



Nikolaus Schafzahl, BSc

Measuring portfolio risk: Application of C-vine, D-vine and R-vine copulas in comparison to multivariate t-copulas and traditional methods

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Finanz- und Versicherungsmathematik

submitted to

Graz University of Technology

Supervisor

Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Eranda Dragoti-Çela

Institute of Discrete Mathematics

Head: Univ.-Prof. Dipl.-Ing Dr.rer.nat. Wolfgang Woess

Graz, March 2018

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date

Signature

Abstract

Measuring portfolio risk using Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) has been a hot issue in both academia and the financial industry for some time. Of special interests for financial institutions is often the influence of coupled risk. This thesis constructs a vine copula based GARCH(1,1) VaR and CVaR model combining the use of GARCH models and copula functions to allow for flexibility in the choice of marginal distributions and for a rich dependence structure. Vine copulas are a cascade of bivariate copula constructions allowing for multivariate dependence modeling by using bivariate copulas as building blocs, and serve as a useful alternative to multivariate copulas for high-dimensional portfolios. We apply this modeling approach, using three different vine copula structures (C-vine, D-vine and R-vine), to a six-dimensional equally weighted portfolio and test the performance of these models in forecasting daily VaR and CVaR over a period of 1757 days. For comparison we perform the same calculations using a multivariate Student's t-copula model. An approach using fixed dependence structure was compared to a 50 days refitted rolling window approach. The empirical results show that refitting the GARCH(1,1) model and the vine copula structure every 50 days can improve the measurement reliability, provided that a sufficiently large number of simulations is used, whereas a fixed model approach might struggle to adjust to market changes, especially in periods of high volatility such as the recent global financial crisis.

Contents

Abstract	iii
1. Introduction	1
1.1. Risk	2
1.2. Risk measures	4
1.2.1. Value-at-Risk	5
1.2.2. Conditional Value-at-Risk	7
1.3. Traditional methods for calculating VaR and CVaR	8
1.3.1. Historical simulation method (HS)	9
1.3.2. Mean variance method (MV)	11
1.4. Methods of backtesting VaR and CVaR	13
1.4.1. VaR: Kupiec and Christoffersen coverage test	13
1.4.2. CVaR: Zero mean test	21
1.5. Copula	22
1.6. Dependence measures	23
1.6.1. Pearson correlation coefficient	23
1.6.2. Spearman's ρ	24
1.6.3. Kendall's τ	25
1.6.4. Tail dependence	26
2. Univariate GARCH model	27
2.1. GARCH(1,1) model	31
2.1.1. Forecasting volatility	33
2.1.2. Estimation of parameters	35
2.1.3. Non-Gaussian error distributions	37
2.2. Ljung-Box test	38
2.3. Limitations of the GARCH model	39

Contents

3. Bivariate copula families	40
3.1. Examples of common elliptical and Archimedean bivariate copulas	44
3.2. Elliptical copulas	45
3.2.1. Gaussian copula	45
3.2.2. Student's t-copula	46
3.3. Archimedean copulas	49
3.3.1. Clayton copula	50
3.3.2. Gumbel copula	52
3.3.3. Frank copula	55
3.3.4. Joe copula	56
3.4. Simulation of Archimedean and elliptical copulas	57
3.4.1. Simulation of elliptical copulas	57
3.4.2. Simulation of Archimedean copulas	58
3.5. Summary of presented copula families and their characteristics	61
4. Vine copula	64
4.1. Pair-copula-construction - PCC	65
4.2. Introduction to vines	68
4.3. Vine copula specification	74
4.3.1. Tree structure selection	75
4.3.2. Copula family selection	77
4.3.3. Parameter estimation	80
5. Empirical studies and analysis	82
5.1. R implementation and packages	82
5.2. Setting up of fixed vs. refitted rolling window models	84
5.2.1. Fixed model	84
5.2.2. Refitted rolling window models	84
5.3. Data and analysis of daily log-returns	85
5.4. Marginal distribution estimates	92
5.5. Finding the vine copula and Student's t-copula models	100
5.5.1. C-vine	101
5.5.2. D-vine	103
5.5.3. R-vine	105
5.5.4. Student's t-copula	107
5.5.5. Comparison of models	108

Contents

5.6. Procedure: Forecasting and backtesting	109
6. Results	115
6.1. Backtesting: Traditional methods - VaR and CVaR	116
6.2. Backtesting: Fixed models - VaR and CVaR	118
6.2.1. Fixed model: Value-at-Risk:	119
6.2.2. Fixed model: Conditional-Value-at-Risk:	120
6.3. Backtesting: 50 days refitted rolling window models - VaR and CVaR	122
6.3.1. 50 days refitted rolling window model: Value-at-Risk:	122
6.3.2. 50 days refitted rolling window model: Conditional- Value-at-Risk:	123
6.4. Graphical comparison of 50 days refitted rolling window model vs. fixed model	125
6.5. Runtime comparison	127
A. Proofs, calculations etc.	134
A.1. Likelihood ratio test	134
A.2. Calculations for likelihood ratio test of unconditional cover- age (UC)	136
A.2.1. Maximum likelihood estimate of π for the likelihood function of the alternative hypothesis	136
A.3. Calculations for likelihood ratio test of independence (IND)	137
A.3.1. Maximum likelihood estimate of π_{01} and π_{11} for the likelihood function of the alternative hypothesis	137
A.3.2. Maximum likelihood estimate of π for the likelihood function of the zero hypothesis	138
A.4. Derivation of h-functions for different copulas	139
A.4.1. Gaussian copula	139
A.4.2. Student's t-copula	141
A.4.3. Clayton copula	144
A.4.4. Gumbel copula	144
A.4.5. Frank copula	145
A.4.6. Joe copula	145
A.5. Inverse Laplace-Stiltjes transforms F	146
A.5.1. Clayton copula	146
A.5.2. Gumbel copula	147

Contents

A.5.3. Frank copula	148
A.5.4. Joe copula	148

Bibliography	150
---------------------	------------

List of Figures

1.1.	Illustration of $VaR_{0.05}$.	6
1.2.	Illustration of $CVaR_{0.05}$.	7
1.3.	Illustration of historical simulation method, from Harper (2017).	10
1.4.	Illustration of mean variance method, from Harper (2017).	11
2.1.	ACF plot: volatility clustering vs. no volatility clustering.	29
3.1.	Perspective plots and contour plots of the comonotonicity copula, the independence copula and the countermonotonicity copula.	43
3.2.	Elliptical copulas: Gaussian copula vs. Student's t-copula.	48
3.3.	Summary of copulas with different dependence parameters.	63
4.1.	An example R-vine on five variables and four trees.	71
4.2.	An example D-vine on five variables and four trees.	72
4.3.	An example C-vine on five variables and four trees.	73
5.1.	Observations used for one-day-ahead forecasting.	85
5.2.	Adjusted closing prices and the corresponding daily log-returns of ALV.DE.	86
5.3.	Adjusted closing prices and the corresponding daily log-returns of BMW.DE.	86
5.4.	Adjusted closing prices and the corresponding daily log-returns of $\hat{G}DAXI$.	87
5.5.	Adjusted closing prices and the corresponding daily log-returns of $\hat{T}NX$.	87
5.6.	Adjusted closing prices and the corresponding daily log-returns of TM.	88
5.7.	Adjusted closing prices and the corresponding daily log-returns of $\hat{N}255$.	88

List of Figures

5.8. Normalized adjusted closing prizes using 100 as the base value at 1/1/2004.	89
5.9. Log-returns of equally weighted portfolio.	89
5.10. Normal QQ-plots of all daily log-returns.	90
5.11. Empirical Kendall's τ of log-returns computed pairwise for the fitting period.	91
5.12. ACF of the squared values of the daily log-returns.	93
5.13. Comparison of GARCH(1,1) residuals vs. GARCH(1,1) standardized residual distribution for each time series for the first 943 daily observations.	94
5.14. GARCH(1,1) standardized residuals for the first 943 days.	95
5.15. Selected diagnostic plots of a GARCH(1,1) model with Student's t error distribution fitted to ALV.DE data from the first 943 days.	96
5.16. Selected diagnostic plots of a GARCH(1,1) model with Student's t error distribution fitted to BMW.DE data from the first 943 days.	96
5.17. Selected diagnostic plots of a GARCH(1,1) model with skew-generalized error distribution fitted to ^GDAXI data from the first 943 days.	97
5.18. Selected diagnostic plots of a GARCH(1,1) model with Student's t error distribution fitted to ^TNX data from the first 943 days.	97
5.19. Selected diagnostic plots of a GARCH(1,1) model with Student's t error distribution fitted to TM data from the first 943 days.	98
5.20. Selected diagnostic plots of a GARCH(1,1) model with skewed Student's t error distribution fitted to ^N225 data from the first 943 days.	98
5.21. C-vine: Tree structure.	101
5.22. C-vine: Contour plots of all pair-copulas.	102
5.23. C-vine: Pair plots and Kendall's τ	102
5.24. D-vine: Tree structure.	103
5.25. D-vine: Contour plots of all pair-copulas.	104
5.26. D-vine: Pair plots and Kendall's τ	104
5.27. R-vine: Tree structure.	105
5.28. R-vine: Contour plots of all pair-copulas.	106

List of Figures

5.29. R-vine: Pair plots and Kendall's τ	106
5.30. Pair plots and Kendall's tau.	107
5.31. First level tree of C/D/R vine structures.	108
5.32. Distribution of daily log-returns of all 6 given assets.	110
6.1. VaR backtest: Historical simulation and mean variance.	117
6.2. VaR backtest: Age-weighted historical simulation.	118
6.3. VaR backtest: Fixed model - C/D/R-vine copula and Student's t-copula.	119
6.4. CVaR backtest: Fixed model - C/D/R-vine copula and Student's t-copula.	121
6.5. VaR backtest: Rolling window - C/D/R-vine copula and Student's t-copula.	123
6.6. VaR backtest: Rolling window - C/D/R-vine copula and Student's t-copula.	124
6.7. VaR backtest: Rolling window vs. fixed - Student's t-copula.	125
6.8. VaR backtest: Rolling window vs. fixed - C-vine copula.	125
6.9. Log-returns of equally weighted portfolio.	126

1. Introduction

There are a number of approaches to quantify risk in modern financial risk management. With the most modern and consistent being quantile based, it is of major interest to us to thoroughly understand the distribution of portfolio returns. Most financial portfolios consist of multiple assets, which could lead to complications. Copulas are an important tool to model dependencies. While it is easy to construct bivariate copulas, the construction of flexible high-dimensional copulas can be a more challenging problem. Although there are a multitude of parametric bivariate copulas, the class of high-dimensional copulas is still rather limited. Vine copula models (proposed by Joe (1997)) offer great advantages in describing the relationship among multi-dimensional portfolios. The introduction of pair-copula-constructions (PCC) (see Aas et al. (2006)) has been an enormous advancement for high-dimensional dependence modeling. It gives us the benefits of rich selection of bivariate copula families, while simultaneously adding flexibility to our model.

In this thesis a model for analyzing stocks and calculating Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR or alternatively ES for Expected-Shortfall), with vine copula based GARCH_(1,1) modeling is presented. VaR and CVaR are widely used instruments to measure market risk. However, risk measures such as VaR and CVaR are only accurate as long as the return distribution is precisely modeled. Typical choices of confidence levels connected with VaR and CVaR are 95%, 97% and 99% respectively.

This thesis consists of six chapters, whereof four are of theoretical nature and the last two deal with empirical studies and analysis and results. The first chapter of this thesis introduces mathematical concepts and definitions which are needed to be able to understand the further chapters. The second chapter gives a brief introduction to the GARCH model especially the

1. Introduction

GARCH(1,1) model, which will later be used to model the marginal distributions of the different stock markets. In Chapter 3 we present an overview of the most commonly used bivariate copula families and their properties. The fourth chapter will be all about the modeling of inter-structure dependence of these stock markets with R-vine, C-vine and D-vine copula models using the bivariate copula families introduced in Chapter 3 as building blocs. Here we will introduce the pair-copula-construction (PCC) approach.

The last two chapters of this master's thesis are concerned with practical analysis of data and results. Here we estimate VaR and CVaR in the C-vine, D-vine and R-vine model based on daily data and compare the results to values calculated from a multivariate Student's t-copula and traditional methods like historical simulation (HS) and the mean variance model (MV). We consider two backtesting scenarios: First we look at a fixed dependence model and in a second approach we re-estimate the dependence structure every 50 days.

