the two copulas with lowest value of ρ and highest value of ν in the vine, indicating low dependence. To summarize, we try the following modifications:

1. $c_{SO|F} = 1$. Result: $l(\hat{\boldsymbol{\Theta}}) = 1011.21$.
2. $c_{SO|F}$ is a Clayton copula. Result: $l(\hat{\boldsymbol{\Theta}}) = 1013.66$, with $\hat{\theta}_{Cl} = 0.08 \Rightarrow \hat{\lambda}_L = 2^{-1/\hat{\theta}} = 1.73 \cdot 10^{-4}$.
3. $c_{EO|SF} = 1$. Result: $l(\hat{\boldsymbol{\Theta}}) = 991.15$.
4. $c_{SO|F} = 1$ and $c_{EO|SF} = 1$. Result: $l(\hat{\boldsymbol{\Theta}}) = 989.40$.
5. $c_{SO|F}$ is a Clayton copula and $c_{EO|SF} = 1$. Result: $l(\hat{\boldsymbol{\Theta}}) = 991.94$.

The parameter values alter slightly when doing these modifications, but we do not give these values here since the alteration is very small. First, we compare the vines were we have set some copulas equal to the independence copula. These vines are nested within the Student copula, and their likelihood values can therefore be compared directly. It is clear that the only choice that seems somewhat reasonable is to set $c_{SO|F} = 1$. It is evidently a better choice than assuming either $c_{EO|SF} = 1$ or both $c_{SO|F}$ and $c_{EO|SF}$ to be the independence copula. This is an interesting observation when considering our approach to choose which type of vine to use in the full model in the previous section.

In Section 8.1 we discussed whether to compare ρ's or df when linking variables in tree 1. We fitted a D-vine with the pairs of variables with lowest df in the root tree, while we fitted a canonical vine with the pairs of variables with highest correlation coefficients in the root tree. Both approaches aimed to choose the three pairs of variables with largest degree of interaction. We ran into the same problem when trying to decide which variables in the vine that could be assumed independent based on the estimated values of either ρ or ν. In Table 8.6 we have recapitulated the information of the two copulas in question.

Table 8.6: Comparison of $C_{SO|F}$ and $C_{EO|SF}$ with regard to independence.

Copula	$\hat{\rho}$	$\hat{\nu}$	$l(\hat{\boldsymbol{\Theta}})_{C=\Pi}$
$C_{SO|F}$	0.03	17.71	1011.21
$C_{EO|SF}$	0.21	17.30	991.15

Both copulas have a high number of df compared to the other copulas in the vine, while $C_{SO|F}$ has a considerably lower value of ρ than $C_{EO|SF}$. When comparing the log-likelihood values, it is clear that we prefer setting $C_{SO|F} = \Pi$ rather than setting $C_{EO|SF} = \Pi$. This indicate that ρ is a better aid when deciding which variables that can be assumed independent. In the end, though, we would keep the full model instead of assuming some of the variables to be independent.

As the result in point 2 indicates, a Clayton copula might fit $C_{SO|F}$ better than a Student copula since it achieves a higher likelihood value. As mentioned in Section 7.2, we cannot compare the likelihood values directly - we need to perform tests of hypotheses. We use approach 1 discussed in Section 7.2.1 to test the hypothesis

$$\begin{aligned} \mathrm{H}_0 &: C_{SO|F} \text{ is a Clayton copula with } \theta = 0.08 \\ \text{vs. } \mathrm{H}_1 &: C_{SO|F} \text{ is } \textit{not} \text{ a Clayton copula with } \theta = 0.08, \end{aligned} \tag{8.1.1}$$

and

$$\begin{aligned} &\mathrm{H}_0^*: C_{SO|F} \text{ is a Student copula with } \rho = 0.03 \text{ and } \nu = 17.71 \\ \text{vs. } &\mathrm{H}_1^*: C_{SO|F} \text{ is } \textit{not} \text{ a Student copula with } \rho = 0.03 \text{ and } \nu = 17.71. \end{aligned} \tag{8.1.2}$$

In addition we also use approach 2 to test H_0. The p-values of the tests are summarized in Table 8.7.

Table 8.7: The estimated p-values of H_0 and H_0^*. We use $\alpha = 0.05$ as level of significance.

Null hypothesis	Approach	K	N_b	$\hat{p}$	Conclusion
H_0	1	10000	-	0.65	Keep H_0
H_0	2	5000	-	0.09	Keep H_0
H_0^*	1	10000	5000	0.63	Keep H_0^*

As is seen from the estimated p-values in Table 8.7, Neither H_0 nor H_0^* is rejected at a 5% level. One could plot the degree of closeness as done in [2] to investigate the two alternatives further, but since the canonical vine in this case seem to outperform the D-vine regardless of the choice of $C_{SO|F}$, we do not pursue the matter any further.

Canonical Vine

Analyzing Figure 8.1.6, plot 1, 2 and 3 coincide quite good with student copulas. The fifth plot also coincide somewhat with a Student copula. The fourth and sixth plots however, seem much more unstructured, and it seems natural to assume that both $c_{SF|E} = 1$ and $c_{SO|FE} = 1$. Again, we try all combinations of the proposed improvements.

1. $c_{SF|E} = 1$. Result: $l(\hat{\mathbf{\Theta}}) = 1019.06$.
2. $c_{SO|FE} = 1$. Result: $l(\hat{\mathbf{\Theta}}) = 1017.03$.
3. $c_{SF|E} = 1$ and $c_{SO|FE} = 1$. Result: $l(\hat{\mathbf{\Theta}}) = 1016.89$.

The conclusions are much the same as with the D-vine; some of the variables could have been assumed conditional independent, but the full model still achieves a slightly higher log-likelihood value. Hence, we keep the full model.

8.1.3 Final Model and Comparison with a Four-Dimensional Student Copula

As discussed in Section 8.1.1, the canonical vine seems to maintain the dependency structure among all four variables in a better way than the D-vine, especially when considering the degrees of freedom. In addition, the canonical vine achieved a higher log-likelihood value than the D-vine regardless of the choice of copula for $C_{SO|F}$[2]. With this in mind, we choose the canonical vine as our final model. As mentioned earlier, common practice when fitting models to financial returns has been two fit multivariate Student copulas to the data. As a final evaluation of our model, we compare it to a four-dimensional Student copula performing a likelihood ratio test. The estimated parameters of the four-dimensional

[2]Note: We cannot compare the likelihood values for the canonical vine and the D-vine with a Clayton copula directly.

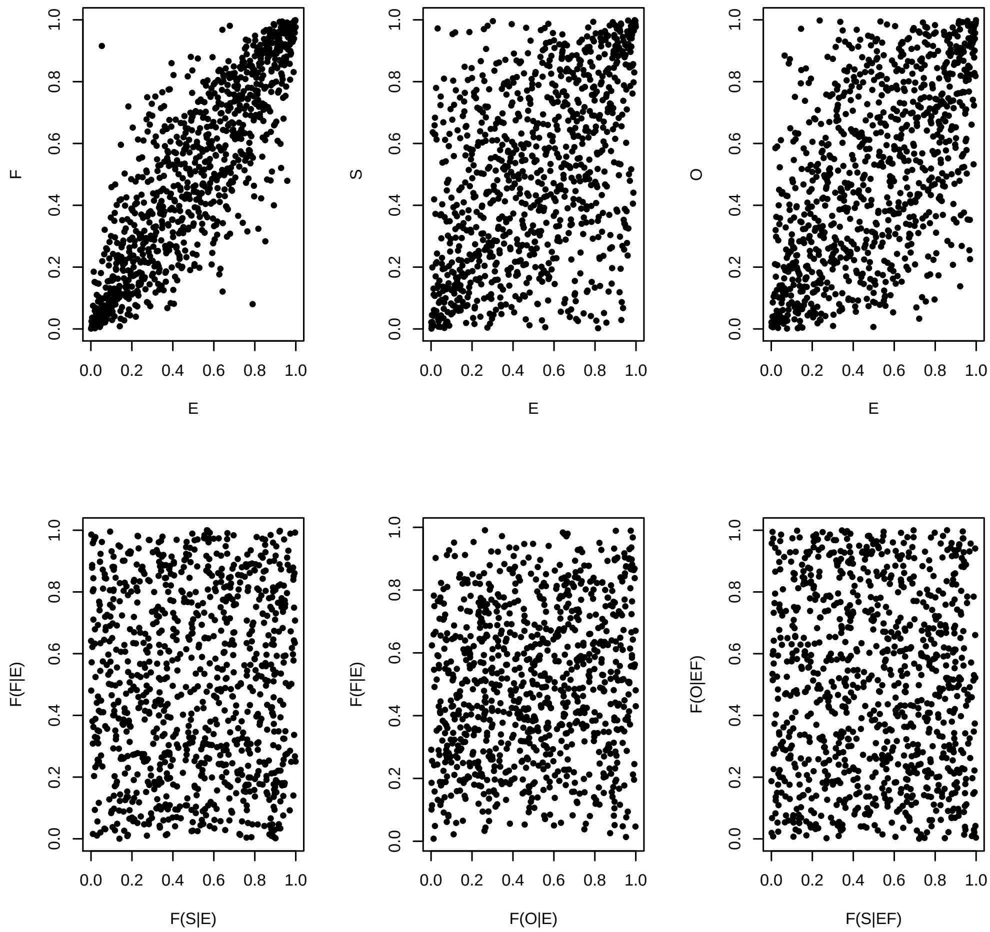


Figure 8.1.6: The data used to fit the initial value fro the canonical vine. The data is achieved by following the procedure in Section 6.5.1.

Student copula are summarized in Table 8.8. We notice that the values of ρ_{ij} coincide with the values in Table 8.2.

Table 8.8: Estimated parameters for the four-dimensional Student copula fitted the data.

ρ_{EO}	ρ_{EF}	ρ_{ES}	ρ_{OF}	ρ_{OS}	ρ_{FS}	ν	$l(\hat{\boldsymbol{\Theta}})$
0.59	0.86	0.50	0.59	0.29	0.44	9.50	1010.65

The likelihood ratio statistics is $2 \cdot (1019.24 - 1010.65) = 17.18$ with $12 - 7 = 5$ degrees of freedom. This gives a p-value of 0.0042. Hence, we can reject the four-dimensional Student copula in favor of the canonical vine.

8.2 Fitting Bivariate Copulas to Intra Day Financial Returns

In the previous section we investigated daily quotes and fitted a model to these. In this section we look at intra day data, that is we look at prices at which a stock has been realized (bought and sold) for during a market day. We look at the following three stocks.

- Citibank, a major international bank, founded in 1812.
- International Business Machines Corporation (IBM), a multinational computer technology and IT consulting corporation.
- General Motors (GM), a global automaker founded in 1908 with headquarters in Detroit, Michigan. It was one of the worlds largest automakers.

We look at the sixteen market days from Wednesday 01.09.99 to Friday 23.09.99 (Monday 06.09.99 was labor day), and we extract the prices every fifth minute from 10.00 to 16.00. These prices are plotted in Figure 8.2.1. The market opens at 09.00 and closes at 17.00, but big buyers tend to manipulate the prices immediately after opening or just before closing, so we omit these periods of the day. This results in 73 quotes each day. However, we analyze the log-returns, and we do not include the log-returns that goes over two days, e.g. $\log\left(r1_{Tuesday}/r73_{Monday}\right)$, since this also would bring in the effect of the possible large jumps at the start or end of each market day. This effect is easily seen at the dotted red lines in Figure 8.2.1. This leaves us with 72 returns on each day, a total of 1152 log-returns for each stock over all sixteen days.

Our main purpose in this section is to investigate how a model on daily data develop over a short time period. We do this by fitting models to the data in the way described in Figure 8.2.2.

Hence, we fit nine models, named Model 1, Model 2, ... , Model 9, each based on 576 returns in which 504 is used in the previous model and 504 in the next model for all models except Model 1 and Model 9.