Summarizing, we will estimate the joint distribution of a portfolio of six assets using R-vine, C-vine and D-vine copula models. After evaluating the joint distribution of this six dimensional portfolio we can predict Value-at-Risk and Conditional-Value-at-Risk and use backtesting procedures to evaluate the accuracy of our estimations. For comparison we used a multivariate Student's t-copula model.

For the sake of completeness and fully understanding the following thesis and results the remainder of this chapter is devoted to some standard mathematical concepts and definitions.

1.1. Risk

Since the financial crisis from 2007 until 2009 and its aftermath, the discussion of systemic risk has become more important than ever before. The default of Lehman Brothers in September 2008 showed strikingly that the abrupt meltdown of one key global financial player does not only affect other financial institutions but can also lead to serious instability and endangerment of real economies in countries all around the globe.

1. Introduction

According to *Merriam Webster Dictionary* risk is defined as the “*possibility of loss or injury*” or alternatively “*the chance that an investment (such as a stock or commodity) will lose value*”. While in the eyes of most people risk represents something negative it can also be an opportunity. A entire industry has formed around financial products and risk management, that would not exist if there was no risk. The products and services offered by those financial institutions usually give their clients and customers more financial stability and certainty.

According to A. J. McNeil, Frey, and Embrechts (2006) there are three main types of risk that are encountered in the context of financial and insurance industries: Market Risk, Operational Risk and Credit Risk.

Especially in banking *Market Risk* is well-known. It is the possibility of loss experienced by an investor due to the financial market in which he is investing and its general performance. It is not possible to eliminate market risk by diversification, although an investor can construct a hedge against it. Changes in value of underlying assets can occur due to several reasons including recession, political turmoil, natural disasters or terrorist attacks.

The second important category is *Credit Risk*, which refers to the risk that a borrower does not repay its promised payments on outstanding investments, including loans, bonds and other obligations. This usually happens due to default of the borrower. It is almost never certain whether the debtors will definitely have the assets to repay their debts or not.

Lastly, *Operational Risks* are the risks undertaken by a company operating within a given field or industry. It can be summarized as the risk of losses resulting from failed or inadequate internal processes, people or systems. External events can also be a source of operational risk. It can be characterized as the risk of business operations failing due to human error. The main differentiator of this risk category to the other two categories are that it is not revenue driven.

We always have to keep in mind that the boundaries between all three risk categories are not precisely defined and not always clear. In this thesis we will analyze a portfolio of different assets and try to evaluate the market risk connected with these assets using so-called risk measures.

1. Introduction

1.2. Risk measures

A risk measure needs to fulfill a list of desirable properties in order to be a good risk measure. The selection of the appropriate risk measure is crucial. In financial economics and actuarial science there are many ways to define risk. In order to clarify the concept Artzner, Delbaen, and Ebner (1999) described a number of properties that a risk measure might have. Let \mathcal{G} be a vector space of random variables that represent portfolio values at a fixed future date.

Definition 1 (Coherent risk measure)

Let $\rho : \mathcal{G} \rightarrow \mathbb{R}$ be a risk measure. A risk measure satisfying the following four axioms is called coherent.

Axiom T - Translation invariance

For all $X \in \mathcal{G}$ and all real numbers α , we have $\rho(X + \alpha) = \rho(X) - \alpha$.

Axiom S - Subadditivity

For all X and $Y \in \mathcal{G}$, we have $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

Axiom PH - Positive homogeneity

For all $\lambda > 0$ and all $X \in \mathcal{G}$, we have $\rho(\lambda X) = \lambda \rho(X)$.

Axiom M - Monotonicity

For all X and $Y \in \mathcal{G}$ with $X \stackrel{a.s.}{\leq} Y$, we have $\rho(Y) \leq \rho(X)$.

A risk measure assigns a monetary value to risk, which has economically speaking approximately the same value as equity needed to hedge against the same risk. Describing risk using a single number is connected to a great loss of information. However the decision of taking risk or not is fundamentally binary. Many times standard deviation (or variance) is a sufficient measure of risk. More recently however, other risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR), also known as Expected-Shortfall (ES), are more and more being used in practice. Also Basel II considers VaR as the preferred risk measure.

1. Introduction

1.2.1. Value-at-Risk

Value-at-Risk(VaR) is widely used as measurement of financial risk and plays a vital role in modern risk measurement. It was first introduced by Baumol (1963). Although VaR can be interpreted without difficult it is not trivial to estimate.

Often VaR is analysed in terms of loss L by using the loss distribution F_L . Here we will follow a different approach by using the actual portfolio distribution of X . In our case X is representing the distribution of change in portfolio value in a given day.

Definition 2 (Value-at-Risk)

Given some confidence level $\alpha \in (0, 1)$ the VaR_α of an underlying X is defined as the smallest number x so that $\mathbb{P}(X \leq x) \geq \alpha$.

$$\begin{aligned} VaR_\alpha(X) &= \inf\{x \in \mathbb{R} | \mathbb{P}(X \leq x) \geq \alpha\} \\ &= \inf\{x \in \mathbb{R} | F_X(x) \geq \alpha\} \end{aligned}$$

Definition 3 (Generalized inverse function)

Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a increasing function ($x \leq y \Rightarrow F(x) \leq F(y)$). The generalized inverse function of F is defined as

$$F^{\leftarrow}(y) := \inf\{x \in \mathbb{R} : F(x) \geq y\}.$$

We use the convention that the infimum of an empty set is ∞ .

Therefore the VaR is simply the α -quantile of the return distribution,

$$VaR_\alpha(L) = q_\alpha(F) = F^{\leftarrow}(\alpha),$$

where

$$F : \mathbb{R} \rightarrow \mathbb{R}, q_\alpha(F) := \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

is the α -quantile of F .

Keep in mind that VaR always relates to some confidence level α , typically around 1 to 5 percent and hence does not give us a maximum loss but rather tells us the worst portfolio outcome that happens once every so many days. Clearly this is not sufficient to tell us everything about risk. Thus one major

1. Introduction

flaw of VaR is that it does not tell us how large the likely magnitude of losses is on the days that the return is worse than the VaR.

Artzner, Delbaen, and Ebner (1999) have shown that VaR fails to satisfy the subadditivity property and therefore is no coherent risk measure as described in the beginning of this section. One way to conquer this shortcoming is the introduction of Conditional-Value-at-Risk (CVaR), as an alternative to traditional VaR, which is defined as the expected return conditional on the return being worse than the VaR.

Banking laws and the *Basel Committee on Banking and Supervision* define VaR as a main component of risk quantification. It regulates that . . . (a) "Value-at-risk" must be computed on a daily basis. (b) In calculating the value-at-risk, a 99-th percentile, one-tailed confidence interval is to be used. . . [Basel Committee on Banking Supervision (2009)] Although there do exist alternatives to VaR, it still remains the most commonly used risk measure.

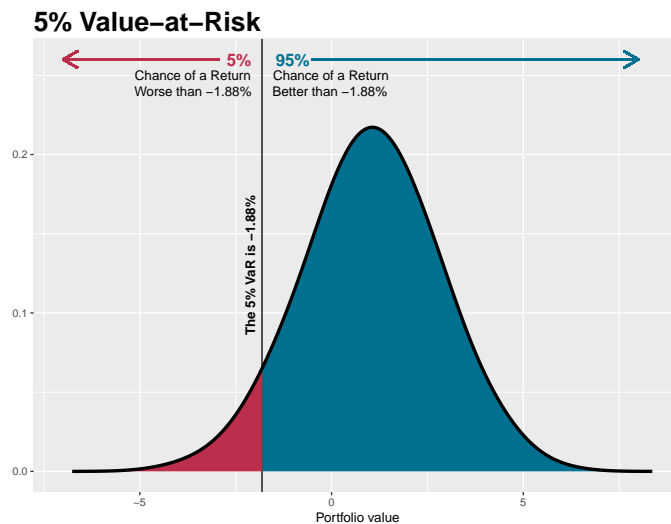


Figure 1.1.: Illustration of $VaR_{0.05}$.

Figure 1.1 illustrates an $\alpha = 5\%$ VaR for a random portfolio of normally distributed variables with a mean of 1 and standard deviation of 1.7. The vertical line marks the $VaR_{0.05}$ or equivalently as mentioned above the 5% quantile.

1. Introduction

1.2.2. Conditional Value-at-Risk

Conditional-Value-at-Risk (CVaR) also known as Expected-Shortfall (ES) is closely linked to Value-at-Risk (VaR). From now on we will refer to it as CVaR.

Definition 4 (Conditional-Value-at-Risk)

Given some confidence level $\alpha \in (0, 1)$ and a continuous portfolio distribution X with distribution function F_X , the $CVaR_\alpha(X)$ is defined as:

$$CVaR_\alpha(X) := ES_\alpha(X) = \mathbb{E}[X|X \leq VaR_\alpha(X)]$$

This can equivalently be written as

$$CVaR_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_p(X) dp.$$

In other words, the "CVaR at level $\alpha\%$ " gives us the expected return in the worst $\alpha\%$ of all cases. It is considered to be a better alternative to VaR, because it takes into account the shape of the left tail of the loss distribution. Pflug (2000) proved that CVaR satisfies all four conditions for a coherent risk measure: translation invariance, subadditivity, positive homogeneity and monotonicity. In contrast to VaR, CVaR therefore is a coherent risk measure.

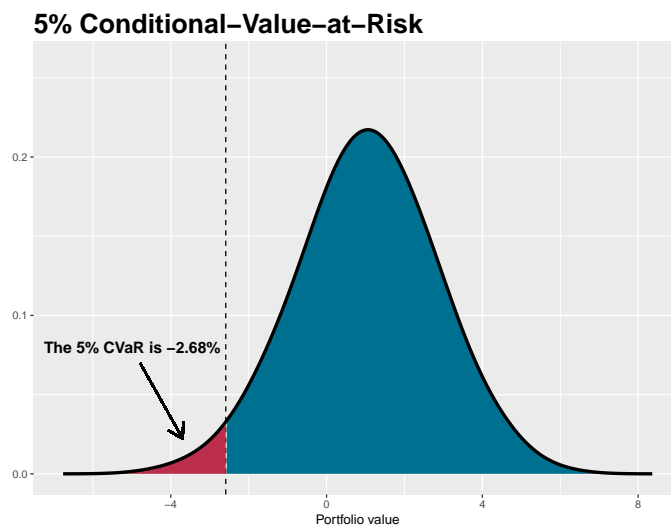


Figure 1.2.: Illustration of $CVaR_{0.05}$.

1. Introduction

Figure 1.2 illustrates an $\alpha = 5\%$ CVaR for a random portfolio of normally distributed variables with a mean of 1 and standard deviation of 1.7.

The CVaR satisfies the following properties:

- $CVaR$ is a coherent risk measure.
- $CVaR_{1.0}$ equals the expected value of the entire portfolio.
- As α increases the $CVaR_\alpha$ increases.
- For any given portfolio we get $CVaR_\alpha \leq VaR_\alpha$ at the same level α .

Historical overview over the most important developments concerning VaR and CVaR:

- 1994: RiskMetrics™ Technical Documents make VaR popular.
- 1996: Basel Committee on Banking Supervision (1996) internal-based approach to capital adequacy - based on VaR.
- 1999: Artzner, Delbaen, and Ebner (1999) develops axioms for sensible risk measures. VaR is criticized for not complying.
- 2000: Pflug (2000) proved that CVaR satisfies all conditions for a coherent risk measure.
- 2000: VaR and CVaR are widely adopted in the financial industry.
- 2013/2016: Basel Committee on Banking Supervision (2013) replaces $VaR_{0.01}$ with $CVaR_{0.025}$. VaR remains important for model backtesting.

1.3. Traditional methods for calculating VaR and CVaR

There are several different ways of calculating VaR. In this section we give a brief introduction to the two most commonly used methods: the historical simulation and the mean variance method. On the basis of VaR we will later be able to further calculate the CVaR of a given multidimensional portfolio. The general outline of this subsection is modeled after works from Xu and Chen (2012).