8.2.1 Complications with Intra Day Data

There are mainly two drawbacks when analyzing intra day prices contra daily prices. Firstly, the fact that stocks are purchased and sold at fixed prices during the day, i.e. prices can only rise or fall in multiples of the tick size (1/16$) [16], cause a problem when we try to fit a GARCH-model on the log-returns. This grid-effect is seen in the scatter plots of the log-returns in Figure 8.2.3. The GARCH-framework is meant for continuous data, hence the analysis with the discrete data becomes very poor. This effect is seen in intra day data and not in the daily quotes used in the previous section due to the small time intervals at which we collect the prices in this section. We try to imitate a continuous distribution for the log-returns by adding noise following (8.2.1).

$$lr_i = lr_i + \xi_i, \quad \xi_i \sim N(0, \sigma^2), \quad \sigma^2 = 4.50 \cdot 10^{-7}, \tag{8.2.1}$$

where lr is the vector of the log-returns for variable i. The new log-returns are plotted in Figure 8.2.4, and we can see that the data no longer appear as discrete at the same time as the structure is somewhat kept.

The second factor that affect our analysis is the size of the fluctuations we have during a market day. This is a feature that is very stock-specific. Not all stocks are exchanged

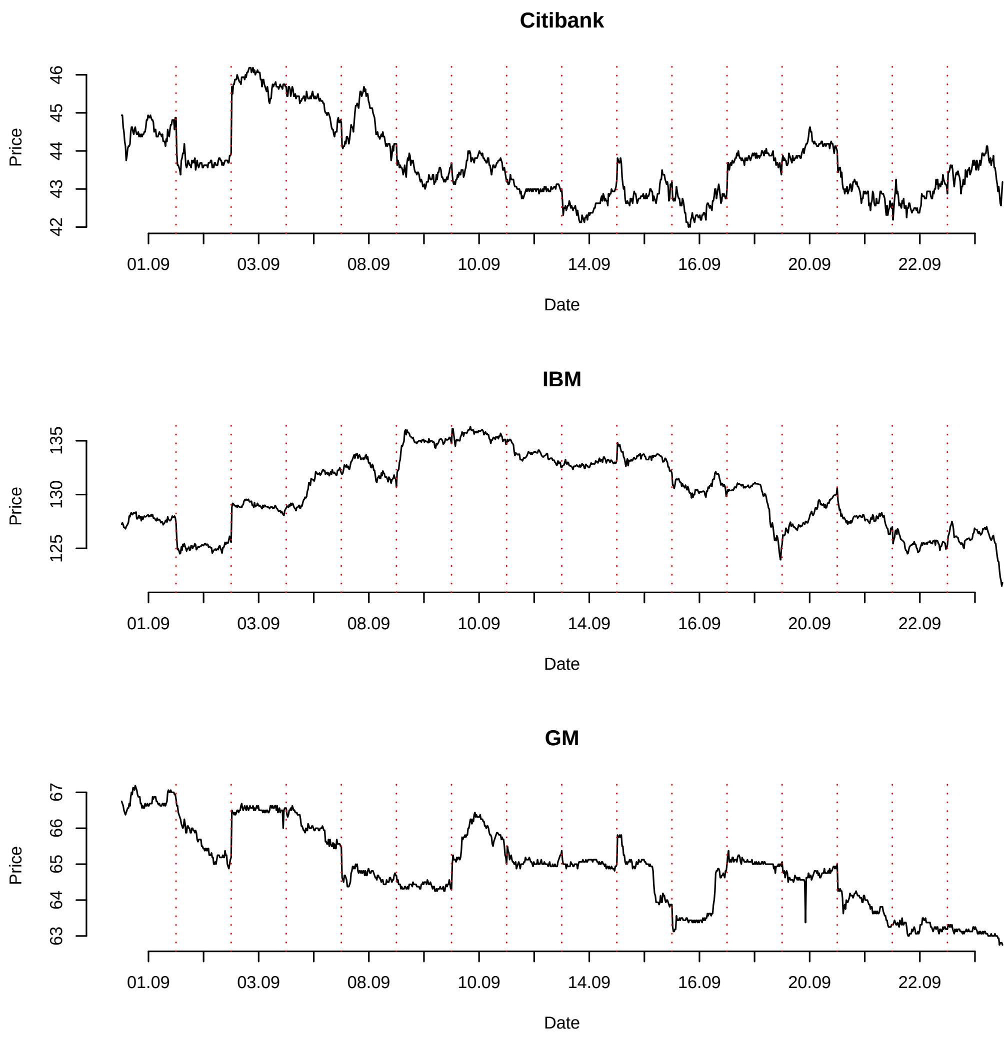


Figure 8.2.1: Prices for which Citibank, IBM and GM were traded every fifth minute from 10.00 to 16.00 from 01.09.99 to 23.09.99. The days on the x-axis are marked in the middle of the day, and two days are separated with a dotted red line.

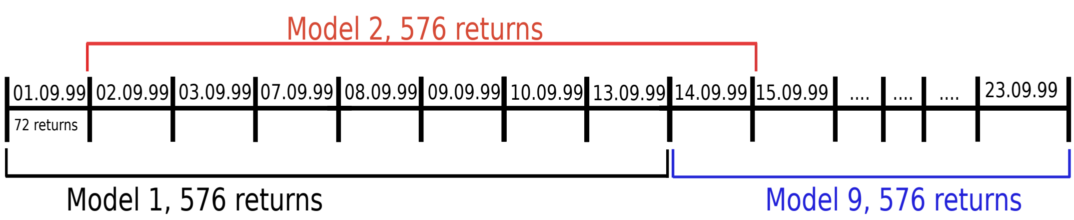
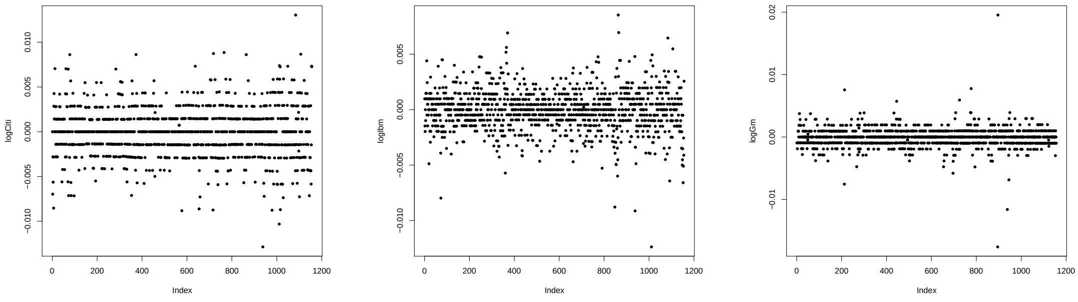


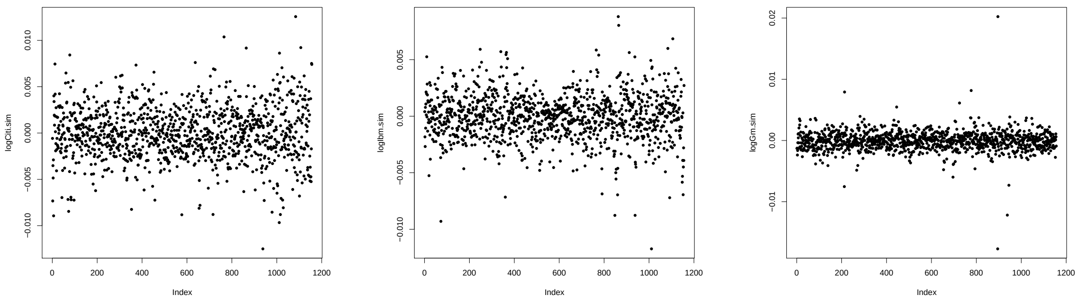
Figure 8.2.2: A graphic representation of the 9 models we fit the data.



(a) (b) (c)

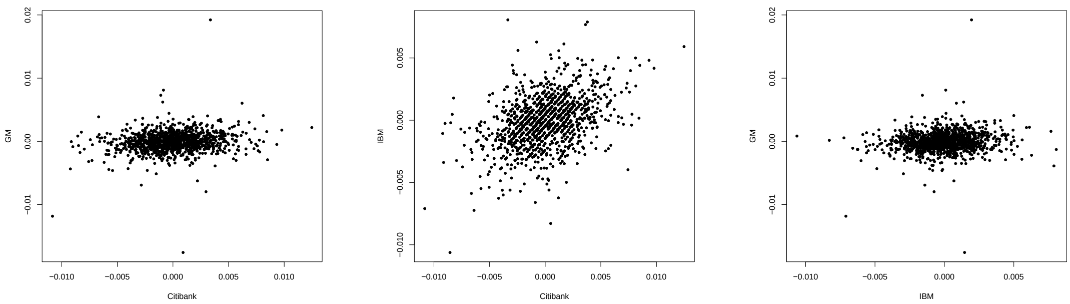
Figure 8.2.3: The log-returns. (a): Citibank (b): IBM (c): GM.

equally frequent, or vary as much in price, and this may lead to poor intra day models. This is seen in the pairwise scatter plot of the log-returns in Figure 8.2.5 and in Table 8.9 where we have summarized the activity on September 1st. GM clearly does not vary much relative to Citibank and IBM. This is an inevitable problem which we have to accept. Of course we could have chosen more liquid stocks to avoid this problem, but by analyzing the stocks in question, we can see how this will effect the final model.



(a) (b) (c)

Figure 8.2.4: The log-returns with added noise. (a): Citibank (b): IBM (c): GM.



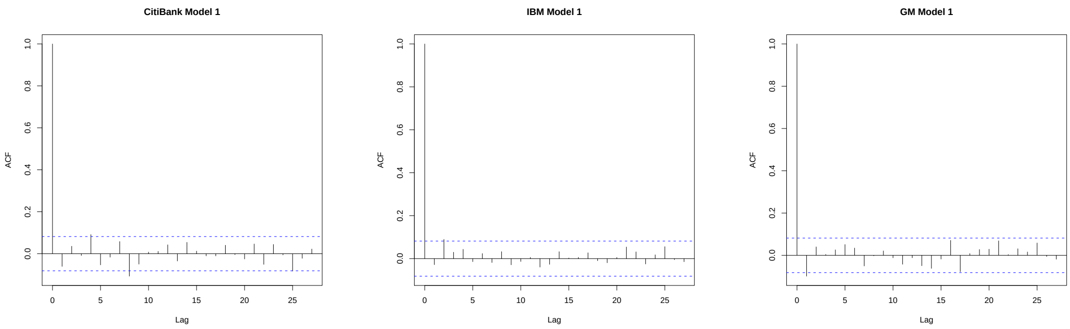
(a) (b) (c)

Figure 8.2.5: Scatter plot of pairwise log-returns. (a): Citibank vs. GM (b): Citibank vs. IBM (c): IBM vs. GM.

In Figure 8.2.6 we see that there is little or no autocorrelation left in the standardized residuals from the GARCH(1,1)-model fitted the log-returns with added noise[3].

Table 8.9: Statistics for the three stocks on September 1st 1999. Note: We have rounded prices off till two decimals even though they are traded in multiples of (1/16$).

Stock	# Trades	min(price)	max(price)	mean(price)	sd(price)
Citibank	3785	43.75	45.13	44.53	0.28
IBM	4744	125.50	128.50	127.53	0.65
GM	1296	65.75	67.38	66.80	0.20



(a) (b) (c)

Figure 8.2.6: The autocorrelation function for the three stocks for Model 1. (a): Citibank (b): IBM (c): GM.

8.2.2 Fitting Vines

Scatter plots of the empirical distribution function for the residuals of the GARCH(1,1) models fitted the log-returns are shown in Figure 8.2.7 for Model 1 and Model 9. These two models do not have any overlapping data. It is clear from these plots that there is some day-to-day-variation, but also that the "big picture" is kept over the sixteen days. We assume all copulas to be Student copulas. This assumption is supported by the initial data for Model 1, shown in Figure 8.2.8.

We tried all $(n!/2) = (3!/2) = 3$ possible vines[4], and achieved highest log-likelihood values with IBM as the kernel variable in tree 1 as illustrated in Figure 8.2.9. The final values are summarized in Table 8.10, and the development of them are visualized in Figure 8.2.10, 8.2.11 and 8.2.12.

There is no obvious joint pattern in the development of the ρ's seen in Figure 8.2.10. We see that ρ_{CI} shows an increasing trend in the latter models, while ρ_{IG} and $\rho_{CG|I}$ show slightly decreasing trends. Looking at the df in Figure 8.2.11 however, all three variables fluctuate in a more similar pattern in the latter models. Both these observations are

[3]Whenever we talk about the log-returns later in this section, we are referring to the log-returns with added noise unless otherwise stated.