1. Introduction

1.3.1. Historical simulation method (HS)

One of the industry standards for computing VaR is the so-called historical simulation (HS). Under this model the forecasting of VaR is solely based on the empirical loss distribution. On this way it is possible to estimate future VaR directly from observed portfolio returns. Hence there is no need for modeling the loss distribution under some statistical model. VaR of a n -dimensional portfolio can be computed in two simple steps.

1. We first need to transform the asset returns into pseudohistorical portfolio returns. Using today's portfolio weights and historical asset returns of the past t days, a series of pseudohistorical portfolio returns, is constructed:

$$R_p^i = \sum_{j=1}^n w_j R_a^{i,j} \text{ for } 1 \leq i \leq t,$$

where R_p^i denotes the pseudohistorical portfolio return at time i , w_j the portfolio weight of financial assets j and $R_a^{i,j}$ denotes the historical asset return of the single portfolio component j at time i . The portfolio weights $w = (w_1, w_2, \dots, w_n)$ give the proportion invested in each individual asset and sum to one.

2. The α -quantile of the pseudohistorical portfolio returns is being calculated, which leads to the empirical VaR_α at time $t + 1$:

$$\widehat{VaR}_\alpha^{t+1} = q_\alpha(R_p^1, R_p^2, \dots, R_p^t)$$

Advocates of the HS model to calculate VaR highlight its *model-free* nature. Nevertheless it is certainly not *assumption-free*. Historical simulation implicitly assumes that the returns are independent and identically distributed, which is unfortunately empirically not the case. This might be very misleading, especially in the case when a period of market turbulence appears after a long period of low volatility. The obvious attractions of the HS method are that it is quite easy to implement and does not depend on any assumptions of the loss distribution.

One should also be very careful when setting the length of the timeframe of historic data used for the estimation. A too long window might lead to a VaR that is based on too old historical data which may not correctly

1. Introduction

represent the current situation on the market anymore and in addition is slow to adjust to changes in volatility. On the other hand a too short window might lead to calculations that are not sufficiently robust.

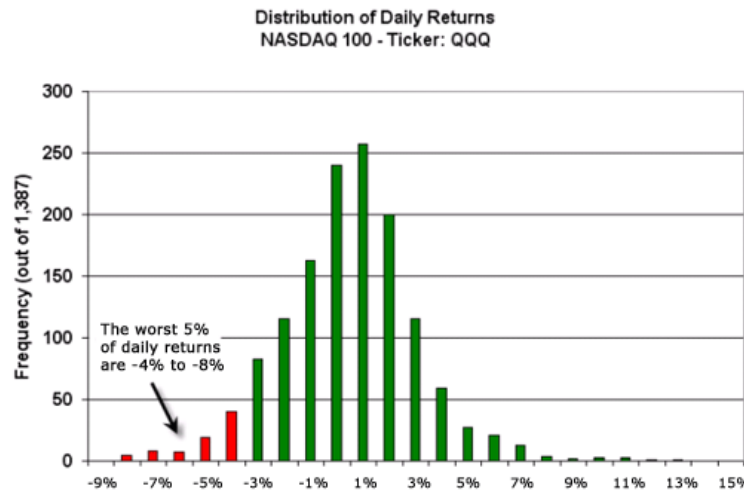


Figure 1.3.: Illustration of historical simulation method, from Harper (2017).

Figure 1.3 illustrates an $\alpha = 5\%$ VaR of daily returns of the NASDAQ 100 (ticker: QQQ) within a period of something more than five years by using a histogram of daily returns. Therefore the red bars on the left of the histogram are indicating the worst 5% of daily returns. In this example we can say with a certainty of 95% that the worst daily loss will not exceed 4%.

This approach struggles from the co-called ghost effects, which refers to the effect of a VaR estimation being unduly high (low) due to a short period of high (low) volatility. The VaR estimation stays high (low) until these observations have fallen out of the sample window. To reduce this Boudoukh, Richardson, and Whitelaw (1997) suggested using weights for each observation according to their age, instead of using equal weights for all historic observations. This makes the VaR estimates more responsive to a large loss observation and also helps with volatility clustering (e.g. clusters of large losses).

1. Introduction

1.3.2. Mean variance method (MV)

This method, used by the RiskMetricsTM methodology, was originally developed by JP Morgan. It assumes that the portfolio returns are multivariate normally distributed with mean vector μ and covariance matrix Σ . This allows us to plot a normal distribution curve against the actual return data as seen below.

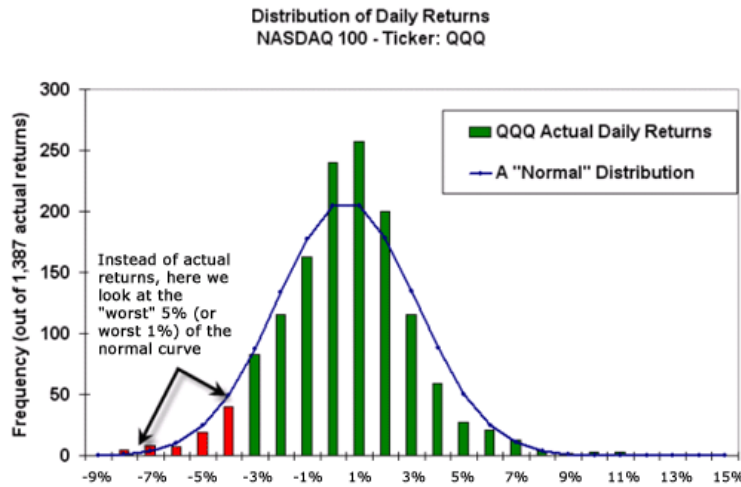


Figure 1.4.: Illustration of mean variance method, from Harper (2017).

The idea behind this method is very similar to the historical simulation, except that we are now using a familiar curve instead of actual market data. On this normal curve it is quite trivial to calculate the worst $\alpha\%$ that we need for VaR estimation. The VaR can be computed in three simple steps:

1. We need to estimate two factors from the portfolio of assets: The vector of expected returns μ and the covariance matrix Σ ,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}.$$

1. Introduction

The mean vector μ represents the average historical returns for each asset, whereas the covariance matrix Σ depicts the interdependencies between each asset. One way to estimate the mean μ_i of the individual asset i is

$$\hat{\mu}_i = \frac{1}{t} \sum_{k=1}^t R_a^{k,i} \text{ for } 1 \leq i \leq n$$

whereas the estimator $\hat{\Sigma} = (\hat{\sigma}_{ij})$ for $\Sigma = (\sigma_{ij})$ is given by:

$$\hat{\sigma}_{ij} = \frac{1}{t-1} \sum_{k=1}^t (R_a^{k,i} - \hat{\mu}_i)(R_a^{k,j} - \hat{\mu}_j) \text{ for } 1 \leq i, j \leq n.$$

- Given a vector of portfolio weights $w = (w_1, w_2, \dots, w_n)$, the estimators of weighted expected portfolio return $\hat{\mu}_p$ and the weighted return variance $\hat{\sigma}_p^2$ are given by:

$$\begin{aligned} \hat{\mu}_p &= w\hat{\mu} \\ \hat{\sigma}_p^2 &= w\hat{\Sigma}w' \end{aligned}$$

- Under the assumption of multivariate normal distribution the estimator \widehat{VaR}_α at time $t + 1$ can now be calculated:

$$\widehat{VaR}_\alpha^{t+1} = \hat{\mu}_p + \hat{\sigma}_p \phi^{-1}(\alpha),$$

where ϕ denotes the cumulative distribution function of a standard normally distributed random variable.

One major drawback of assuming normal distribution for VaR calculation is that portfolio asset are, in the best case scenario, only approximately normally distributed. In the worst case asset values could follow a totally different distribution.

1.4. Methods of backtesting VaR and CVaR

Backtesting procedures are essential tools for determining the accuracy of VaR and CVaR models. In the context of financial risk modeling, the main idea of backtesting is to compare the calculated VaR_α^t measures to the actual observed gains (or losses) x_t realized on the portfolio for $1 \leq t \leq n$, to assess the performance of the used VaR/CVaR model. In practice, there are several different metrics and statistical tests¹ to evaluate the performance of risk measures, whereof we are using a widely used test introduced by P. F. Christoffersen (1998) and Kupiec (1995) for VaR and the Zero Mean test proposed by A. J. McNeil and Frey (2000) for CVaR.

1.4.1. VaR: Kupiec and Christoffersen coverage test

VaR_α^t is defined in such a way that the probability of realizing a loss in excess of the reported VaR_α^t (also known as a violation) is matching the implied probability α from the VaR confidence level. $\{\mathbb{1}_t\}_{t=1}^n$ is defined as a indicator sequence across n days indicating if such a violation at day t occurred or not:

$$\mathbb{1}_t = \begin{cases} 1, & \text{if } x_t < VaR_\alpha^t \\ 0, & \text{if } x_t \geq VaR_\alpha^t, \end{cases} \quad (1.1)$$

where x_t is denoting the actual gain (or loss) realized on the portfolio, whereas VaR_α^t denotes the calculated $\alpha\%$ VaR value for day t . This series of indicators, which is often referred to as hitting sequence, returns 1 on day t if the actual return on that day is lower than the VaR forecast for that day. Vice versa, if the forecasted VaR for day t is smaller than the observed return, then the hitting sequence returns 0. According to P. F. Christoffersen (1998), the problem of determining the quality of a VaR model can be reduced to the question whether the hitting sequence $\{\mathbb{1}_t\}_{t=1}^n$ fulfills two properties:

1. *Unconditional coverage property:*

The likelihood of realizing a loss in excess of the reported VaR_α^t (also

¹See Campbell (2005) and Berkowitz, P. Christoffersen, and Pelletier (2011) for further details on several different backtesting procedures.

1. Introduction

known as violations) must match the implied probability α from the VaR confidence level. Formally

$$\mathbb{P}(\mathbb{1}_t = 1) = \alpha \text{ for } 1 \leq t \leq n, \quad (1.2)$$

makes sure that the probability of violation during the observed period equals the α level of the VaR estimation.

2. Independence property:

Up to this point we have only placed restrictions on how often a VaR violation may occur. The independence property states additionally restrictions regarding the way in which the violations may occur. The VaR violations observed on two different observations for the same coverage rate of α must be independently distributed. This means that a variable $\mathbb{1}_t$ must be independent of the variable $\mathbb{1}_{t-j}, \forall j \neq 0$. Hence past unconditional coverage violations should not carry any information about possible future violations.

Christoffersen points out that VaR forecasts are only valid if both properties are met, this meaning that the hitting sequence satisfies the unconditional coverage hypothesis (UC) and the independence hypothesis (IND). In other words, the hitting sequence should be completely unpredictable and hence independently distributed over time, following an i.i.d. Bernoulli distribution that assumes 1 with probability α and 0 with probability $(1 - \alpha)$. This can be written as:

$$H_0 : \mathbb{1}_t \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(\alpha)$$

The probability distribution function of a Bernoulli distribution with $\alpha \in (0, 1)$ is given as:

$$f(\mathbb{1}_t | \alpha) = (1 - \alpha)^{1 - \mathbb{1}_t} \alpha^{\mathbb{1}_t}$$

Overall the Christoffersen frequency test consists of two separate tests which can be joined together to a conditional coverage test (CC).

The likelihood ratio test of unconditional coverage (UC):

The unconditional coverage test evaluates the unconditional coverage properties of a series of VaR_α^t forecasts for $1 \leq t \leq n$.

We want to test if the fraction of violations obtained for a certain observed risk model (we will call it π) is significantly different from the expected

1. Introduction

fraction α . Hence we set the null and alternative hypothesis as:

$$\begin{aligned} H_0 : \mathbb{1}_t &\overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\alpha) \\ H_1 : \mathbb{1}_t &\overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\pi) \end{aligned}$$

The likelihood under H_0 is simply the likelihood of an i.i.d. Bernoulli(α) hitting sequence

$$L(\alpha | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) = \prod_{t=1}^n (1 - \alpha)^{1 - \mathbb{1}_t} \alpha^{\mathbb{1}_t} = (1 - \alpha)^{n_0} \alpha^{n_1},$$

and under the alternative hypothesis H_1

$$L(\pi | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) = \prod_{t=1}^n (1 - \pi)^{1 - \mathbb{1}_t} \pi^{\mathbb{1}_t} = (1 - \pi)^{n_0} \pi^{n_1},$$

where n is the total number of VaR forecasts in the series and n_0 and n_1 the number of non-violations and violations in the sample. π can easily be estimated from the observed fraction of violation in the sequence using maximum likelihood estimation:

$$\hat{\pi} = \frac{n_1}{n_0 + n_1} = \frac{n_1}{n}$$

Detailed calculations can be found in Appendix A.2.1.