[4]Remember that for three variables, D-vines can be represented as canonical vines and vice versa, hence we have we only have (3!/2) possible vines.

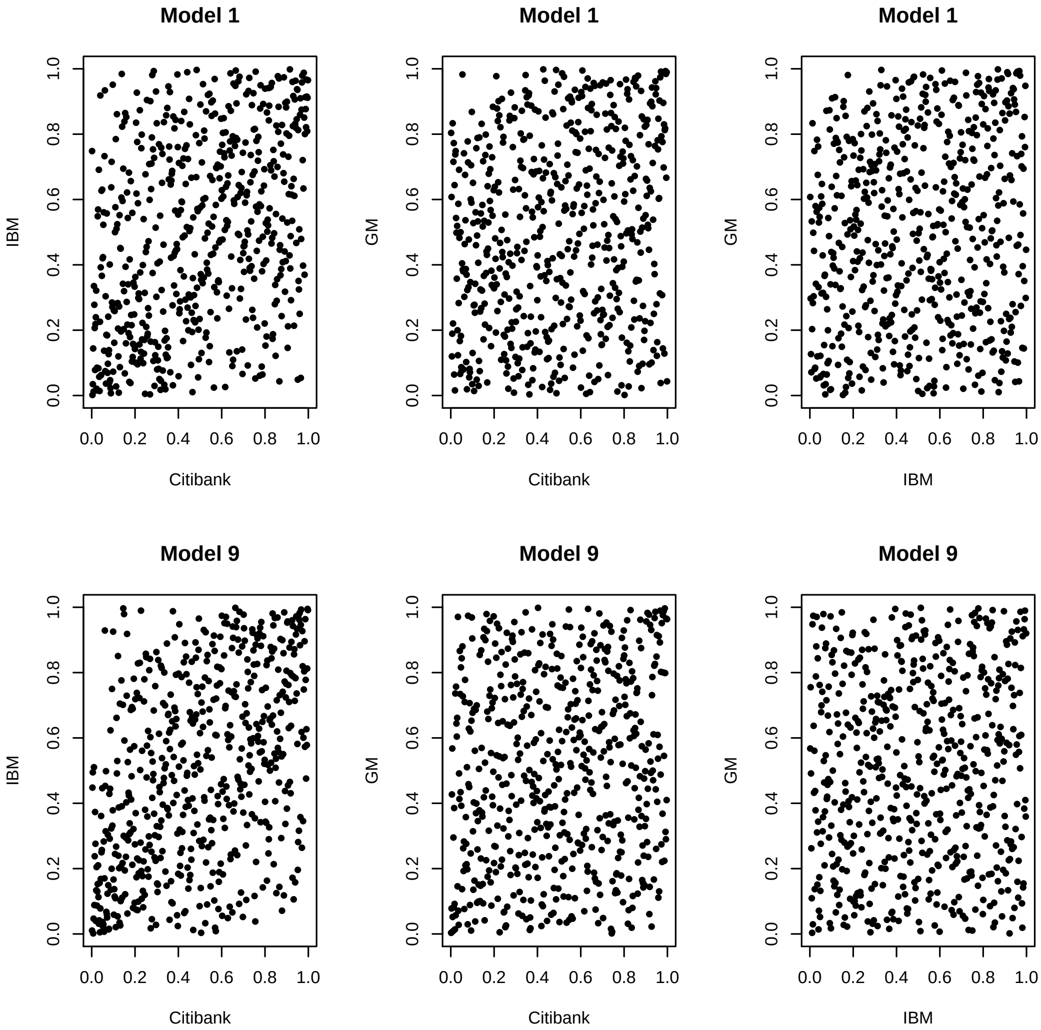
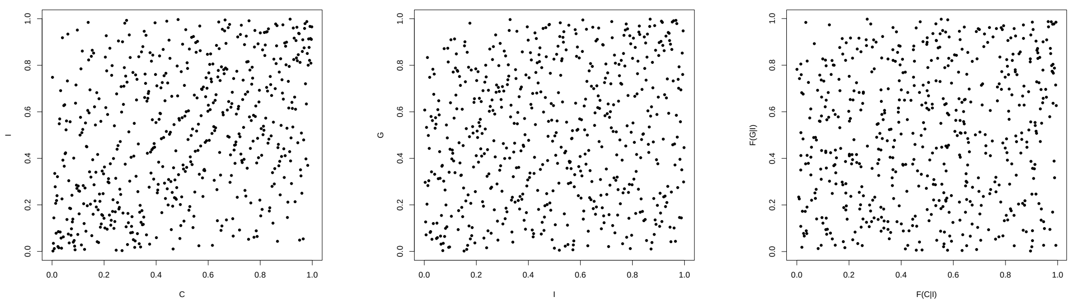


Figure 8.2.7: Scatter plots of the empirical distribution function for the residuals of the GARCH(1,1) models fitted the log-returns for Model 1 and Model 9.



(a) (b) (c)

Figure 8.2.8: The initial data for Model 1 when assumed Student copulas for all copulas. (a): C vs. I (b): I vs. G (c): $F(C|I)$ vs. $F(G|I)$, $C =$ Citibank, $I =$ IBM, $G =$ GM.

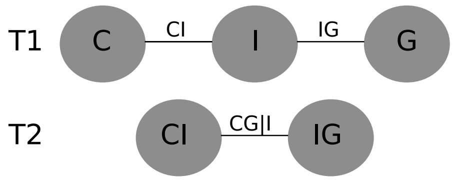


Figure 8.2.9: The vine used in this model. Note: I is here a kernel node in a canonical vine even though we have illustrated it as a D-vine.

Table 8.10: Final parameters and log-likelihood values for the nine models fitted the data. Initial values are not accounted for here.

	M1	M2	M3	M4	M5	M6	M7	M8	M9
ρ_{CI}	0.39	0.38	0.35	0.38	0.39	0.39	0.38	0.40	0.45
ρ_{IG}	0.17	0.15	0.15	0.14	0.12	0.06	0.07	0.06	0.08
$\rho_{CG\|I}$	0.14	0.11	0.12	0.13	0.18	0.14	0.16	0.15	0.15
ν_{CI}	13.61	20.85	27.10	36.75	20.63	17.35	11.03	14.44	11.55
ν_{IG}	300.00	300.00	300.00	74.27	137.84	26.27	13.61	13.44	9.86
$\nu_{CG\|I}$	50.95	300.00	300.00	300.00	37.86	26.50	14.30	18.83	19.50
$l(\hat{\mathbf{\Theta}})$	59.70	52.83	45.96	54.49	61.91	54.36	59.17	60.94	75.39

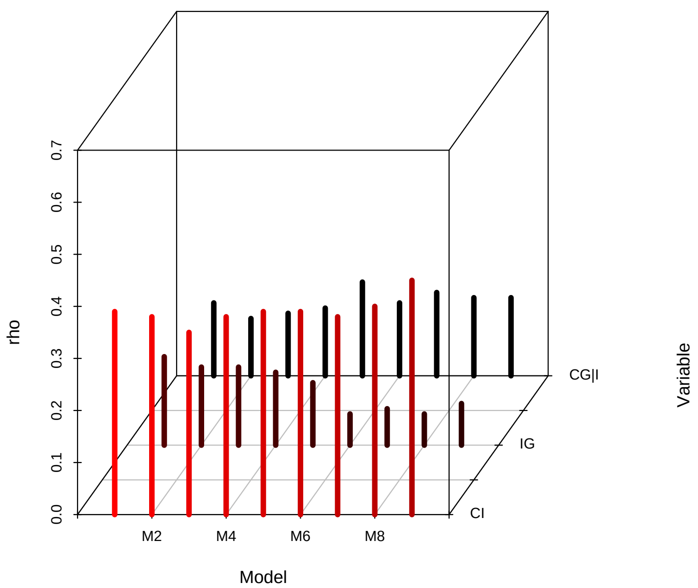


Figure 8.2.10: The development for the different correlations for all nine models.

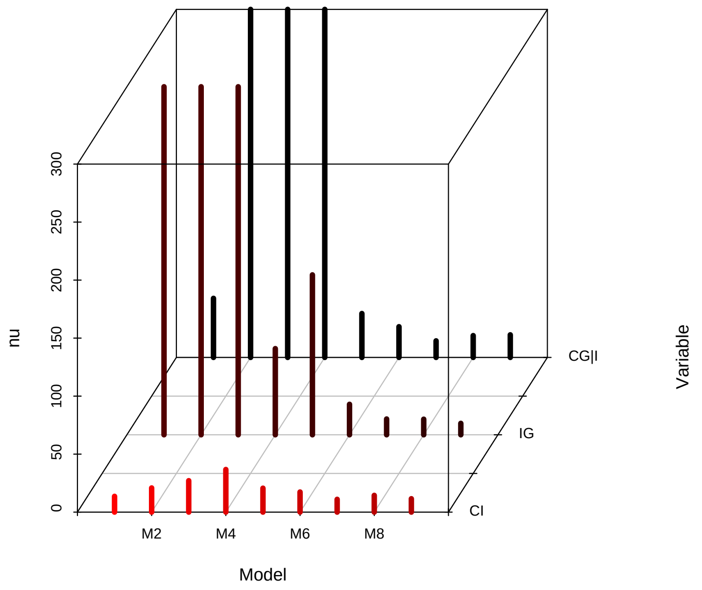


Figure 8.2.11: The development for the different df for all nine models.

verified in Table 8.11, which shows the changes in percent between neighboring models. The lack of internal interaction between ρ and df for each of the three copulas is seen in Figure 8.2.12.

If we only consider the correlation coefficients, the most dramatic change in the parameters occur when shifting from Model 5 (period: 08.09.99-17.09.99) to Model 6 (period: 09.09.99-20.09.99). If we look at the prices in Figure 8.2.1, we see that on the two days that separates these models, the eighth and twentieth September, the intra day movements are quite different for the three stocks. On the eighth, Citibank has a large peak in the price in the middle of the day, IBM slightly decreases in price during the whole day and GM has only minor fluctuations. On the twentieth we see some of the same characteristics with Citibank, while IBM, contrary to the eighth, increases during the whole day and GM has a downward peak in the middle of the day. These changes causes ρ_{IG} to fall with 50% and $\rho_{CG|I}$ with 22%. Similar characteristics in the movement of the prices can be found to explain the other jumps in the parameter values.

Again, as when fitting Student copulas to daily quotes, the df is much more unsteady when optimizing the vine with respect to the log-likelihood value.

8.2.3 Comparison with a Three-Dimensional Student Copula

As we did in Section 8.1.3, we compare our pairwise decomposition with a multivariate Student copula. The parameter values with corresponding log-likelihood values are summarized in Table 8.12. We have also calculated the likelihood ratio statistics and contrary to the model on daily quotes, we fail to reject the multivariate Student copulas in all nine models.

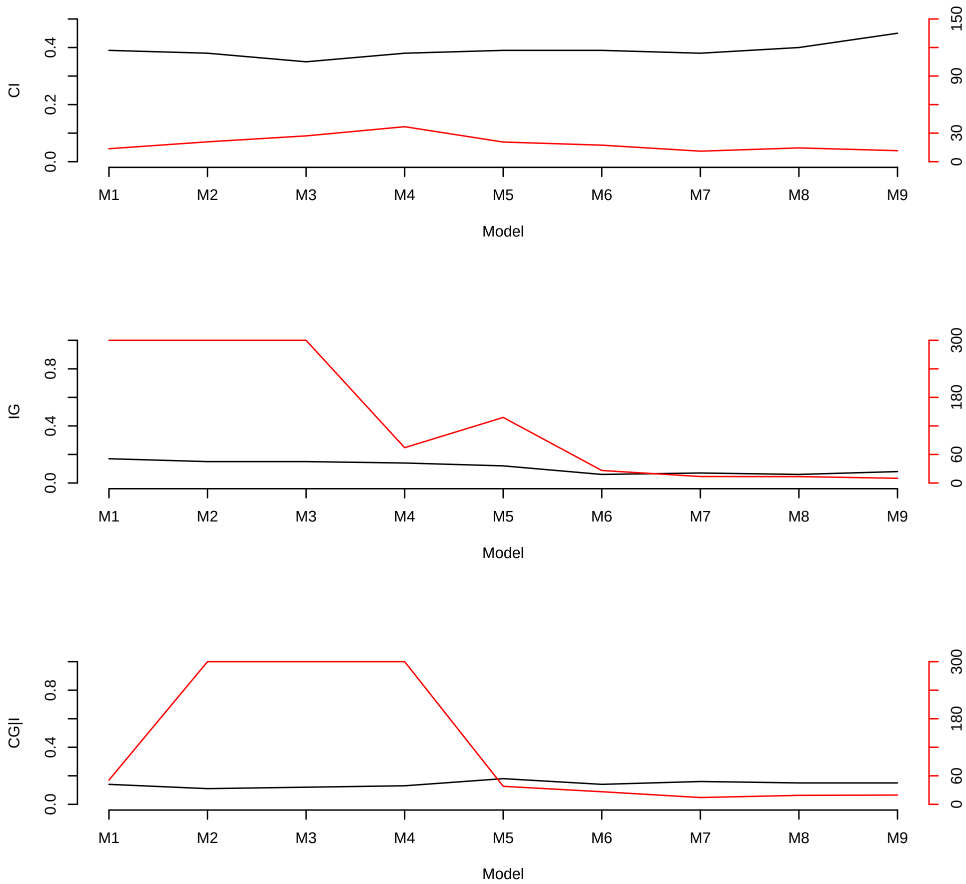


Figure 8.2.12: The development for the different correlations (black lines) and degrees of freedom (red) for all nine models.

Table 8.11: Changes in the variables between the different models in percentage.

Variable	M1/M2	M2/M3	M3/M4	M4/M5	M5/M6	M6/M7	M7/M8	M8/M9
ρ_{CI}	-3	-8	9	3	0	-3	5	12
ρ_{IG}	-12	0	-7	-14	-50	17	-14	33
$\rho_{CG\|I}$	-21	9	8	38	-22	14	-6	0
ν_{CI}	53	30	36	-44	-16	-36	31	-20
ν_{IG}	0	0	-75	86	-81	-48	-1	-27
$\nu_{CG\|I}$	489	0	0	-87	-30	-46	32	4

Table 8.12: Estimated parameters using ML for three-dimensional Student copulas fitted the nine models. LR = likelihood ratio statistic. The test of hypothesis is performed with $\nu = 6 - 4 = 2$ degrees of freedom.

	M1	M2	M3	M4	M5	M6	M7	M8	M9
ρ_{CI}	0.39	0.38	0.35	0.38	0.39	0.39	0.38	0.40	0.44
ρ_{CG}	0.19	0.15	0.16	0.17	0.21	0.15	0.17	0.16	0.16
ρ_{IG}	0.16	0.15	0.15	0.14	0.12	0.06	0.07	0.06	0.07
ν	35.15	79.94	79.32	55.95	33.85	22.04	12.47	14.70	11.70
$l(\hat{\mathbf{\Theta}})$	58.93	52.35	45.58	54.30	61.74	54.31	59.13	60.92	75.23
LR	1.54	0.96	0.76	0.38	0.34	0.10	0.08	0.04	0.32
p	0.46	0.62	0.68	0.83	0.84	0.95	0.96	0.98	0.85

8.2.4 Altering the Model

Since the vine in Section 8.2.2 did not lead the rejection of the three-dimensional Student copula, we decided to alter the data set in different ways to see if we could find a decomposed model that would lead to the rejection of a multivariate Student copula. We tried the following two changes.

1. We added noise with greater variance to the log-returns, i.e. we changed σ in (8.2.1). We used $\sigma = 4.5 \cdot 10^{-6}$.

2. We expanded the data set with an additional stock, and examined the impact this had in the likelihood ratio tests. We added Apple[5] into our data set, and used $\sigma = 4.5 \cdot 10^{-6}$ in (8.2.1).

We chose to add another stock as we suspected the poor fit in Section 8.2.2 to be related to the degrees of freedom. That is, in the vine, we had six parameters; Three correlation coefficients and three df-parameters. In the three-dimensional Student copula we had four variables; three correlation coefficients and one df-parameter. Our theory was that there is so little information in the degrees of freedom that the additional two df-parameters in the three-dimensional vine does not make a large enough impact to reject the three-dimensional copula in favor of the decomposed model.

By expanding the model to contain four variables, we can see the impact the additional df-parameters have. The results are summarized in Table 8.13 and 8.14 (parameter values are not accounted for).

By analyzing the estimated p-values in both changes, the result is clear: We cannot reject the multivariate Student copulas in favor of the decomposed models. More discussion about the result follows in Chapter 9.

[5]Apple is an American multinational corporation which designs and manufactures consumer electronics and software products.

Table 8.13: The effect of greater noise on the residuals. Log-likelihood values for the vine in Figure 8.2.9 with $\sigma = 4.5 \cdot 10^{-6}$ in (8.2.1) and a three-dimensional Student copula on the same data. The p-value is calculated for a likelihood ratio test with $\nu = 6 - 4 = 2$ degrees of freedom.

Model	M1	M2	M3	M4	M5	M6	M7	M8	M9
Vine	147.88	167.85	153.69	153.08	156.70	158.04	165.80	162.23	166.96
Copula	145.79	166.57	153.20	152.79	155.13	157.21	164.84	161.11	165.63
p-value	0.12	0.28	0.61	0.75	0.21	0.44	0.38	0.33	0.26

Table 8.14: The effect of adding Apple to the data. Log-likelihood values for a D-vine with $G \leftrightarrow C \leftrightarrow I \leftrightarrow A$ in tree 1 with $\sigma = 4.5 \cdot 10^{-6}$ in (8.2.1) and a four-dimensional Student copula on the same data. The p-value is calculated for a likelihood ratio test with $\nu = 12 - 7 = 5$ degrees of freedom.

Model	M1	M2	M3	M4	M5	M6	M7	M8	M9
Vine	68.30	50.39	46.25	61.07	64.37	50.11	66.34	75.99	96.23
Copula	67.10	49.69	46.07	60.77	63.61	49.14	65.38	74.96	95.34
p-value	0.79	0.92	1.00	0.99	0.91	0.86	0.86	0.84	0.88

Chapter 9

Evaluation of the Numerical Experiments

In this chapter we discuss the results obtained in Section 8.1 and 8.2. Section 9.1 is devoted to discussion and conclusions, while we in Section 9.2 propose further work that is possible to perform with the material covered in this thesis as a basis. If possible, we recommend direct improvements or changes in some of the procedures performed in this thesis.

9.1 Discussion and Conclusions

We divide this section into two parts; the first part is regarding both of the models in Chapter 8, and the other is about discoveries regarding only the model on intra day data.

9.1.1 Both Models

In this thesis we have performed two main experiments. We have fitted one model on daily quotes and one model on intra day transaction prices. Models regarding daily quotes are starting to get well explored in the literature. One of our main sources of inspiration, the work of Aas et al. (2007) [2], has also tested the theory on daily quotes. The conclusion in both our experiments and in Aas et al. (2007) is evident. On daily quotes, a decomposed model consisting of bivariate copulas only is clearly preferable to multivariate Student copulas with a level of significance less than 0.01. However, there are still interesting observations to discuss.

One interesting aspect regarding the two vines we fitted in Section 8.1 is how good the initializing procedure described in Section 6.5.1 performs; Using the initial parameter values in the four-dimensional canonical vine in Section 8.1 gives a p-value of 0.0049 for the same likelihood ratio test performed in Section 8.1.3. This might indicate that the optimization of the log-likelihood is redundant. However, if we on the contrary were to use the D-vine with only Student copulas rather than the canonical vine, we would get p-values of 1.00 and 0.79 using the initial and final parameters respectively. In both cases the four-dimensional Student copula would be preferred, but that is not the point. The point is that there is a much greater difference in the two p-values than in the case of the canonical vine. One cannot know in advance *how* good the initial parameter estimates are. Our conclusion is that even though the initial estimates are “good”, one still has to perform the optimization procedure.

In Section 6.5.2 we discussed the introduction of a penalizing function for the degrees of freedom when optimizing the likelihood with respect to $\mathbf{\Theta}$. Even though we discarded our attempt to penalize ν's larger than 15, we have experienced some difficulties when optimizing the likelihood. This became evident in the analysis of the intra day data where the df varied greatly even in neighboring models. More on this topic in the next section.

One other topic discussed in Section 8.1 was which measure to use in the process of deciding the structure in tree 1 of vines consisting of Student copulas. In Section 8.1.2 we spoke in favor of using tables regarding ρ, since this resulted in the canonical vine that we later found to be superior compared with the D-vine. However, we should be careful to only use the ρ's in this procedure since they are merely a linear correlation coefficient. Degrees of freedom tells us something about the dependency in the tails, unquestionably a very important feature when discussing copulas. It may also be taken into consideration that for models with low dimensions one can fit all possible vines, an choose the one with highest likelihood. However, with five or more variables, this becomes extremely costly.

9.1.2 Intra Day Data

We were not capable of making a decomposed model that outperformed a multivariate vine in Section 8.2. We tried with both three- and four-dimensional models, with different variances in the noise added to the log-returns. Neither had the desired effect. There are several factors that can have lead to the rejection of our model. The size of the data set can be one of them. While we on the daily quotes used 934 data points, we only used 576 for each model on the intra day data. We did not test other model sizes, so we do not know whether this could have lead to other conclusions.

Another complication with the intra day model was of course the data itself. As discussed in Section 8.2.1, the fact that the shares rise and fall in discrete steps leads to poor fits using the GARCH-framework. There might be other, perhaps better, ways to adjust the data than we did. We discuss this further in the next section. The intra day models also seemed extra sensitive to changes in the degrees of freedom when optimizing the likelihood with respect to $\mathbf{\Theta}$. This can either be linked with the data itself, or perhaps to outliers such as the great fall we had in Model 6 on GM as discussed in Section 8.2.2.

One of the most interesting findings analyzing the intra day data (even though we rejected our model) is the fact that we must use dynamic models. It is evident when analyzing the parameter in both the vine and the three-dimensional Student copula that none of the parameters stays constant in time. Even in neighboring models the changes can be dramatic. This is clearly a subject of interest for further development.

9.2 Improvements and Further Work

Due to time limitations we have not been able to investigate all of the obtained results as thorough as desired. In this section we list some of the subjects that we think should be further investigated. Most of them is regarding the decomposed models on intra day data since this, to our knowledge, is not well documented in the literature.

Direct Changes of the Intra Day Data Set

As mentioned in the previous section, larger dynamic models would be preferable. That is, larger data sets should be used. In addition, experiments with changes in the number of days in each model and number of variables should be performed. There can also be performed alterations regarding the time intervals during the day. We have used 5 minute intervals, but this can easily be altered. The interval has to be large enough so that there is sufficiently trading in each interval to provide useful information, but at the same time small enough to capture a short-term relationship.

We have only considered the transactions between 10.00 and 16.00 even though the marked is open between 09.00 and 17.00. Although this is usual to do when regarding intra day models, it could be interesting to look into models where all data is accounted for. One can for instance divide each day into three sections; morning, noon and afternoon and have three models for each day. Two models might be enough, if we link one days' afternoon with the next days' morning transactions.

Other Adjustments and Additional Work

Instead of adding noise to the log-returns as we did, it would be interesting to fit a discrete GARCH-model to the data as discussed in Amilon (2003) [17]. Note: This may be a dead end, but is worth looking into.

We also think that it is worth looking closer into the process of estimating the degrees of freedom. That is, look at other possible ways to restrict the degrees of freedom or in some way remove the instability it brings into the optimization procedure.

Finally the models should be used in further analysis such as for example calculating Value at Risk discussed in [5]. There are many articles and books available discussing the use of copulas in financial theory, and it would be interesting to compare the decomposed models suggested here with the usual multivariate models on such theory as well. See for example Cherubini, Luciano and Vecchiato's *Copula Methods in Finance* [18].

Bibliography

[1] Assaf Zeevi and Roy Mashal. The generalized hyperbolic skew studentâs t-distribution, 2002.