The test statistic can now be formulated as a standard likelihood ratio test²:

$$\begin{aligned} LR_{uc} &= -2 \log \left[\frac{L(\alpha | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)}{L(\hat{\pi} | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)} \right] \\ &= -2 \log \left[\frac{(1 - \alpha)^{n_0} \alpha^{n_1}}{\left(1 - \frac{n_1}{n}\right)^{n_0} \left(\frac{n_1}{n}\right)^{n_1}} \right] \end{aligned}$$

The likelihood ratio test compares two nested models, hence the simpler model in the numerator with fewer parameters is a special case of the more complex model in the denominator consisting of more parameters. We now

²A short introduction to the likelihood ratio test can be found in Appendix A.1.

1. Introduction

use Wilks' Theorem introduced in Wilks (1938). The theorem states that the distribution of the statistic $-2 \log \left[\frac{L(\theta_0|x)}{L(\theta|x)} \right]$ for a nested model converges to a χ^2_ν distribution as the sample size $n \rightarrow \infty$, given H_0 is true. The degrees of freedom ν of the χ^2_ν distribution are equal to the difference in dimension of the parameter space of θ and θ_0 .

With this theorem in hand (and for large n) we can now asymptotically compare LR_{uc} to the χ^2_ν distributions with $\nu = (2 - 1)$ degrees of freedom.

$$LR_{uc} \stackrel{\text{asy}}{\sim} \chi^2_{2-1} = \chi^2_1$$

Example for a significance level of $\beta = 5\%$:

As the LR_{uc} value gets larger the more unlikely H_0 is to be true. Under the significance level of $\beta = 5\%$, we will have a critical value of 3.841 from the χ^2_1 distribution. Hence the null hypothesis is rejected if $LR_{uc} > LR_{critical} = 3.841$.

In another approach we can calculate the p-value associated with the likelihood ratio test statistic, which is defined as the probability, given the null hypothesis H_0 is true, of obtaining a sample that conforms even less to the null hypothesis H_0 than the sample we actually got.

In our particular setting the p-value is calculated as:

$$\text{p-value} : 1 - F_{\chi^2_1}(LR_{uc})$$

$F_{\chi^2_1}(\cdot)$ is denoting the cumulative density function of a χ^2 variable using one degree of freedom. The null hypothesis is rejected if the p-value is below the significance level β . If we now presume a test value of $LR_{uc} = 4.1$, then we obtain a p-value of:

$$\text{p-value} : 1 - F_{\chi^2_1}(4.1) = 1 - 0.9571 = 0.0429$$

Hence, when using a significance level of $\beta = 5\%$, we can reject the null hypothesis.

Nevertheless simply testing for correct unconditional coverage is insufficient. The UC hypothesis gives no indication whether the violations come clustered together in a time-dependent fashion. The clustering of VaR violations

1. Introduction

can be a serious warning signal of risk model misspecification. This often happens when using VaR forecasting models that are slow at responding to volatility increases in the market like e.g. historical simulation (HS). We need one more test to make up for this deficiency.

The likelihood ratio test of independence (IND):

Our goal is to construct a statistical test that is able to reject a series of VaR estimates that exhibit clustered violations. The VaR violations observed on two different observations for the same coverage rate of $\alpha\%$ must be independently distributed. This means that a variable $\mathbb{1}_t$ (indicating a VaR violation at time t) must be independent of the variable $\mathbb{1}_{t-j}, \forall j \neq 0$. Hence past UC violations should not carry any information about possible future violations. We now consider a first-order Markov chain $\{\mathbb{1}_t\}_{t=1}^n$ with transition matrix

$$\Pi_1 = \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix},$$

where

$$\pi_{ij} = \mathbb{P}(\mathbb{1}_t = j | \mathbb{1}_{t-1} = i).$$

Here we assume that only today's outcome has relevance for tomorrow's outcome. The likelihood function for this first-order Markov process is

$$L(\Pi_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}, \quad (1.3)$$

where n_{ij} for $i, j \in \{0, 1\}$ refers to the number of observations where j follows i . We can solve for the maximum likelihood estimates by taking the first derivatives with respect to π_{01} and π_{11} and setting these derivatives to zero:

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}},$$

$$\hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}.$$

Detailed calculations can be found in Appendix A.3.1. The probabilities need to sum up to one, hence we get:

$$\hat{\Pi}_1 = \begin{pmatrix} \hat{\pi}_{00} & \hat{\pi}_{01} \\ \hat{\pi}_{10} & \hat{\pi}_{11} \end{pmatrix} = \begin{pmatrix} 1 - \hat{\pi}_{01} & \hat{\pi}_{01} \\ 1 - \hat{\pi}_{11} & \hat{\pi}_{11} \end{pmatrix} = \begin{pmatrix} \frac{n_{00}}{n_{00} + n_{01}} & \frac{n_{01}}{n_{00} + n_{01}} \\ \frac{n_{10}}{n_{10} + n_{11}} & \frac{n_{11}}{n_{10} + n_{11}} \end{pmatrix}$$

1. Introduction

Dependence in the hitting sequence corresponds to allowing π_{01} and π_{11} to be different. In general, we are mostly concerned about positive dependence, meaning that it is more likely that a violation is following a violation (π_{11}) than a non-violation (π_{01}). However, in the case of independent violations over time, the probabilities of a violation tomorrow does not depend on the occurrence of a violation today. Therefore we set the null hypothesis as:

H_0 : The occurrences of violations and non-violations in the hitting sequence $\{\mathbb{1}_t\}_{t=1}^n$ are independently distributed over time, hence $\pi_{01} = \pi_{11} = \pi$.

This leads to the transition matrix

$$\Pi = \begin{pmatrix} 1 - \pi & \pi \\ 1 - \pi & \pi \end{pmatrix}.$$

Under the null hypothesis the likelihood function becomes

$$L(\Pi | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) = (1 - \pi)^{(n_{00} + n_{10})} \pi^{(n_{01} + n_{11})},$$

and the maximum likelihood estimate is

$$\hat{\pi} = \frac{n_{01} + n_{11}}{n_{00} + n_{10} + n_{01} + n_{11}} = \frac{n_1}{n}.$$

Detailed calculations can be found in Appendix A.3.2. We set the alternative hypothesis as:

H_1 : The occurrences of violations and non-violations in the hitting sequence $\{\mathbb{1}_t\}_{t=1}^n$ are not independently distributed over time.

Under the alternative hypothesis the likelihood function is given by Equation (1.3).

The following likelihood ratio test tests the independence hypothesis that $\pi_{01} = \pi_{11}$:

$$\begin{aligned} LR_{ind} &= -2 \log \left[\frac{L(\hat{\Pi} | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)}{L(\hat{\Pi}_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)} \right] \\ &= -2 \log \left[\frac{(1 - \hat{\pi})^{(n_{00} + n_{10})} \hat{\pi}^{(n_{01} + n_{11})}}{(1 - \hat{\pi}_{01})^{n_{00}} \hat{\pi}_{01}^{n_{01}} (1 - \hat{\pi}_{11})^{n_{10}} \hat{\pi}_{11}^{n_{11}}} \right] \stackrel{\text{asy}}{\sim} \chi_1^2 \end{aligned}$$

1. Introduction

In the practical implementation of LR_{ind} one might find samples where $n_{11} = 0$ (the existence of no instances where violations follow directly onto violations), which leads to numerical problems.

Example for a significance level of $\beta = 5\%$:

The larger the LR_{ind} value is the more unlikely the independence hypothesis is to be true. Under the significance level of $\beta = 5\%$, we will have a critical value of 3.841 from the χ_1^2 distribution. Hence the independence hypothesis is rejected if $LR_{ind} > LR_{critical} = 3.841$.

Alternatively the null hypothesis is rejected if the p-value is below the significance level β . If we now presume a test value of $LR_{ind} = 4.1$, then we obtain a p-value of:

$$\text{p-value} : 1 - F_{\chi_1^2}(4.1) = 1 - 0.9571 = 0.0429$$

Hence, when using a significance level of $\beta = 5\%$, we can reject the null hypothesis.

This test does not depend on the true coverage rate α and therefore it is only a test for the independence part of our hypothesis.

The likelihood ratio test of conditional coverage (CC):

An accurate *VaR* measure must satisfy both the unconditional coverage (UC) and independence (IND) property. The conditional coverage (CC) test forms a complete test by jointly testing for the unconditional coverage (UC) and the independence (IND) property. The previously discussed independence test (IND) examined whether the likelihood of a violation following on a previous violation is equal to the likelihood of a violation following on a previous non-violation: $\pi_{01} = \pi_{11}$. If we additionally require the VaR estimate to exhibit the unconditional coverage (UC) property, then the above likelihoods need to correspond with the total proportion of violation α , hence: $\pi_{01} = \pi_{11} = \alpha$.

Accordingly, the joint conditional coverage (CC) test determines if there is a difference in likelihood of a VaR violation following a VaR violation or a non-violation and examines whether the probability is significantly different from α . Therefore we set the null hypothesis as:

$$H_0 : \pi_{01} = \pi_{11} = \alpha$$

1. Introduction

The conditional coverage test statistic is given as:

$$LR_{cc} = -2 \log \left[\frac{L(\alpha | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)}{L(\hat{\Pi}_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)} \right] \stackrel{\text{asy}}{\sim} \chi_2^2$$

The likelihood ratio test statistic for the unconditional coverage test (CC) is asymptotically χ_2^2 distributed as seen in Equation (1.4).

The unconditional coverage test (UC), the independence test (IND) and the conditional coverage test (CC) are related by the following identity:

$$\begin{aligned} LR_{cc} &= -2 \log \left[\frac{L(\alpha | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)}{L(\hat{\Pi}_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)} \right] \\ &= -2 \log \left[\frac{L(\alpha | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)}{L(\hat{\pi} | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)} \cdot \frac{L(\hat{\Pi}_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)}{L(\hat{\Pi}_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)} \right] \\ &= -2 \log \left[\frac{L(\alpha | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)}{L(\hat{\Pi}_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)} \right] - 2 \log \left[\frac{L(\hat{\Pi}_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)}{L(\hat{\Pi}_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n)} \right] \\ &= LR_{uc} + LR_{ind} \stackrel{\text{asy}}{\sim} \chi_2^2 \end{aligned} \quad (1.4)$$

Example for a significance level of $\beta = 5\%$:

Under significance level of $\beta = 5\%$, we will have a critical value of 5.991 from the χ_2^2 distribution. The VaR model passes the conditional coverage test if $LR_{cc} \leq LR_{critical} = 5.991$.

Alternatively the null hypothesis is rejected if the p-value is below the significance level β . If we now presume a test value of $LR_{cc} = 8.2$, then we obtain a p-value of:

$$\text{p-value} : 1 - F_{\chi_2^2}(8.2) = 1 - 0.9834 = 0.0166$$

Hence, when using a significance level of $\beta = 5\%$, we can reject the null hypothesis.

1. Introduction

1.4.2. CVaR: Zero mean test

According to the definition of Conditional-Value-at-Risk given in Artzner, Delbaen, and Ebner (1999), CVaR is the conditional expectation of loss, given the loss of an portfolio of assets is greater than the forecasted VaR. CVaR is a risk measure for forecasting a specific loss on the basis of VaR. Therefore we cannot discuss the CVaR forecast for models which are not even suitable to properly forecast VaR of a portfolio. A. J. McNeil and Frey (2000) have proposed a backtesting procedure for CVaR using the zero mean test. We observe that $CVaR_\alpha$ can be written as

$$CVaR_\alpha = VaR_\alpha + \underbrace{(CVaR_\alpha - VaR_\alpha)}_{\text{Excess Shortfall}}.$$

Given that the VaR estimation passed the previously in Subsection 1.4.1 described tests, the excess shortfall part can be backtested using the following test statistic:

$$S_t = (CVaR_\alpha^t - x_t) \mathbb{1}_{\{x_t < VaR_\alpha^t\}} \text{ for } 1 \leq t \leq n,$$

where x_t denotes the actual gains (losses) realized on the portfolio, whereas VaR_α^t and $CVaR_\alpha^t$ denote the calculated risk measures at a $\alpha\%$ level for day t . A. J. McNeil and Frey (2000) pointed out that S_t should be i.i.d. distributed over time with a mean of 0, in the case that the CVaR is forecasted correctly. Therefore the null hypothesis H_0 states that the excess conditional shortfall (excess of the actual series when VaR is violated), is i.i.d. and has a mean of 0. We can conduct a standard one-sided t-test against the alternative hypothesis H_1 that the excess shortfall has mean greater than 0, meaning that the CVaR is systemically underestimated, since this presents the most likely direction of failure.

1.5. Copula

Definition 5 (Copula)

A n -dimensional copula C is a multivariate cumulative distribution function on $[0, 1]^n$,

$$C : [0, 1]^n \rightarrow [0, 1],$$

with univariate uniform distribution margins.

In other terms, $C : [0, 1]^n \rightarrow [0, 1]$ is a n -dimensional copula if:

- $C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_n) = 0$ for all $i \in \{1, \dots, n\}$, $u_i \in [0, 1]$. If one of the arguments is zero then the copula is equal to zero.
- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ for all $i \in \{1, \dots, n\}$, $u_i \in [0, 1]$. If one of the arguments of the copula is u_i and all others are 1 then the copula is equal to u_i .
- C is d -non-decreasing, i.e. for each hyperrectangle

$$R = \prod_{i=1}^n [x_i, y_i] \subseteq [0, 1]^n \text{ with } 0 \leq x_i \leq y_i \leq 1, \text{ for all } i \in \{1, \dots, n\}$$

the C -volume of R is non-negative:

$$V_C(R) := \int_R dC(u) = \sum_{\mathbf{z} \in \prod_{i=1}^n \{x_i, y_i\}} (-1)^{N(\mathbf{z})} C(\mathbf{z}) \geq 0,$$

where the $N(\mathbf{z}) = \#\{i : z_i = x_i\}$.

In the 2-dimensional case, $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a bivariate copula if

- $C(0, u) = C(u, 0) = 0$, for all $u \in [0, 1]$
- $C(1, u) = C(u, 1) = u$, for all $u \in [0, 1]$
- C is 2-non-decreasing:
 $V_C([u_1, u_2] \times [v_1, v_2]) = C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ for all $0 \leq u_1 \leq u_2 \leq 1$ and $0 \leq v_1 \leq v_2 \leq 1$.

For further details on copulas and bivariate copula families see Chapter 3.

1.6. Dependence measures

In this section we will briefly introduce and describe the dependence measures Pearson's ρ , Spearman's ρ and Kendall's τ . The following subsection is based on Embrechts, Lindskog, and A. McNeil (2003) Chapter 5.

Definition 6 (Concordant / Discordant)

Let $(x_1, x_2)^T$ and $(\tilde{x}_1, \tilde{x}_2)^T$ be two observations from a vector $(X_1, X_2)^T$ of continuous random variables. $(x_1, x_2)^T$ and $(\tilde{x}_1, \tilde{x}_2)^T$ are called concordant if

$$(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0.$$

Similarly $(x_1, x_2)^T$ and $(\tilde{x}_1, \tilde{x}_2)^T$ are called discordant if

$$(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0.$$

Two important measures of concordance we will look at in this section are Kendall's τ and Spearman's ρ . Further details on dependence measures can be found in Joe (1997).

1.6.1. Pearson correlation coefficient

The Pearson correlation coefficient is a measure of linear dependence of two variables X_1 and X_2 .

Definition 7 (Pearson correlation coefficient / Linear correlation coefficient)

Let $\text{Var}(X_1), \text{Var}(X_2) \in (0, \infty)$. The Pearson correlation coefficient also known as linear correlation coefficient $\rho_L(X_1, X_2) \in [-1, 1]$ is defined as

$$\rho_L(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\sigma_{X_1}^2 \sigma_{X_2}^2}},$$

where $\text{Cov}(\cdot)$ is the covariance, σ_{X_1} the standard deviation of X_1 and σ_{X_2} the standard deviation of X_2 .

This kind of linear correlation coefficient has some limitations and therefore has to face some criticism:

1. Introduction

- It measures linear dependence, which can be a drawback in cases where dependence is non-linear.
- Zero correlation in general does not imply independence.
- It is sensitive to outliers.
- May be misleading if the multivariate distribution function is not elliptical.
- The Pearson correlation coefficient is not defined for variables with non-finite variance.
- It is not invariant under *non-linear* strictly increasing transformations (e.g. X_1 and X_2 do not yield the same correlation as $\log(X_1)$ and $\log(X_2)$). Invariance only holds in case of strictly increasing *linear* transformations. Hence for two random variables X_1 and X_2 and real constants $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}, \beta_1 > 0$ and $\beta_2 > 0$ we have:

$$\rho_L(\alpha_1 + \beta_1 X_1, \alpha_2 + \beta_2 X_2) = \rho_L(X_1, X_2)$$

Copulas on the other hand are invariant under *non-linear* continuous and increasing transformation of the margins. It is common practice in the analysis of financial risk factors to use log-returns on asset and commodity prices, which may present a problem when using linear correlation. So it is apparent that we will now focus on copula-based dependence measures as alternative to linear correlation based measures.

1.6.2. Spearman's ρ

Definition 8 (Spearman's ρ)

Let F be a continuous bivariate cumulative distribution function and let $(X_1, X_2)^T, (\tilde{X}_1, \tilde{X}_2)^T$ and $(\check{X}_1, \check{X}_2)$ be independent random pairs with distribution F . Then formally Spearman's ρ denoted by $\rho_S(X_1, X_2) \in [0, 1]$ is defined as:

$$\rho_S(X_1, X_2) = 3 (\mathbb{P}((X_1 - \tilde{X}_1)(X_2 - \check{X}_2) > 0) - \mathbb{P}((X_1 - \tilde{X}_1)(X_2 - \check{X}_2) < 0))$$

Equivalently, let F_1 and F_2 be the continuous margins of $(X_1, X_2)^T$, then Spearman's ρ is the Pearson correlation of $F_1(X_1)$ and $F_2(X_2)$ ($\rho_S(X_1, X_2) = \rho_L(F_1(X_1), F_2(X_2))$). In other words Spearman's ρ is the linear correlation of the unique copula of $(X_1, X_2)^T$.

1. Introduction

Theorem 1

Let $(X_1, X_2)^T$ be a random vector with continuous margins and a unique copula C , then we can write Spearman's ρ as:

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2 \quad (1.5)$$

$$= 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3 \quad (1.6)$$

Proof. The proof to this theorem can be found in A. J. McNeil, Frey, and Embrechts (2006), Proposition 5.29. \square

1.6.3. Kendall's τ

Definition 9 (Kendall's τ)

Let F be a continuous bivariate cumulative distribution function and let $(X_1, X_2)^T$, $(\tilde{X}_1, \tilde{X}_2)^T$ be independent random pairs with distribution F . Then formally Kendall's τ denoted by $\rho_\tau(X_1, X_2) \in [0, 1]$ is defined as:

$$\rho_\tau(X_1, X_2) = \mathbb{P}((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0) - \mathbb{P}((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0)$$

Theorem 2

Let $(X_1, X_2)^T$ be a random vector with continuous margins and a unique copula C , then we can write Kendall's τ as:

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1 \quad (1.7)$$

Proof. The proof to this theorem can be found in A. J. McNeil, Frey, and Embrechts (2006), Proposition 5.29. \square

We can interpret Kendall's τ as the difference of probability of two random concordant pairs and the probability of two random discordant pairs.

Spearman's ρ and Kendall's τ are rank correlations, which do not directly depend on the particular distribution of the observed data but on the copula of the bivariate random vector. Therefore they are invariant under strictly monotone transformation.

1. Introduction

1.6.4. Tail dependence

This concept refers to the degree of dependence in the lower-left quadrant and the upper-right quadrant of a bivariate distribution and is therefore important for measuring dependence between extreme events.

Definition 10 (Upper tail dependence)

Let $(X_1, X_2)^T$ be a random vector with marginal distribution F_1 and F_2 . The coefficient of upper tail dependence of $(X_1, X_2)^T$ is defined as:

$$\lambda_U(X_1, X_2) = \lim_{u \rightarrow 1^-} \mathbb{P}(X_2 > F_2^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)),$$

given that the limit exists.

Definition 11 (Lower tail dependence)

Let $(X_1, X_2)^T$ be a random vector with marginal distribution F_1 and F_2 . The coefficient of lower tail dependence of $(X_1, X_2)^T$ is defined as:

$$\lambda_L(X_1, X_2) = \lim_{u \rightarrow 0^+} \mathbb{P}(X_2 \leq F_2^{\leftarrow}(u) \mid X_1 \leq F_1^{\leftarrow}(u)),$$

given that the limit exists.

If $\lambda_U > 0$ ($\lambda_L > 0$) we say that $(X_1, X_2)^T$ has a upper (lower) tail dependence.

Theorem 3

Let $(X_1, X_2)^T$ be a random vector with continuous marginal distribution and C a copula. If the limits exist we get

$$\lambda_U(X_1, X_2) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}$$
$$\lambda_L(X_1, X_2) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}.$$

Proof. The proof to this theorem can be found in Dragoti-Çela (2009). \square

If the copula in question has a simple closed form, the calculation of these coefficients is usually straightforward as we will see in Chapter 3.

2. Univariate GARCH model

An important concept in financial modeling that measures the movement in market prices is volatility. Time series of financial asset returns often tend to volatility clustering. This is the tendency of large price changes of financial assets to cluster together, meaning that the variance of log-prices is high for extended periods and then low for extended periods.

This phenomenon has been of great interest to researchers and has influenced the development of several stochastic models in finance. While many econometric models take the assumption of constant variance, the *Autoregressive Conditional Heteroscedasticity* (ARCH) model introduced by Engel (1982) models the conditional variance as a function of past errors, hence allowing it to change over time.

Since the development of the ARCH model there have been several extensions in order to improve the model, one of which is the *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) model by Bollerslev (1986). The GARCH process is often used by financial professionals because it provides good real-world properties when trying to estimate prices and rates of financial instruments. In our context we will mainly focus on the univariate GARCH(1,1) model. Parts of this chapter are based on works from Orskaug (2009), Dißmann (2010) and Brechmann (2010).

Definition 12 (Autocovariance function)

If $(X_t)_{t \in \mathbb{Z}}$ is a stochastic process with mean $\mathbb{E}[X_t^2] < \infty, \forall t \in \mathbb{Z}$, then the autocovariance function γ_X of $(X_t)_{t \in \mathbb{Z}}$ is defined as

$$\begin{aligned}\gamma_X(r, s) &= \text{Cov}(X_r, X_s) \\ &= \mathbb{E}[(X_r - \mu_r)(X_s - \mu_s)],\end{aligned}$$

with $r, s \in \mathbb{Z}$ and $\mu_t = \mathbb{E}[X_t]$.

2. Univariate GARCH model

Definition 13 (Stationarity¹)

The time series $(X_t)_{t \in \mathbb{Z}}$ is called stationary if it satisfies the following:

- $\mathbb{E}[X_t^2] < \infty \forall t \in \mathbb{Z}$
- $\mathbb{E}[X_t] = \mu \in \mathbb{R}, \forall t \in \mathbb{Z}$ and
- $\gamma_X(r, s) = \gamma_X(r + t, s + t) \forall r, s, t \in \mathbb{Z}$

Obviously, for a stationary process the autocovariance function as defined in Definition 12 only depends on $|r - s|$, since $\gamma_X(r, s) = \gamma_X(r - s, 0)$. Hence we can redefine the autocovariance function for a stationary process as:

$$\gamma_X(h) := \gamma_X(h, 0) = \text{Cov}(X_{t+h}, X_t), \forall t, h \in \mathbb{Z}.$$

Definition 14 (Autocorrelation function (ACF))

The autocorrelation function of a stochastic process with mean $\mathbb{E}[X_t^2] < \infty, \forall t \in \mathbb{Z}$, is the normed autocovariance function:

$$\rho_X(r, s) = \frac{\gamma_X(r, s)}{\sqrt{\text{Var}(X_r)\text{Var}(X_s)'}}$$

where $-1 \leq \rho_X(r, s) \leq 1$.