[2] Kjersti Aas, Claudio Czado, Arnoldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependences. *Insurance: Mathematics and Economics (2007)*, 2007.

[3] Daniel Berg. Copula goodness-of-fit testing: An overview and power comparison, 2009.

[4] Tim Bedford and Roger M. Cooke. Vines - a new graphical model for dependent random variables, 2002.

[5] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantative Risk Management*. Princeton, 2005.

[6] Roger B. Nelsen. *An Introduction to Copulas*. Springer, 1999.

[7] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye. *Probability and Statistics for Engineers and Scientists*. Prentice Hall, 2002.

[8] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: Properties and pitfalls, 1999.

[9] Paul Embrechts, Filip Linskog, and Alexander McNeil. Modelling dependence with copulas and applications to risk management, 2001.

[10] H. Joe. Familie of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters, 1996.

[11] David M. Gay. Usage summary for selected optimization routines, 1990.

[12] Daniel Berg. Copula goodness-of-fit testing: An overview and power comparison, 2007.

[13] `http://www.stoxx.com/index.html`.

[14] Claudia Czado, Florian Gärtner, and Aleksey Min. Analysis of australian electricity loads using joint bayesian inference of d-vines with autoregressive margins, 2008.

[15] Claudia Czado, Aleksey Min, Tanja Baumann, and Rada Dakovic. Pair-copula constructions for modeling exchange rate dependence, 2008.

[16] `http://www.nyse.com/`.

[17] Henrik Amilon. Garch estimation and discrete stock prices: an application to low-priced australian stocks, 2003.

[18] Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula Methods in Finance*. John Wiley & Sons, Ltd, 2004.

[19] H. Joe. Gary g. venter, guy carpenter instrat, 2001.

[20] Craig Blocker, John Conway, Luc Demortier, Joel Heinrich, Tom Junk, Louis Lyons, and Giovanni Punzi. Simple facts about p -values, 2006.

Appendix A

Classes and Families of Copulas

A.1 Archimedean Copulas

There are many different types of copulas that fall into the Archimedean class, see [6] et al. for a thorough discussion. Three well known members are the Gumbel copula, the Clayton copula and the Frank copula. The formulas for the densities and h-functions are collected from [2] and [19].

Gumbel:

$$\begin{aligned}
C_\theta^{\mathrm{Gu}}(u,v) &= \exp\Big(-\big((-\log u)^\theta + (-\log v)^\theta\big)^{1/\theta}\Big), \quad 1 \le \theta < \infty,\\
c_\theta^{\mathrm{Gu}}(u,v) &= C_\theta^{\mathrm{Gu}}(u,v)\cdot(uv)^{-1}\cdot\{(-\log u)^\theta + (-\log v)^\theta\}^{-2+2/\theta}\\
&\quad \times (\log u \log v)^{\theta-1}\cdot\{1+(\theta-1)\cdot((-\log u)^\theta + (-\log v)^\theta)^{-1/\theta}\},\\
h_{\mathrm{Gu}}(u,v,\theta) &= C_\theta^{\mathrm{Gu}}(u,v)\cdot\frac{1}{v}\cdot(-\log v)^{\theta-1}\cdot\{(-\log u)^\theta + (-\log v)^\theta\}^{1/\theta-1},\\
h_{\mathrm{Gu}}^{-1}(u,v,\theta) &\text{ must be obtained numerically.}
\end{aligned}$$

Clayton:

$$\begin{aligned}
C_\theta^{\mathrm{Cl}}(u,v) &= (u^{-\theta}+v^{-\theta}-1)^{-1/\theta}, \quad 0<\theta<\infty,\\
c_\theta^{\mathrm{Cl}}(u,v) &= (1+\theta)\cdot(uv)^{-1-\theta}\cdot(u^{-\theta}+v^{-\theta}-1)^{-1/\theta-2},\\
h_{\mathrm{Cl}}(u,v,\theta) &= v^{-\theta-1}(u^{-\theta}+v^{-\theta}-1)^{-1-1/\theta},\\
h_{\mathrm{Cl}}^{-1}(u,v,\theta) &= \Big((uv^{\theta+1})^{-\frac{\theta}{\theta+1}} + 1 - v^{-\theta}\Big)^{-1/\theta}.
\end{aligned}$$

Frank:

$$\begin{aligned}
C_\theta^{\mathrm{Fr}}(u,v) &= -\frac{1}{\theta}\log\left(1+\frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right), \quad \theta\in\mathbb{R}.\\
c_\theta^{\mathrm{Fr}}(u,v) &= \frac{-\theta g_1(1+g_{u+v})}{(g_u g_v + g_1)^2}, \quad \text{where } g_y = e^{-\theta y}-1,\\
h_{\mathrm{Fr}}(u,v,\theta) &= \frac{g_u g_v + g_u}{g_u g_v + g_1},\\
h_{\mathrm{Fr}}^{-1}(u,v,\theta) &= -\frac{1}{\theta}\ln\left\{1+\frac{u g_1}{1+g_v(1-u)}\right\}.
\end{aligned}$$

Graphs of these copulas are shown in Figure A.1.1, A.1.2 and A.1.3 for different parameter values. Since $h_{\mathrm{Gu}}^{-1}(u,v,\theta)$ must be obtained numerically, [2] suggest it might be better to fit a Clayton survival copula instead, since it possesses the same property - a heavy right tail. The survival copula is defined in [6] as $\hat{C}(u,v) = u+v-1+C(1-u,1-v)$. We then get

$$\begin{aligned}
\hat{C}_\theta^{\mathrm{Cl}}(u,v) &= u+v-1+\left[(1-u)^{-\theta}+(1-v)^{-\theta}-1\right]^{-1/\theta},\\
\hat{c}_\theta^{\mathrm{Cl}}(u,v) &= (1+\theta)\left[(1-u)(1-v)\right]^{-1-\theta}\left[(1-u)^{-\theta}+(1-v)^{-\theta}-1\right]^{-1/\theta-2}.
\end{aligned}$$

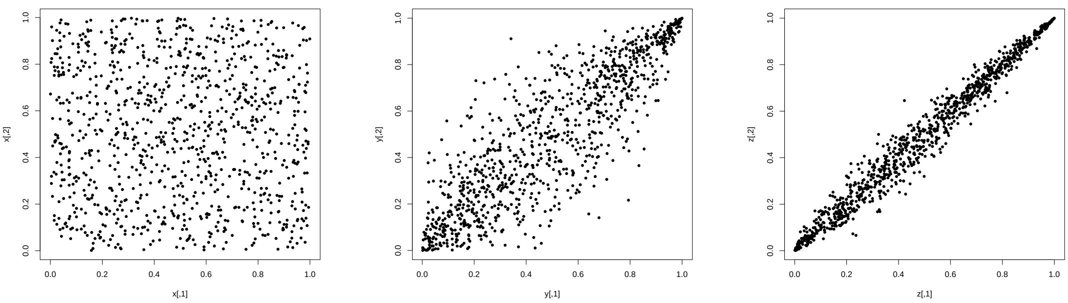
Note that $\hat{c}(u,v) = c(1-u,1-v)$, since

$$\begin{aligned}
\hat{c}(u,v) &= \frac{\partial \hat{C}(u,v)}{\partial u \partial v} = \frac{\partial}{\partial u \partial v}\Big[u+v-1+C(1-u,1-v)\Big]\\
&= \frac{\partial C(1-u,1-v)}{\partial u \partial v} = c(1-u,1-v)\cdot(-1)\cdot(-1) = c(1-u,1-v).
\end{aligned}$$

The h-function and its inverse are

$$\begin{aligned}
\hat{h}_{\mathrm{Cl}}(u,v,\theta) &= 1-(1-v)^{-1-\theta}\Big[(1-u)^{-\theta}+(1-v)^{-\theta}-1\Big]^{-1/\theta-1},\\
\hat{h}_{\mathrm{Cl}}(u,v,\theta)^{-1} &= 1-\left[\left((1-u)(1-v)^{\theta+1}\right)^{\frac{-\theta}{\theta+1}}+1-(1-v)^{-\theta}\right]^{-1/\theta}.
\end{aligned}$$

The following theorem states the construction of an Archimedean copula.



(a) (b) (c)

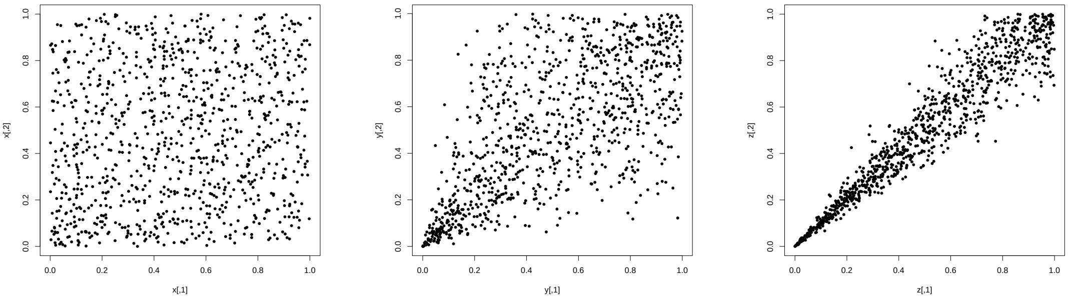
Figure A.1.1: 1000 random variates generated from a Gumbel copula. (a): $\theta = 1.01$ (b): $\theta = 3$ (c): $\theta = 10$.

Theorem A.1.1. Archimedean copula. *Let $\phi : [0,1] \to [0,\infty]$ be continuous and strictly decreasing with $\phi(1) = 0$ and $\phi^{[-1]}(t)$ the pseudo-inverse, see* [6]. *Then*

$$C(u,v) = \phi^{[-1]}(\phi(u)+\phi(v)) \tag{A.1.1}$$

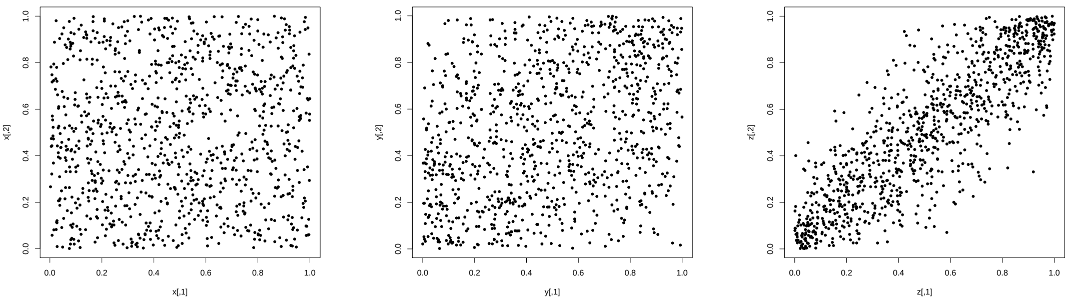
is a copula if and only if ϕ is convex.

A copula constructed according to (A.1.1) is called an Archimedean copula and $\phi(t)$ is known as the **generator** of the copula. That is, if you find a convex function ϕ satisfying (A.1.1), $C(u,v)$ is a copula. If $\phi(0) = \infty$, we say that ϕ is a strict generator. A thorough list of copulas and their respective generators (with boundaries) can be found in Nelsen (1999, pp. 116-119)[6].



(a) (b) (c)

Figure A.1.2: 1000 random variates generated from a Clayton copula. (a): $\theta = 0.2$ (b): $\theta = 2$ (c): $\theta = 10$.



(a) (b) (c)

Figure A.1.3: 1000 random variates generated from Frank a copula. (a): $\theta = 0.2$ (b): $\theta = 2$ (c): $\theta = 10$.