As already stated above for a stationary process the autocovariance function only depends on the difference in time of r and s . Therefore the standard deviation is independent of time and the product of the standard deviations in the denominator corresponds to $\sqrt{\text{Var}X_t^2} = \text{Var}(X_t) = \text{Cov}(X_t, X_t) = \gamma_X(0)$, independent of t . Therefore the autocorrelation function of a stationary process simplifies to

$$\rho_X(h) := \rho_X(h, 0) = \frac{\text{Cov}(X_t, X_{t-h})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-h})}} = \frac{\gamma_X(h)}{\gamma_X(0)}, \forall t, h \in \mathbb{Z},$$

where $h \geq 1$ denotes the lag and $\rho(h) = \rho(-h)$.

Autocorrelation is a representation of the degree of similarity between a given time series and a lagged version of itself over successive time intervals.

¹In some literature the definition of stationarity that is used here is referred to as weak stationarity.

2. Univariate GARCH model

It is the same as calculating the correlation between two different time series, except that the same time series is used twice: once in its original form and once lagged one or more time periods. Large values for an extended amount of lags in the autocorrelation function imply volatility clustering. Figure 2.1 displays two examples of ACF plots, one with volatility clustering and one without volatility clustering.

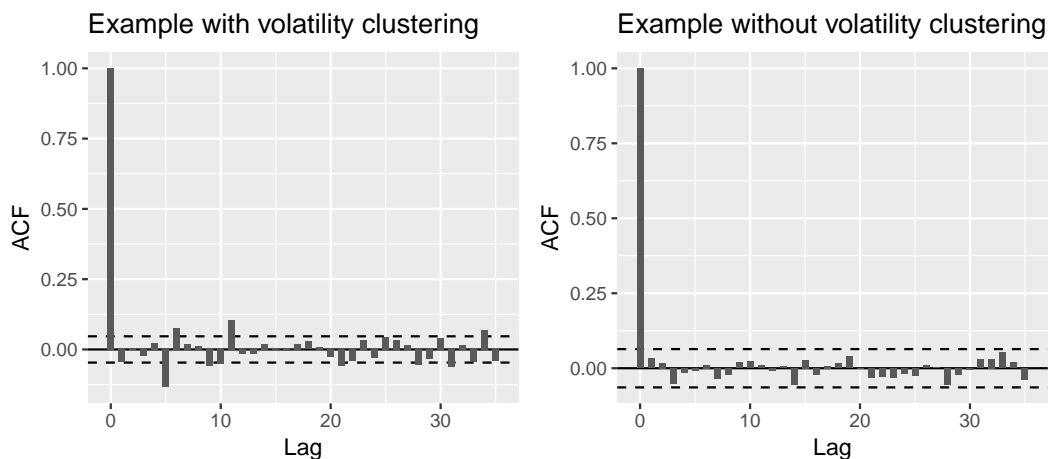


Figure 2.1.: Empirical ACF of log-returns with lags h up to 35 - volatility clustering vs. without volatility clustering.

The by all appearances random variations of asset prices share some non trivial statistical characteristics. In general such properties (as described in Cont (2001)), common across a wide range of financial markets and instruments such as log-returns on indices, interest rates and commodity prices are called *stylized facts*:

- **Volatility clustering:** Volatility of the log-return series is time varying and appears to be clustered.
- **Heavy tails:** The distribution of return series seems to display heavy-tails. Hence normality has to be rejected.
- **Leverage effect:** The changes in stock prices tend to be negatively correlated with the changes in volatility.
- **Mean reverting:** The Volatility is mean-reverting, meaning that there is a normal level to which volatility will eventually return.

2. Univariate GARCH model

- **Serial correlation:** Especially absolute and squared values of the return series display serial correlation.

The time series of daily log-returns r_t is defined as $r_t = \ln\left(\frac{S_t}{S_{t-1}}\right)$, where S_t are observed market prices at time t . In the following chapters, the term *returns* always refers to *log-returns*, unless otherwise specified. The variable r_t can then be modeled as

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \\ \epsilon_t &= \sigma_t z_t \text{ with } z_t \sim D(0, 1), \end{aligned}$$

where μ_t describes the expected value of conditional r_t ($\mathbb{E}[r_t | \mathcal{F}_{t-1}] = \mu_t$), σ_t^2 the conditional variance ($\mathbb{E}[r_t^2 | \mathcal{F}_{t-1}] = \sigma_t^2$) and z_t is an i.i.d. process of some distribution $D(0, 1)$ with zero mean and unit variance. \mathcal{F}_t denotes the information set available at time t , i.e. $\mathcal{F}_t = \{r_k | k \leq t\}$.

Definition 15 (The ARCH(q) model)

A process $(\epsilon_t)_{t \in \mathbb{Z}}$ is called a ARCH(q) process if for every t ,

$$\begin{aligned} \epsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2, \end{aligned}$$

where $q \in \mathbb{N}$, $\alpha_0, \alpha_1, \dots, \alpha_q \geq 0$ and $(z_t)_{t \in \mathbb{Z}}$ an i.i.d. sequence with zero mean and unit variance independent of $\{\epsilon_{t-k}, k \geq 1\}$, $\forall t \in \mathbb{Z}$.

Using the information \mathcal{F}_{t-1} , given at time $t - 1$, we can calculate the conditional mean and variance of ϵ_t :

$$\mathbb{E}[\epsilon_t | \mathcal{F}_{t-1}] = \mathbb{E}[\sigma_t z_t | \mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[z_t | \mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[z_t] = 0 \quad (2.1)$$

=0, using (2.1)

$$\text{Var}[\epsilon_t | \mathcal{F}_{t-1}] = \mathbb{E}[\sigma_t^2 z_t^2 | \mathcal{F}_{t-1}] - \underbrace{(\mathbb{E}[\sigma_t z_t | \mathcal{F}_{t-1}])^2}_{=0} = \sigma_t^2 \mathbb{E}[z_t^2 | \mathcal{F}_{t-1}] \quad (2.2)$$

$$= \sigma_t^2 (\text{Var}[z_t | \mathcal{F}_{t-1}] + \mathbb{E}[z_t | \mathcal{F}_{t-1}]^2) = \sigma_t^2 \text{Var}[z_t] = \sigma_t^2 \quad (2.3)$$

The Generalized Autoregressive Conditional Heteroscedasticity model (GARCH), which allows for a longer memory and more flexible lag structure, is an extension of the ARCH model.

2. Univariate GARCH model

Definition 16 (The GARCH(p,q) model)

A process $(\epsilon_t)_{t \in \mathbb{Z}}$ is called a GARCH(p,q) process if for every t ,

$$\begin{aligned}\epsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2,\end{aligned}$$

where $p \in \mathbb{N}_0$, $q \in \mathbb{N}$, $\alpha_0, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p \geq 0$ and $(z_t)_{t \in \mathbb{Z}}$ an i.i.d. sequence with zero mean and unit variance independent of $\{\epsilon_{t-k}, k \geq 1\}$, $\forall t \in \mathbb{Z}$.

In the ARCH(q) model the conditional variance is specified as a linear function of the past sample variances, while in the GARCH(p,q) model lagged conditional variances are also taken into account.

2.1. GARCH(1,1) model

The GARCH(1,1) is the simplest but many times still very useful GARCH model. We will use this model in our practical applications. The GARCH(1,1) model is given by

$$\begin{aligned}\epsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,\end{aligned}$$

where $(z_t)_{t \in \mathbb{Z}}$ is an i.i.d. sequence with zero mean and unit variance.

Using the so called "surprise in squared returns" method one can rewrite the GARCH(1,1) model to an ARMA(1,1) model on squared residuals ϵ_t^2 by making the substitution $\eta_t = \epsilon_t^2 - \sigma_t^2$. We can show that $\{\eta_t\}$ is a serially uncorrelated sequence with zero mean and uncorrelated with past squared returns.

2. Univariate GARCH model

$$\begin{aligned}\epsilon_t^2 &= \sigma_t^2 + \eta_t \\ &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \eta_t\end{aligned}$$

Using $\eta_{t-1} = \epsilon_{t-1}^2 - \sigma_{t-1}^2$ or equivalently $\sigma_{t-1}^2 = \epsilon_{t-1}^2 - \eta_{t-1}$ one gets

$$\begin{aligned}\epsilon_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 (\epsilon_{t-1}^2 - \eta_{t-1}) + \eta_t \\ &= \alpha_0 + \eta_t + (\alpha_1 + \beta_1) \epsilon_{t-1}^2 - \beta_1 \eta_{t-1},\end{aligned}$$

which is a $\text{ARMA}(1,1)^2$ model for the squared residuals.

We see that a $\text{GARCH}(1,1)$ model for the returns is the same as a $\text{ARMA}(1,1)$ model for the squared residuals.

The unconditional mean of ϵ_t is:

$$\begin{aligned}\mathbb{E}[\epsilon_t] &= \mathbb{E}[\mathbb{E}[\epsilon_t | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[\mathbb{E}[\sigma_t z_t | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[\sigma_t \mathbb{E}[z_t | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[\sigma_t \mathbb{E}[z_t]] = 0\end{aligned}$$

Here we use that $\mathbb{E}[z_t] = 0, \forall t \in \mathbb{Z}$ and the independence of z_t and σ_t .

The unconditional variance of ϵ_t is:

$$\begin{aligned}\sigma^2 = \text{Var}[\epsilon_t] &= \mathbb{E}[\epsilon_t^2] - (\mathbb{E}[\epsilon_t])^2 \\ &= \mathbb{E}[\epsilon_t^2] \\ &= \mathbb{E}[\mathbb{E}[\epsilon_t^2 | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[\mathbb{E}[\sigma_t^2 z_t^2 | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[\sigma_t^2 \mathbb{E}[z_t^2]] \\ &= \mathbb{E}[\sigma_t^2] \\ &= \alpha_0 + \alpha_1 \mathbb{E}[\epsilon_{t-1}^2] + \beta_1 \mathbb{E}[\sigma_{t-1}^2] \\ &= \alpha_0 + \alpha_1 \underbrace{\text{Var}[\epsilon_{t-1}]}_{=\sigma^2} + \beta_1 \underbrace{\mathbb{E}[\sigma_{t-1}^2]}_{=\sigma^2},\end{aligned}$$

²A $\text{ARMA}(p,q)$ model refers to a model with p autoregressive terms and q moving average terms and is given by: $r_t = c + \epsilon_t + \sum_{i=1}^p a_i r_{t-i} + \sum_{j=1}^q b_j \epsilon_{t-j}$

2. Univariate GARCH model

Here we use that $\mathbb{E}[z_t^2] = \mathbb{E}[z_t^2] - E[z_t]^2 = \text{Var}[z_t] = 1$ and the independence of z_t^2 and σ_t^2 . If we assume $\{\epsilon_t\}$ to be a stationary process, we get $\sigma^2 = \text{Var}[\epsilon_t] = \text{Var}[\epsilon_{t-1}] = \dots = \text{Var}[\epsilon_{t-q}]$, which leads to the unconditional variance, by solving with respect to σ^2 :

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \quad (2.4)$$

The unconditional autocovariance of ϵ_t is:

$$\mathbb{E}[\epsilon_{t+k}\epsilon_{t-1}] = \mathbb{E}[\mathbb{E}[\epsilon_{t+k}\epsilon_{t+1}|\mathcal{F}_{t-1}]] = \mathbb{E}[\epsilon_{t+k} \underbrace{\mathbb{E}[\epsilon_{t+1}|\mathcal{F}_{t-1}]}_{=0}] = 0 \text{ for } k = 0, 1, 2, \dots$$

Stationarity

According to Theorem 1 in Bollerslev (1986) a GARCH(p,q) process is stationary if $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$. Therefore a GARCH(1,1) process is stationary if $\alpha_1 + \beta_1 < 1$.

2.1.1. Forecasting volatility

One major use of the GARCH model is the prediction of future volatilities. We are especially interested in predicting future volatilities k-steps ahead, hence in the estimation of σ_{t+k} for $k \geq 1$. In the assumed GARCH(1,1) we have:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

2. Univariate GARCH model

In particular for GARCH(1,1) and when looking $k \geq 1$ steps ahead using the information at time t this yields a general formulation as follows:

$$\begin{aligned}
 \mathbb{E}[\sigma_{t+1}^2 | \mathcal{F}_t] &= \alpha_0 + \alpha_1 \mathbb{E}[\epsilon_t^2 | \mathcal{F}_t] + \beta_1 \mathbb{E}[\sigma_t^2 | \mathcal{F}_t] \\
 &= \alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2 \\
 \mathbb{E}[\sigma_{t+2}^2 | \mathcal{F}_t] &= \alpha_0 + \alpha_1 \mathbb{E}[\epsilon_{t+1}^2 | \mathcal{F}_t] + \beta_1 \mathbb{E}[\sigma_{t+1}^2 | \mathcal{F}_t] \\
 &= \alpha_0 + (\alpha_1 + \beta_1) \mathbb{E}[\sigma_{t+1}^2 | \mathcal{F}_t] \\
 &= \alpha_0 + (\alpha_1 + \beta_1) [\alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2] \\
 &= \alpha_0 [1 + (\alpha_1 + \beta_1)] + (\alpha_1 + \beta_1)^2 \sigma_t^2 \\
 \mathbb{E}[\sigma_{t+3}^2 | \mathcal{F}_t] &= \alpha_0 + \alpha_1 \mathbb{E}[\epsilon_{t+2}^2 | \mathcal{F}_t] + \beta_1 \mathbb{E}[\sigma_{t+2}^2 | \mathcal{F}_t] \\
 &= \alpha_0 + (\alpha_1 + \beta_1) \mathbb{E}[\sigma_{t+2}^2 | \mathcal{F}_t] \\
 &= \alpha_0 + (\alpha_1 + \beta_1) [\alpha_0 [1 + (\alpha_1 + \beta_1)] + (\alpha_1 + \beta_1)^2 \sigma_t^2] \\
 &= \alpha_0 [1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2] + (\alpha_1 + \beta_1)^3 \sigma_t^2 \\
 &\quad \vdots \\
 \mathbb{E}[\sigma_{t+k}^2 | \mathcal{F}_t] &= \alpha_0 [1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2 + \dots + (\alpha_1 + \beta_1)^{k-1}] + (\alpha_1 + \beta_1)^k \sigma_t^2 \\
 &= \alpha_0 \sum_{i=0}^{k-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^k \sigma_t^2 \\
 &= \alpha_0 \left(\frac{1 - (\alpha_1 + \beta_1)^k}{1 - \alpha_1 - \beta_1} \right) + (\alpha_1 + \beta_1)^k \sigma_t^2 \\
 &\stackrel{(2.4)}{=} \sigma^2 (1 - (\alpha_1 + \beta_1)^k) + (\alpha_1 + \beta_1)^k \sigma_t^2 \\
 &= \sigma^2 + (\alpha_1 + \beta_1)^k [\sigma_t^2 - \sigma^2]
 \end{aligned}$$

Here in each step and for $k \geq 0$ we use:

$$\begin{aligned}
 \mathbb{E}[\epsilon_{t+k}^2 | \mathcal{F}_t] &= \mathbb{E}[\sigma_{t+k}^2 z_{t+k}^2 | \mathcal{F}_t] \\
 &= \mathbb{E}[\sigma_{t+k}^2 | \mathcal{F}_t] \mathbb{E}[z_{t+k}^2 | \mathcal{F}_t] \\
 &= \mathbb{E}[\sigma_{t+k}^2 | \mathcal{F}_t] (\text{Var}[z_{t+k} | \mathcal{F}_t] + (\mathbb{E}[z_{t+k} | \mathcal{F}_t])^2) \\
 &= \mathbb{E}[\sigma_{t+k}^2 | \mathcal{F}_t]
 \end{aligned}$$

We also used the geometric series $\sum_{i=0}^{k-1} (\alpha_1 + \beta_1)^i$ with $0 < (\alpha_1 + \beta_1) < 1$. In the second to last step we have substituted for the unconditional variance,

2. Univariate GARCH model

$\sigma^2 = \alpha_0 / (1 - \alpha_1 - \beta_1)$. Since $(\alpha_1 + \beta_1) < 1$, we have $(\alpha_1 + \beta_1)^k \rightarrow 0$ as $k \rightarrow \infty$. We obtain the stationary variance as $\mathbb{E}[\sigma_{t+k}^2 | \mathcal{F}_t] \xrightarrow{k \rightarrow \infty} \frac{\alpha_0}{1 - \alpha_1 - \beta_1} = \sigma^2$ as the forecast horizon goes to infinity. The sum of the parameters α_1 and β_1 indicate how quickly the variance forecast converges to the unconditional variance.

2.1.2. Estimation of parameters

The following subsection is based on Subsection (4.3.4) of A. J. McNeil, Frey, and Embrechts (2006). In practice, the parameters of the GARCH(1,1) model can be estimated by a maximum likelihood approach on the basis of historical data. Suppose we have a total of $n + 1$ data values $\mathcal{E}_0, \dots, \mathcal{E}_n$. Rewriting the joint density function of the corresponding random variables yields:

$$\begin{aligned}
 f_{\mathcal{E}_0, \dots, \mathcal{E}_n}(\epsilon_0, \dots, \epsilon_n) &= f_{\mathcal{E}_n | \mathcal{E}_{n-1}, \dots, \mathcal{E}_0}(\epsilon_n | \epsilon_{n-1}, \dots, \epsilon_0) f_{\mathcal{E}_0, \dots, \mathcal{E}_{n-1}}(\epsilon_0, \dots, \epsilon_{n-1}) \\
 &= f_{\mathcal{E}_{n-1} | \mathcal{E}_{n-2}, \dots, \mathcal{E}_0}(\epsilon_{n-1} | \epsilon_{n-2}, \dots, \epsilon_0) \\
 &\quad \times f_{\mathcal{E}_n | \mathcal{E}_{n-1}, \dots, \mathcal{E}_0}(\epsilon_n | \epsilon_{n-1}, \dots, \epsilon_0) \\
 &\quad \times f_{\mathcal{E}_0, \dots, \mathcal{E}_{n-2}}(\epsilon_0, \dots, \epsilon_{n-2}) \\
 &\quad \vdots \\
 &= f_{\mathcal{E}_0}(x_0) \prod_{t=1}^n f_{\mathcal{E}_t | \mathcal{E}_{t-1}, \dots, \mathcal{E}_0}(\epsilon_t | \epsilon_{t-1}, \dots, \epsilon_0) \tag{2.5}
 \end{aligned}$$

For determining the likelihood function we have to assume a model for z_t . The simplest choice of an error distribution would be to assume z_t to be i.i.d standard normal distributed ($z_t \sim N(0, 1)$), hence ϵ_t conditional on the past information will follow a Gaussian distribution with mean zero and variance σ_t^2 ($\epsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$, since $\epsilon_t = \sigma_t z_t$). Examples of other error distributions can be found in Subsection 2.1.3.

The likelihood function however can not simply be based on (2.5), because the marginal density function $f_{\mathcal{E}_0}$ is not known in a tractable closed form for the GARCH model. For a GARCH(1,1) model σ_t is defined recursively

2. Univariate GARCH model

through σ_{t-1} . For that reason we will now construct the joint density function of $\mathcal{E}_1, \dots, \mathcal{E}_n$ conditional on the realized values of σ_0 and \mathcal{E}_0 , which is given as:

$$f_{\mathcal{E}_1, \dots, \mathcal{E}_n | \mathcal{E}_0, \sigma_0}(\epsilon_1, \dots, \epsilon_n | \epsilon_0, \sigma_0) = \prod_{t=1}^n f_{\mathcal{E}_t | \mathcal{E}_{t-1}, \dots, \mathcal{E}_0, \sigma_0}(\epsilon_t | \epsilon_{t-1}, \dots, \epsilon_0, \sigma_0)$$

Each conditional density $f_{\mathcal{E}_t | \mathcal{E}_{t-1}, \dots, \mathcal{E}_0, \sigma_0}$ is depending on the past only via σ_t , which is given by the recursion $\sigma_t = \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2}$ from $\sigma_0, \mathcal{E}_0, \dots, \mathcal{E}_{t-1}$. The conditional likelihood function for a GARCH(1,1) model using a Gaussian distributed error function can now be stated as:

$$\begin{aligned} L(\alpha_0, \alpha_1, \beta_1 | \epsilon_0, \dots, \epsilon_n) &= f_{\mathcal{E}_1, \dots, \mathcal{E}_n | \mathcal{E}_0, \sigma_0}(\epsilon_1, \dots, \epsilon_n | \epsilon_0, \sigma_0) \\ &= \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left\{ -\frac{\epsilon_t^2}{2\sigma_t^2} \right\}, \end{aligned}$$

where $\sigma_t = \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2}$.

However, the problem that the value of σ_0^2 is not observed still remains. In practice this can be solved by using a starting value e.g. the sample variance of $\mathcal{E}_1, \dots, \mathcal{E}_n$ or in some cases even 0.

The logarithm $\ln(\cdot)$ is a strictly increasing function, therefore we can maximize the logarithm of the conditional likelihood function instead of the conditional likelihood function itself. This is preferable since the log-likelihood function is usually easier to maximize.

$$\begin{aligned} \ln(L(\alpha_0, \alpha_1, \beta_1 | \epsilon_0, \dots, \epsilon_n)) &= \ln \left(\prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left\{ -\frac{\epsilon_t^2}{2\sigma_t^2} \right\} \right) \\ &= -\frac{1}{2} \sum_{t=1}^n \left[\ln(2\pi) + \ln(\sigma_t^2) + \frac{\epsilon_t^2}{\sigma_t^2} \right] \quad (2.6) \\ &= -\frac{1}{2} \sum_{t=1}^n \left[\ln(\sigma_t^2) + \frac{\epsilon_t^2}{\sigma_t^2} \right] + \text{constant}, \end{aligned}$$

2. Univariate GARCH model

where $\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \alpha_j \sigma_{t-j}^2$ is calculated recursively. The first term of the likelihood function in (2.6) is a constant and therefore independent of the selected model parameters. By using numerical optimization methods we can find parameter $\alpha_0, \alpha_1, \beta_1$ that maximize the log-likelihood function and we refer to the optimal parameters as maximum likelihood estimates.

2.1.3. Non-Gaussian error distributions

Financial data however does often not only exhibits volatility clustering but also a skewed and heavy-tailed distribution such as:

- (i) **Student's t-distribution:** The density is given by (see Bollerslev (1987)):

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu s}} \left(1 + \frac{x^2}{s\nu}\right)^{-\frac{\nu+1}{2}},$$

with degree of freedom $\nu > 0$ and scale parameter s . The classical Student's t-density is bell-shaped and symmetric like the normal distribution, but has heavier tails. The t-distribution approaches the normal distribution as the number of degrees of freedom ν increases. The mean and variance of the Student's t-distribution are

$$\begin{aligned}\mathbb{E}[z_t] &= 0, \nu > 0 \\ \text{Var}[z_t] &= \frac{s\nu}{\nu - 2},\end{aligned}$$

hence to achieve zero mean and unit variance for the error term z_t in the GARCH model, the scale parameter s must be set to $\frac{\nu-2}{\nu}$.

- (ii) **Skewed Student's t-distribution:** The density is given by (see Fernandez and Steel (1998)):

$$f(x) = \frac{2}{\gamma + \frac{1}{\gamma}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\rho}} \left[1 + \frac{(x-\mu)^2}{\nu\rho} \left(\frac{1}{\gamma^2} \mathbb{1}_{[0,\infty)}(x-\mu) + \gamma^2 \mathbb{1}_{(-\infty,0)}(x-\mu)\right)\right]^{-\frac{\nu+1}{2}},$$

with location parameter $\mu \in \mathbb{R}$, $\gamma > 0$ the skewness parameter, ρ the scale parameter and $\nu > 0$ the degrees of freedom. If $\gamma = 1$ the skewness is 0 and f becomes the non-standardized Student's t-distribution with ν degrees of freedom.

2. Univariate GARCH model

(iii) **Normal inverse Gaussian (NIG) distribution:** The density is given by (see Andersson (2001)):

$$f(x) = \frac{\alpha \delta K_1 \left(\alpha \sqrt{\delta^2 + (x - \mu)^2} \right)}{\pi \sqrt{\delta^2 + (x - \mu)^2}} \exp \left[\delta \sqrt{\alpha^2 - \beta^2} + \beta(x - \mu) \right],$$

with location parameter $\mu \in \mathbb{R}$, scale parameter $\gamma > 0$ and asymmetry parameter α and tail heaviness parameter β , which satisfy $0 \leq |\beta| \leq \alpha$. K_1 is the modified third order Bessel function of index 1. As a special case for $\alpha \rightarrow \infty$ and $\delta/\alpha = \sigma^2$ we obtain a normal distribution $N(\mu, \sigma^2)$.