A.2 Gaussian Copulas

In ordinary probability theory one cannot avoid running into the normal distribution. In copula theory, the normal distribution is encountered when discussing the so-called Gauss copula. If $\mathbf{Y} \sim N_d(\mu, \mathbf{\Sigma})$, and $\mathbf{Y}$'s copula is the Gauss copula, then, due to the invariance property of copulas, so is $\mathbf{X}$'s copula when $\mathbf{X} \sim N_d(\mathbf{0}, P)$. P is the correlation matrix of $\mathbf{Y}$. Applying Definition 2.1.1, we get the Gauss copula:

$$\begin{aligned} C_P^{Ga}(\mathbf{u}) &= P(F_1(X_1) \le u_1, \ldots, F_d(X_d) \le u_d) \\ &= P(\Phi(X_1) \le u_1, \ldots, \Phi(X_n) \le u_d) \\ &= P(X_1 \le \Phi^{-1}(u_1), \ldots, X_d \le \Phi^{-1}(u_d)) \\ &= \mathbf{\Phi}_P(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), \end{aligned}$$

where $\mathbf{\Phi}_P$ is the joint distribution function of $\mathbf{X}$ and Φ denotes the standard univariate normal distribution function. The Gauss copula does not have a simple closed form, but can, in two dimensions, be expressed by [5]

$$C_\rho^{Ga}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{\frac{-(u^2 - 2\rho uv + v^2)}{2(1-\rho^2)}\right\} \text{dudv}.$$

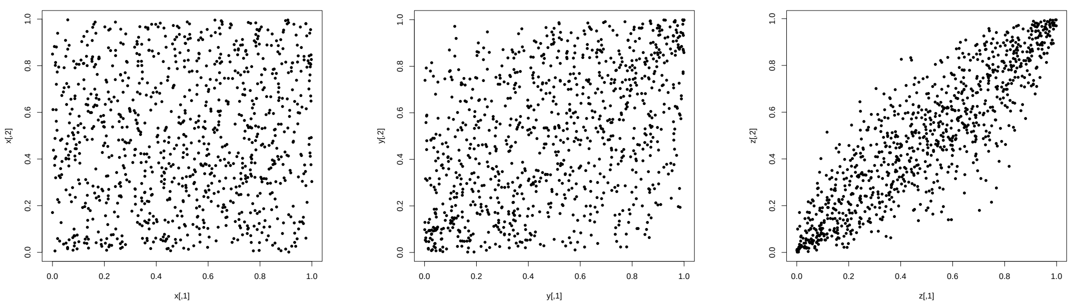
The density is then given by [2]

$$c(u,v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{\rho([\Phi^{-1}(u)]^2 + [\Phi^{-1}(v)]^2) - 2\rho\Phi^{-1}(u)\Phi^{-1}(v)}{2(1-\rho^2)}\right\}.$$

The h-function and its inverse are

$$\begin{aligned} h(u,v,\rho) &= \Phi\left(\frac{\Phi^{-1}(u) - \rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}}\right), \\ h^{-1}(u,v,\rho) &= \Phi\left\{\Phi^{-1}(u)\sqrt{1-\rho^2} + \rho\Phi^{-1}(v)\right\}. \end{aligned}$$

In Figure A.2.1 below is three scatter plots of a normal copula with three different parameter values.



(a) (b) (c)

Figure A.2.1: 1000 random variates generated from a normal copula. (a): $\rho = 0.1$ (b): $\rho = 0.5$ (c): $\rho = 0.9$.

The likelihood function in (3.1.5) becomes [5]

$$l(P; \hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_n) = \sum_{t=1}^{n} f_P(\Phi^{-1}(\hat{U}_{t,1}), \ldots, \Phi^{-1}(\hat{U}_{t,d})) - \sum_{t=1}^{n}\sum_{j=1}^{d} \ln \phi(\Phi^{-1}(\hat{U}_{t,j})).$$

When maximizing this with respect to P, the last summation gets canceled out, and the MLE becomes

$$\hat{P} = \arg\max_{\Sigma \in \mathcal{P}} \sum_{t=1}^{n} \ln f_\Sigma(\mathbf{Y_t}),$$

where f_Σ is a joint density function for a variable $\mathbf{X} \sim N_d(\mathbf{0}, \Sigma)$, $\mathbf{Y_t} = (Y_{t,1}, \ldots, Y_{t,d}) = (\Phi^{-1}(\hat{U}_{t,j}), \ldots, \Phi^{-1}(\hat{U}_{t,d}))$ and $\mathcal{P}$ is the set of all possible linear correlation matrices. [5] give procedures for both finding the exact solution of $\hat{P}$, and an estimate of it.

A.3 Student Copulas

The Student copula (often known as the Student t or just the t copula) is derived in a similar manner as the Gauss copula. It is extracted from the multivariate t-distribution [5], i.e

$$C^t_{\nu,P}(\mathbf{u}) = \mathbf{t}_{\nu,P}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-d}(u_d)),$$

where t_ν is the distribution function of a standard univariate t distribution with ν degrees of freedom, expected value 0 and variance $\frac{\nu}{\nu-2}$. t_ν^{-1} is its quantile function, i.e. the inverse of the cumulative distribution function. $\mathbf{t}_{\nu,P}$ is the joint distribution function of the vector $\mathbf{X} \sim t_d(\nu, \mathbf{0}, P)$, where P is a correlation matrix. As with the Gaussian copula, the Student copulas do not have a simple closed form. The density and h-function are as follows [2]

$$\begin{aligned}
c(u,v) &= \frac{\Gamma(\frac{\nu+2}{2})/\Gamma(\frac{\nu}{2})}{\nu\pi \cdot \mathrm{dt}(t_\nu^{-1}(u),\nu) \cdot \mathrm{dt}(t_\nu^{-1}(v),\nu) \cdot \sqrt{1-\rho^2}} \\
&\quad \times \left\{1 + \frac{[t_\nu^{-1}(u)]^2 + [t_\nu^{-1}(v)]^2 - 2\rho t_\nu^{-1}(u) t_\nu^{-1}(v)}{\nu(1-\rho^2)}\right\}, \\
h(u,v,\rho,\nu) &= t_{\nu+1}\left\{\frac{t_\nu^{-1}(u) - \rho t_\nu^{-1}(v)}{\sqrt{\frac{\left(\nu + [t_\nu^{-1}(v)]^2\right)(1-\rho^2)}{\nu+1}}}\right\}, \\
h^{-1}(u,v,\rho,\nu) &= t_\nu\left\{t_{\nu+1}^{-1}(u)\sqrt{\frac{\left(\nu + [t_\nu^{-1}(v)]^2\right)(1-\rho^2)}{\nu+1}} + \rho t_\nu^{-1}(v)\right\}.
\end{aligned}$$

$\mathrm{dt}(\cdot,\nu)$ is the probability density and $t_\nu^{-1}(\cdot)$ is the quantile function. In Figure A.3.1 below is six scatter plots of a Student copula with different parameter values. Each value of ρ is plotted with two values of ν.

The likelihood function in (3.1.5) becomes [5]

$$l(\nu, P; \hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_n) = \sum_{t=1}^{n} g_{\nu,P}(t_\nu^{-1}(\hat{U}_{t,1}), \ldots, t_\nu^{-1}(\hat{U}_{t,d})) - \sum_{t=1}^{n}\sum_{j=1}^{d} \ln g_\nu(t_\nu^{-1}(\hat{U}_{t,j})),$$

where $g_{\nu,P}$ is the joint density for a variable $\mathbf{X} \sim t_d(\nu, \mathbf{0}, P)$, P is a linear correlation matrix, g_ν is the density of a variable $X \sim t_1(\nu, 0, 1)$ and t_ν^{-1} is the corresponding quantile function.

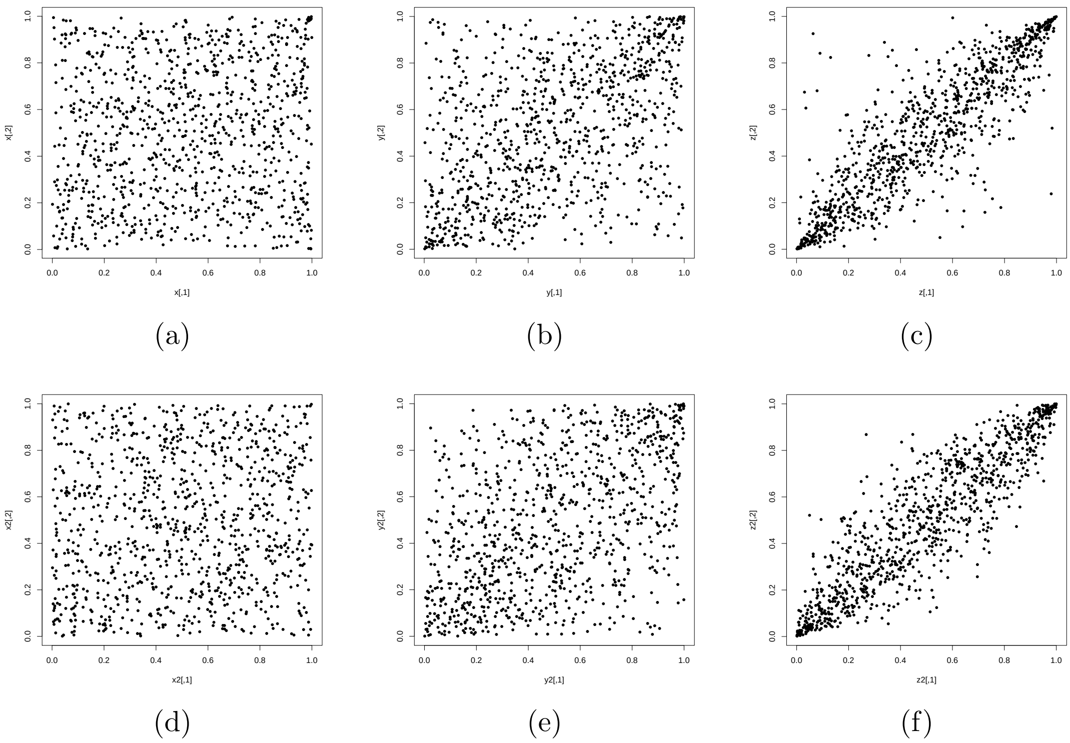


Figure A.3.1: 1000 random variates generated from a Student copula. (a): $\rho = 0.1$, $\nu = 3$ (b): $\rho = 0.5$, $\nu = 3$ (c): $\rho = 0.9$, $\nu = 3$ (d): $\rho = 0.1$, $\nu = 15$, $\rho = 0.5$, $\nu = 15$, $\rho = 0.9$, $\nu = 15$.

Appendix B

Derivations

In this section we derive the h-function described in Section 4.3.

B.1 The h-function

We define $C_1(u,v)$ to be the partial derivative of $C(u,v)$ with respect to its first component, i.e.

$$C_1(u,v) = \frac{\partial C(u,v)}{\partial u},$$

and $C_2(u,v)$ is the derivative with respect to the second component. We have

$$F(x_1|x_2) = \int_{-\infty}^{x_1} f(\tilde{x_1}|x_2)\mathrm{d}\tilde{x_1} = \frac{1}{f(x_2)}\int_{-\infty}^{x_1} f(\tilde{x_1}, x_2)\mathrm{d}\tilde{x_1} = \frac{1}{f(x_2)}\frac{\partial F(x_1,x_2)}{\partial x_2}, \quad \text{(B.1.1)}$$

where the last equality is seen by observing that

$$\frac{\partial F(x_1,x_2)}{\partial x_2} = \frac{\partial}{\partial x_2}\left(\int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f(\tilde{x_1},\tilde{x_2})\mathrm{d}\tilde{x_1}\mathrm{d}\tilde{x_2}\right) = \int_{-\infty}^{x_1} f(\tilde{x_1},x_2)\mathrm{d}\tilde{x_1}.$$

Expanding the conditioning set in (B.1.1), we get

$$\begin{aligned} F(x_1|x_2,x_3) &= \int_{-\infty}^{x_1} f(\tilde{x_1}|x_2,x_3)\mathrm{d}\tilde{x_1} = \frac{1}{f(x_3|x_2)}\int_{-\infty}^{x_1} f(\tilde{x_1},x_3|x_2)\mathrm{d}\tilde{x_1} \\ &= \frac{1}{f(x_3|x_2)}\frac{\partial F(x_1,x_3|x_2)}{\partial x_3}, \end{aligned}$$

where we have used the relation

$$\frac{\partial F(x_1,x_3|x_2)}{\partial x_3} = \frac{\partial}{\partial x_3}\left(\int_{-\infty}^{x_3}\int_{-\infty}^{x_1} f(\tilde{x_1},\tilde{x_3}|x_2)\mathrm{d}\tilde{x_1}\mathrm{d}\tilde{x_3}\right) = \int_{-\infty}^{x_1} f(\tilde{x_1},x_3|x_2)\mathrm{d}\tilde{x_1}.$$

Using the simplifying notations from Section 4.1, we have

$$\text{i)}\quad F_{1|2} = \frac{1}{f_2}\frac{\partial F_{12}}{\partial x_2} = \frac{1}{f_2}\frac{\partial C_{12}}{\partial x_2} = \frac{1}{f_2}\frac{\partial C_{12}}{\partial F_2}\frac{\partial F_2}{\partial x_2} = \frac{\partial C_{12}}{\partial F_2} = C_2(F_1, F_2),$$

$$\text{ii)}\quad F_{1|23} = \frac{1}{f_{3|2}}\frac{\partial F_{13|2}}{\partial x_3} = \frac{1}{f_{3|2}}\frac{\partial C_{13|2}}{\partial x_3} = \frac{1}{f_{3|2}}\frac{\partial C_{13|2}}{\partial F_{3|2}}\frac{\partial F_{3|2}}{\partial x_3} = C_2(F_{1|2}, F_{3|2}),$$

$$F_{3|12} = \frac{1}{f_{1|2}}\frac{\partial F_{13|2}}{\partial x_1} = \frac{1}{f_{1|2}}\frac{\partial C_{13|2}}{\partial x_1} = \frac{1}{f_{1|2}}\frac{\partial C_{13|2}}{\partial F_{1|2}}\frac{\partial F_{1|2}}{\partial x_1} = C_1(F_{1|2}, F_{3|2}),$$

and by induction, see [10]:

$$\text{iii)}\quad F_{1|2,\dots,m} = \frac{\partial}{\partial u_2} C_{1m}(F_{1|2,\dots,m-1}, F_{m|2,\dots,m-1}),$$

where $\frac{\partial}{\partial u_2} C_{1m}(F_{1|2,\dots,m-1}, F_{m|2,\dots,m-1})$ means the partial derivative with respect to the second component.

Appendix C

Sampling From Four- and Five-Dimensional Vines

In this section we derive the expressions for x_4 and x_5 for a canonical vine. Recall that the difference between sampling from a canonical and a D-vine is the choice of v_j in (4.3.1). The indices of Θ_{ij} are chosen to represent the copula it is used in for the benefit of the reader. That is, we have not used the notations in either of the algorithms in Section 6.1.

$$
\begin{aligned}
x_1 =& w_1,\\
x_2 =& F^{-1}(w_2|x_1)\\
&= h^{-1}(w_2, x_1, \Theta_{11}),\\
x_3 =& F^{-1}(w_3|x_1, x_2)\\
&= h^{-1}\left[h^{-1}(w_3, h(x_2, x_1, \Theta_{11}), \Theta_{21}), x_1, \Theta_{12}\right],\\
x_4 =& F^{-1}(w_4|x_1, x_2, x_3) =\\
&h^{-1}\Bigg\{h^{-1}\Bigg[h^{-1}\Bigg\{w_4, h\Bigg[h(x_3, x_1, \Theta_{13}), h(x_2, x_1, \Theta_{12}), \Theta_{23|1}\Bigg], \Theta_{34|12}\Bigg\}, h(x_2, x_1, \Theta_{12}),\\
&\Theta_{24|1}\Bigg], x_1, \Theta_{14}\Bigg\},\\
x_5 =& F^{-1}(w_5|x_1, \ldots, x_4) =\\
&h^{-1}\Bigg(h^{-1}\Bigg\{h^{-1}\Bigg[h^{-1}\Bigg\{w_5, h\Bigg[h\Bigg\{h(x_4, x_1, \Theta_{14}), h(x_2, x_1, \Theta_{12}), \Theta_{24|1}\Bigg\},\\
&h\Bigg\{h(x_3, x_1, \Theta_{13}), h(x_2, x_1, \Theta_{12}), \Theta_{23|1}\Bigg\}, \Theta_{34|12}\Bigg], \Theta_{45|123}\Bigg\},\\
&h\Bigg\{h(x_3, x_1, \Theta_{13}), h(x_2, x_1, \Theta_{12}), \Theta_{23|1}\Bigg\}, \Theta_{35|12}\Bigg], h(x_2, x_1, \Theta_{12}), \Theta_{25|1}\Bigg\}, x_1, \Theta_{15}\Bigg).
\end{aligned}
$$

The expressions for x_4 and x_5 are derived in the two next sections.

C.1 Four Dimensions

$$
\begin{aligned}
w_4 =& F(x_4|x_1,x_2,x_3)\\
=& \frac{\partial C_{x_4x_3|x_1x_2}(F_{x_4|x_1x_2},F_{x_3|x_1x_2})}{\partial F_{x_3|x_1x_2}}\\
=& \frac{\partial C_{x_4x_3|x_1x_2}\left(\frac{\partial C_{x_4x_2|x_1}(F_{x_4|x_1},F_{x_2|x_1})}{\partial F_{x_2|x_1}},\frac{\partial C_{x_3x_2|x_1}(F_{x_3|x_1},F_{x_2|x_1})}{\partial F_{x_2|x_1}}\right)}{\partial F_{x_3|x_1x_2}}\\
=& h\Big\{h\Big[h(x_4,x_1,\Theta_{14}),h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big],h\Big[h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big],\\
&\Theta_{34|12}\Big\}
\end{aligned}
$$

We solve for x_4:

$$
\begin{aligned}
&h\Big[h(x_4,x_1,\Theta_{14}),h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big] = h^{-1}\Big\{w_4,h\Big[h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big],\\
&\Theta_{34|12}\Big\}\\
\Rightarrow\ & h(x_4,x_1,\Theta_{14}) =\\
&h^{-1}\Big[h^{-1}\Big\{w_4,h\Big[h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big],\Theta_{34|12}\Big\},h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big]\\
\Rightarrow\ & x_4 =\\
&h^{-1}\Big\{h^{-1}\Big[h^{-1}\Big\{w_4,h\Big[h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big],\Theta_{34|12}\Big\},h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big],\\
&x_1,\Theta_{14}\Big\}.
\end{aligned}
$$

C.2 Five Dimensions

$$
\begin{aligned}
w_5 =& F(x_5|x_1,\ldots,x_4)\\
=& \frac{\partial C_{x_5x_4|x_1x_2x_3}(F_{x_5|x_1x_2x_3},F_{x_4|x_1x_2x_3})}{\partial F_{x_4|x_1x_2x_3}}\\
=& \frac{\partial C_{x_5x_4|x_1x_2x_3}\left(\frac{\partial C_{x_5x_3|x_1x_2}(F_{x_5|x_1x_2},F_{x_3|x_1x_2})}{\partial F_{x_3|x_1x_2}},\frac{\partial C_{x_4x_3|x_1x_2}(F_{x_4|x_1x_2},F_{x_3|x_1x_2})}{\partial F_{x_3|x_1x_2}}\right)}{\partial F_{x_4|x_1x_2x_3}}\\
=& \partial C_{x_5x_4|x_1x_2x_3}\left(\frac{\partial C_{x_5x_3|x_1x_2}\left(\frac{\partial C_{x_5x_2|x_1}(F_{x_5|x_1},F_{x_2|x_1})}{\partial F_{x_2|x_1}},\frac{\partial C_{x_3x_2|x_1}(F_{x_3|x_1},F_{x_2|x_1})}{\partial F_{x_2|x_1}}\right)}{\partial F_{x_3|x_1x_2}},\right.\\
&\left.\frac{\partial C_{x_4x_3|x_1x_2}\left(\frac{\partial C_{x_4x_2|x_1}(F_{x_4|x_1},F_{x_2|x_1})}{\partial F_{x_2|x_1}},\frac{\partial C_{x_3x_2|x_1}(F_{x_3|x_1},F_{x_2|x_1})}{\partial F_{x_2|x_1}}\right)}{\partial F_{x_3|x_1x_2}}\right)\Bigg/\partial F_{x_4|x_1x_2x_3}
\end{aligned}
$$

$$=h\Big\{h\Big[h\Big\{h(x_5,x_1,\Theta_{15}),h(x_2,x_1,\Theta_{12}),\Theta_{25|1}\Big\},h\Big\{h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big\},$$
$$\Theta_{35|12}\Big],7h\Big[h\Big\{h(x_4,x_1,\Theta_{14}),h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big\},h\Big\{h\big[x_3,x_1,\Theta_{13}\big],h\big[x_2,x_1,\Theta_{12}\big],$$
$$\Theta_{23|1}\Big\},\Theta_{34|12}\Big],\Theta_{45|123}\Big\}.$$

We solve for x_5:

$$h\Big[h\Big\{h(x_5,x_1,\Theta_{15}),h(x_2,x_1,\Theta_{12}),\Theta_{25|1}\Big\},h\Big\{h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big\},\Theta_{35|12}\Big]$$
$$=h^{-1}\Big\{w_5,h\Big[h\Big\{h(x_4,x_1,\Theta_{14}),h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big\},h\Big\{h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big\},$$
$$\Theta_{34|12}\Big],\Theta_{45|123}\Big\}$$
$$\Rightarrow h\Big\{h(x_5,x_1,\Theta_{15}),h(x_2,x_1,\Theta_{12}),\Theta_{25|1}\Big\}$$
$$=h^{-1}\Big[h^{-1}\Big\{w_5,h\Big[h\Big\{h(x_4,x_1,\Theta_{14}),h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big\},h\Big\{h\big[x_3,x_1,\Theta_{13}\big],h\big[x_2,x_1,\Theta_{12}\big],$$
$$\Theta_{23|1}\Big\},\Theta_{34|12}\Big],\Theta_{45|123}\Big\},h\Big\{h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big\},\Theta_{35|12}\Big]$$
$$\Rightarrow h(x_5,x_1,\Theta_{15})=$$
$$h^{-1}\Big\{h^{-1}\Big[h^{-1}\Big\{w_5,h\Big[h\Big\{h(x_4,x_1,\Theta_{14}),h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big\},h\Big\{h(x_3,x_1,\Theta_{13}),$$
$$h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big\},\Theta_{34|12}\Big],\Theta_{45|123}\Big\},h\Big\{h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big\},\Theta_{35|12}\Big],$$
$$h(x_2,x_1,\Theta_{12}),\Theta_{25|1}\Big\}$$
$$\Rightarrow x_5=$$
$$h^{-1}\Big(h^{-1}\Big\{h^{-1}\Big[h^{-1}\Big\{w_5,h\Big[h\Big\{h(x_4,x_1,\Theta_{14}),h(x_2,x_1,\Theta_{12}),\Theta_{24|1}\Big\},h\Big\{h(x_3,x_1,\Theta_{13}),$$
$$h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big\},\Theta_{34|12}\Big],\Theta_{45|123}\Big\},h\Big\{h(x_3,x_1,\Theta_{13}),h(x_2,x_1,\Theta_{12}),\Theta_{23|1}\Big\},\Theta_{35|12}\Big],$$
$$h(x_2,x_1,\Theta_{12}),\Theta_{25|1}\Big\},x_1,\Theta_{15}\Big).$$

Appendix D

Verifying Approach 1 and 2

In this chapter we verify the two approaches discussed in Chapter 7.2.1. We do this by performing tests of hypotheses on data arising from a known copula. We use the two hypotheses

$$\begin{aligned} \mathrm{H}_0 &: C \text{ is a Clayton copula with } \theta = 2.00 \\ \text{vs. } \mathrm{H}_1 &: C \text{ is } \textit{not} \text{ a Clayton copula with } \theta = 2.00 \end{aligned}$$

and

$$\begin{aligned} \mathrm{H}_0^* &: C \text{ is a Student copula with } \rho = 0.50 \text{ and } \nu = 6 \\ \text{vs. } \mathrm{H}_1^* &: C \text{ is a } \textit{not} \text{ Student copula with } \rho = 0.50 \text{ and } \nu = 6 \end{aligned}$$

in our verification procedure.

D.1 Results

We test H_0 with approach 1 and 2, and H_0^* with approach 1. We verify the approaches by testing an on before-hand known true null hypothesis. The test statistic (p-value) is uniformly distributed between 0 and 1 if the null hypothesis is true [20]. Hence approximately 5% of the tests should be rejected at a significance level on 0.05. The results together with the values used on constants in each test are listed in Figure D.1.1, D.1.2 and D.1.3. We conclude that our implementation of both approaches are valid.

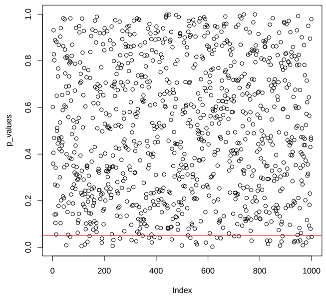


Figure D.1.1: Estimated p-values when testing H_0 with approach 1. The data arises from a Clayton copula with $\theta = 2.00$. $n = 500$ and $K = 1000$. # Hypotheses tested: 1000. % null hypotheses rejected $= 5.4$.

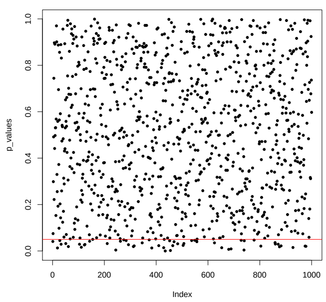


Figure D.1.2: Estimated p-values when testing H_0^* with approach 1. The data arises from a Student copula with $\rho = 0.50$ and $\nu = 6$. $n = 100$, $K = 1000$ and $N_b = 2500$. # Hypotheses tested: 1000. % null hypotheses rejected $= 4.0$.

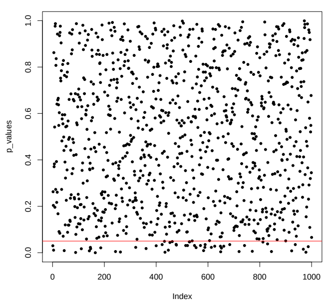


Figure D.1.3: Estimated p-values when testing H_0 with approach 2. The data arises from a Clayton copula with $\theta = 2.00$. $n = 100$ and $K = 1000$. # Hypotheses tested: 1000. % null hypotheses rejected $= 4.4$.

Appendix E

R-code

In this chapter we present a graphic overview of our implemented R-code.

E.1 Graphic Overview

Functions/packages marked with * are built-in R-functions. An arrow imply that the method pointed at is used in the method. The code itself with examples can be seen by contacting Håvard Rue or the author.

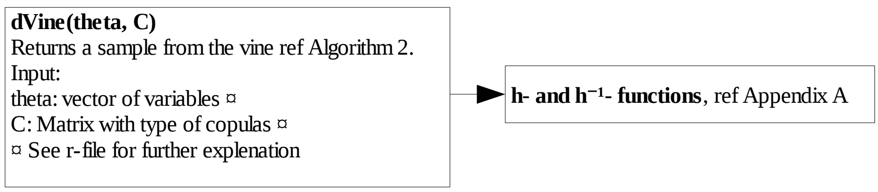


Figure E.1.1: Implementation of Algorithm 2

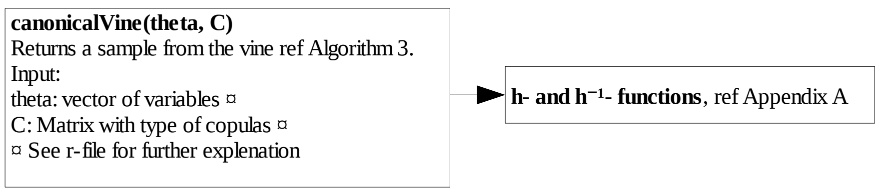


Figure E.1.2: Implementation of Algorithm 3

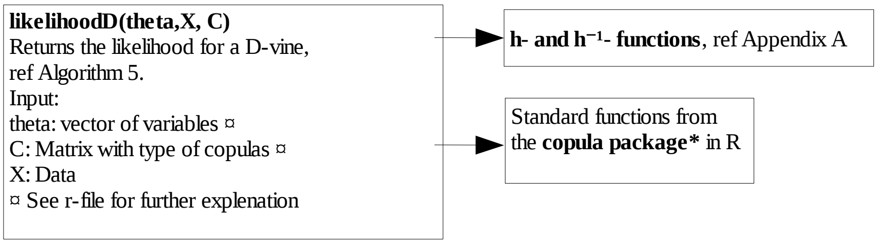


Figure E.1.3: Implementation of Algorithm 4

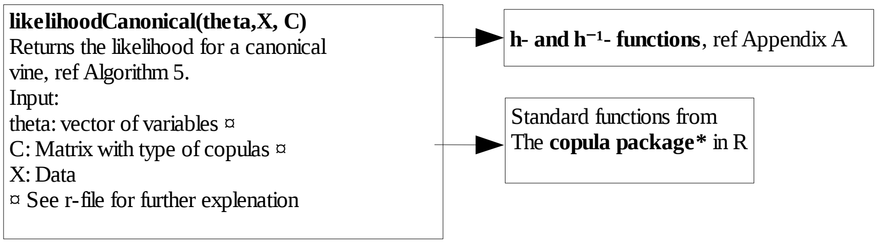


Figure E.1.4: Implementation of Algorithm 5

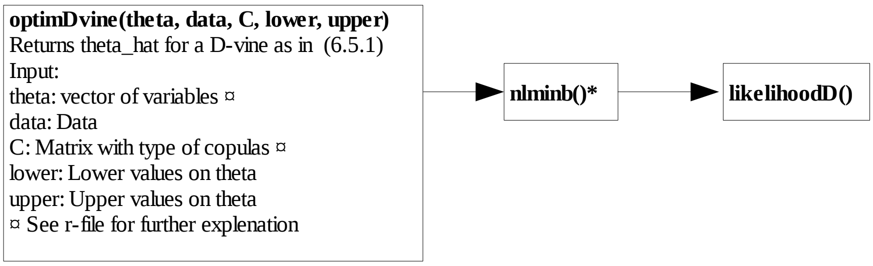


Figure E.1.5: Implementation of the optimization of the log-likelihood of a D-vine.

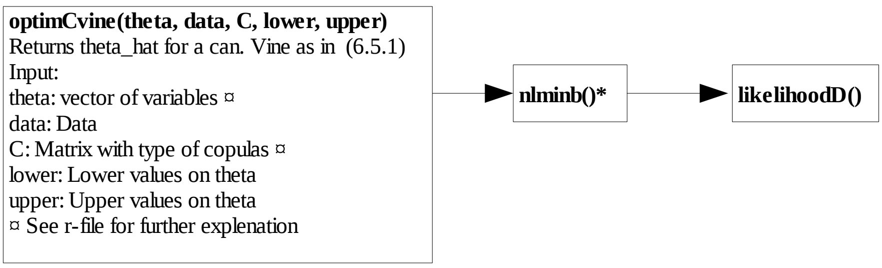


Figure E.1.6: Implementation of the optimization of the log-likelihood of a canonical vine.

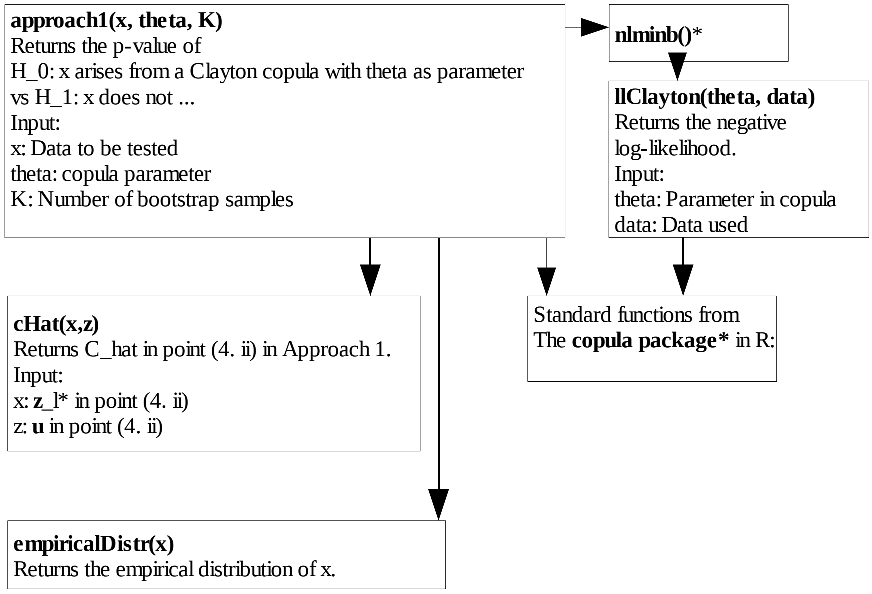


Figure E.1.7: Implementation of approach 1 in Section 7.2.1 for a Clayton copula.

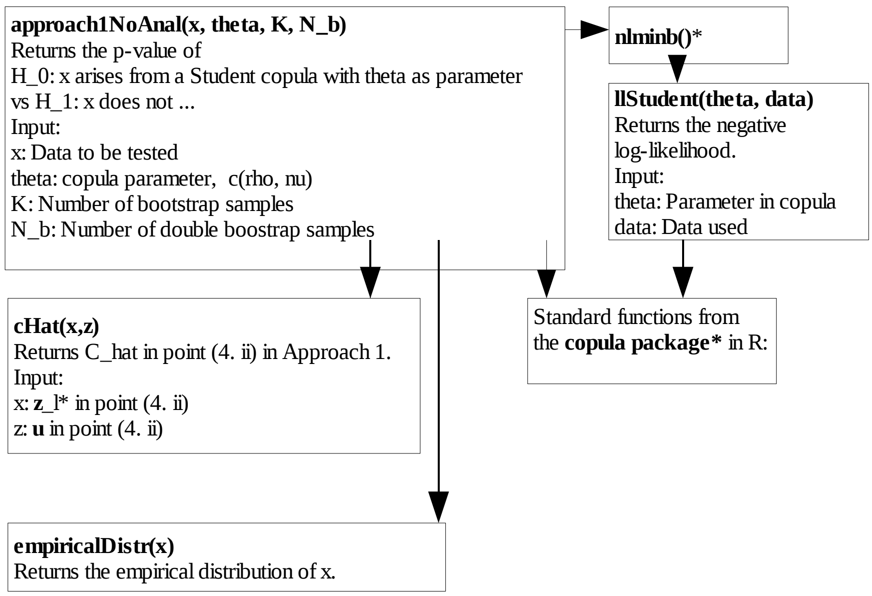


Figure E.1.8: Implementation of approach 1 in Section 7.2.1 for a Student copula.

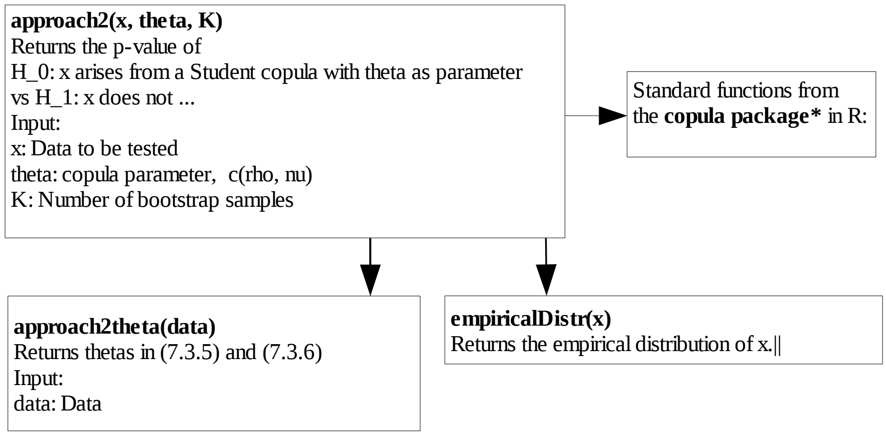


Figure E.1.9: Implementation of approach 2 in Section 7.2.1 for a Clayton copula.

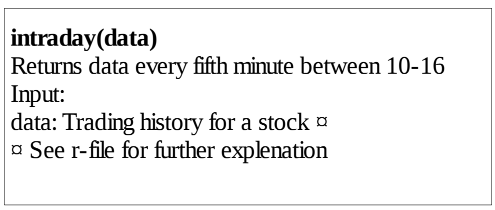


Figure E.1.10: Script for extracting intra day prices.