We need to use standardized versions of these distributions since we require z_t do be of mean zero and unit variance.

2.2. Ljung-Box test

After specifying a time series model for $(r_t)_{t=1, \dots, n}$ we are investigating the appropriateness of the chosen model, i.e. the goodness-of-fit. The Ljung-Box test (see Ljung and Box (1978)) applied to the residual series, tests whether the residuals of the fitted model still exhibit serial dependence, or in other words it tests whether the first m sample autocorrelations $\hat{\rho}_k^2$ for $k = 1, \dots, m$ from a process are collectively small in magnitude.

H_0 : The data is independently distributed, i.e. the correlations in the population from which the sample is taken are 0. Hence any observed correlations in the data comes from randomness of the sampling process.

H_1 : Not H_0 . At least one correlation in the population from which the sample is taken is not 0. Hence the data is not independently distributed and exhibits serial correlation.

For a fixed sufficiently large number m of autocorrelations, the Ljung-Box statistic Q_m is given by

$$Q_m = n(n+2) \sum_{h=1}^m \frac{\hat{\rho}_h^2}{n-h}$$

2. Univariate GARCH model

where n denotes the sample size and $\hat{\rho}_h$ is the sample autocorrelation at lag h . This test jointly considers the autocorrelations of the first m lag, where $1 \leq m \leq n - 1$. Under the null hypothesis the statistic Q_m asymptotically follows a χ^2 distribution, with m degrees of freedom. The null hypothesis H_0 can be rejected for some chosen significance level α if

$$Q_m > \chi_{1-\alpha, m}^2$$

where m is number of degrees of freedom and $\chi_{1-\alpha, m}^2$ is the $(1 - \alpha)$ -quantile of the chi-squared distribution.

2.3. Limitations of the GARCH model

Despite the usefulness of the GARCH model across a wide range of applications, due to the quite simple structure according to Rossi (2004) some important limitations arise.

- GARCH models often fail to fully capture observed heavy-tails in the return series of assets. Although heteroscedasticity explains some of the heavy-tail behavior it explains typically not all of it. Heavy-tailed distributions, such as Student's t-distribution, have been applied in GARCH modeling. However the right choice of distribution is often a matter of trial and error.
- GARCH models usually solve only one part of the modeling problem. Financial decisions are rarely based solely on expected returns and volatilities.
- GARCH models basically specify the behavior of squared data. This leads to the problem that a few large observations could dominate the sample.

These and other limitations are the main reason for the development of extensions to the GARCH model such as Exponential GARCH (EGARCH), Threshold GARCH (TGARCH), GARCH-in-Mean (GARCH-M) and many others. A overview of several different types of GARCH models can be found in Ali (2013).

3. Bivariate copula families

Copulas can be described as functions that allow multivariate distribution functions to be obtained from the marginal distributions. Although copulas have been formally defined in 1959, the first (published) application in risk management was as recent as the early 1990s. Copulas are a tool that helps in understanding dependence at a deeper level and allows us to see the potential perils of approaches to dependence that only take correlation into account and show us a way of defining a number of useful alternative dependence measures. Copulas are described by Nelsen (2006) as "*functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions*", which explains the name "copula". Parts of this chapter are based on A. J. McNeil, Frey, and Embrechts (2006). Copulas have already been specified in Definition 5 in Section 1.5.

A very important result is Sklar's Theorem which provides the theoretical foundation for the application of copulas. It states that a joint distribution function can be written using a copula and the marginal distribution functions.

Theorem 4 (Sklar (1959))

Let $F : \mathbb{R}^n \rightarrow [0, 1]$ be a n -dimensional distribution function with margins F_1, \dots, F_n . Then there exists a copula C such that for all $x = (x_1, \dots, x_n)' \in \mathbb{R}^n$,

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (3.1)$$

If F_1, \dots, F_n are continuous, then C is unique.

Vice versa if C is a copula and F_1, \dots, F_n are univariate distribution functions, then the function F defined in (3.1) is a joint distribution function with margins F_1, \dots, F_n .

3. Bivariate copula families

From this we see that for continuous multivariate distribution functions, the univariate margins and the multivariate dependence structure can be separated. The dependence structure can be represented by a copula.

If we evaluate the equation in Sklar's Theorem at the arguments $x_i = F_i^{\leftarrow}(u_i)$, $0 \leq u_i \leq 1$, $i = 1, \dots, n$ we obtain the following theorem.

Theorem 5

Let $F : \mathbb{R}^n \rightarrow [0, 1]$ be a distribution function with continuous margins F_1, \dots, F_n . The unique copula C is defined through

$$C(u_1, \dots, u_n) = F(F_1^{\leftarrow}(u_1), \dots, F_n^{\leftarrow}(u_n)),$$

where F_i^{\leftarrow} are the generalized inverse functions of the margins as defined in Definition 3.

If the multivariate distribution function F is in addition absolutely continuous with strictly increasing, continuous margins F_1, \dots, F_n , then the copula density can (see A. J. McNeil, Frey, and Embrechts (2006)) be expressed as

$$\begin{aligned} c(F_1(x_1), \dots, F_n(x_n)) &= \frac{\partial^n C(F_1(x_1), \dots, F_n(x_n))}{\partial F_1(x_1) \dots \partial F_n(x_n)} \\ &\stackrel{\text{Sklar's Theorem}}{=} \frac{\partial^n F(x_1, \dots, x_n)}{\partial F_1(x_1) \dots \partial F_n(x_n)} \\ &= \frac{f(x_1, \dots, x_n)}{f_1(x_1) \dots f_n(x_n)}, \end{aligned}$$

where f is the joint density of F and f_1, \dots, f_n are the marginal densities. Hence for the 2-dimensional case this simplifies to

$$c_{12}(u_1, u_2) = \frac{\partial^2 C_{12}(u_1, u_2)}{\partial u_1 \partial u_2}. \quad (3.2)$$

For the joint density function $f(x_1, \dots, x_n)$ we have

$$f(x_1, \dots, x_n) = c_{12\dots n}(F_1(x_1), \dots, F_n(x_n))f_1(x_1) \dots f_n(x_n), \quad (3.3)$$

for some uniquely defined n-variate copula density $c_{12\dots n}$. For the 2-dimensional case we get

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2).$$

3. Bivariate copula families

Theorem 6 (Fréchet–Hoeffding (1957))

For any copula $C : [0, 1]^n \rightarrow [0, 1]$ and any $(u_1, \dots, u_n) \in [0, 1]^n$ the following bounds hold:

$$W^n(u_1, \dots, u_n) \leq C(u_1, \dots, u_n) \leq M^n(u_1, \dots, u_n)$$

The function W^n is called the lower Fréchet–Hoeffding bound and is defined as

$$W^n(u_1, \dots, u_n) = \max\{u_1 + \dots + u_n - n + 1, 0\},$$

whereas the function M^n is called the upper Fréchet–Hoeffding bound which is defined as

$$M^n(u_1, \dots, u_n) = \min\{u_1, \dots, u_n\}.$$

Proof.

$$\begin{aligned} C(u_1, \dots, u_n) &= \mathbb{P}(U_1 \leq u_1, \dots, U_n \leq u_n) = \mathbb{P}\left(\bigcap_{i=1}^n \{U_i \leq u_i\}\right) \\ &\leq \mathbb{P}(U_i \leq u_i) = u_i, \forall i \end{aligned}$$

From this it follows that $C(u_1, \dots, u_n) \leq \min\{u_1, \dots, u_n\} =: M^n(u_1, \dots, u_n)$.

$$\begin{aligned} C(u_1, \dots, u_n) &= \mathbb{P}\left(\bigcap_{i=1}^n \{U_i \leq u_i\}\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^n \{U_i > u_i\}\right) \\ &\geq 1 - \sum_{i=1}^n \mathbb{P}(U_i > u_i) = 1 - \sum_{i=1}^n (1 - u_i) = 1 - n + \sum_{i=1}^n u_i \end{aligned}$$

However, any copula is a probability function and therefore $C \geq 0$, hence $C(u_1, \dots, u_n) \geq \max\{u_1 + \dots + u_n - n + 1, 0\} =: W^n(u_1, \dots, u_n)$. \square

Comonotonicity copula and countermonotonicity copula

In a 2-dimensional setting the upper and lower bounds are themselves copulas. The 2-dimensional Fréchet–Hoeffding lower bound $W^2(u, v)$ is called *countermonotonicity copula* and describes *perfect negative dependence* whereas the Fréchet–Hoeffding upper bound $M^2(u, v)$ is called *comonotonicity copula* and it describes *perfect positive dependence*.

In the 2-dimensional case, the Fréchet–Hoeffding Theorem reduces to:

$$W^2(u, v) = \max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\} = M^2(u, v)$$

3. Bivariate copula families

Independence copula

The most simple of all copula models is the so-called independence copula:

$$C^{Ind}(u_1, \dots, u_n) = \prod_{i=1}^n u_i \quad (3.4)$$

It immediately follows from Sklar's Theorem, especially from equation (3.1), that random variables with continuous distributions are independent if and only if their dependence structure is given by the independence copula (3.4).

Figure 3.1 depicts the perspective plots and the contour plots of the comonotonicity copula $M^2(u, v)$, the independence copula $C^{Ind}(u, v)$ and the countermonotonicity copula $W^2(u, v)$.

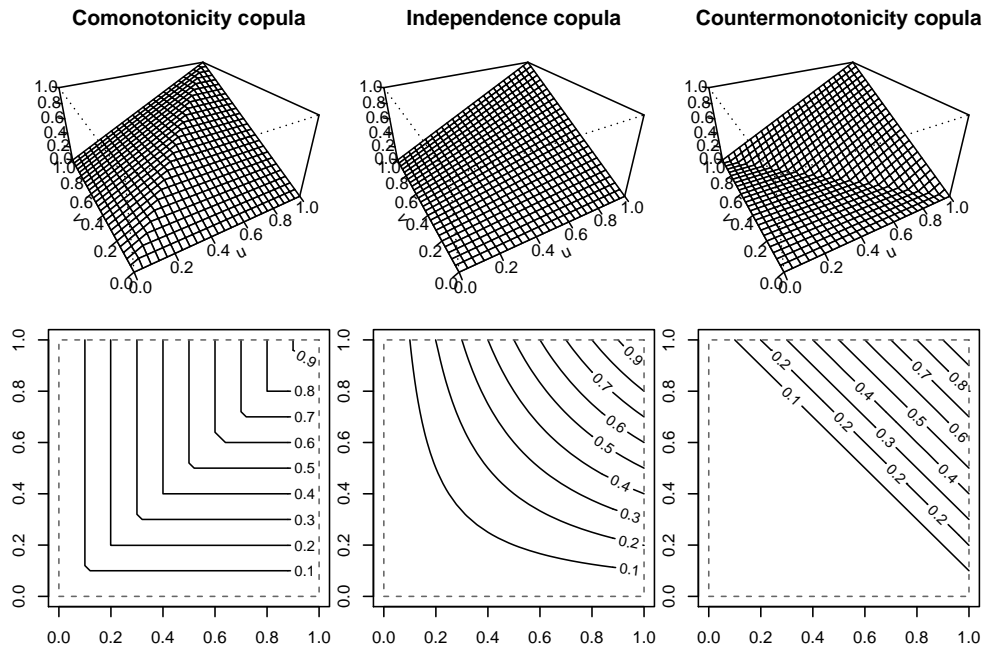


Figure 3.1.: Perspective plots and contour plots of the comonotonicity copula, the independence copula and the countermonotonicity copula.

3.1. Examples of common elliptical and Archimedean bivariate copulas

Bivariate copulas, also known as pair-copulas, are the basic building blocs used for the construction of multivariate models in the following chapters. In this chapter we present six of the most commonly used and discussed copulas: Gaussian, Student's t , Clayton, Gumbel, Frank and Joe copula. The first two belong to the family of elliptical copulas. They both do not have a simple closed form and are hence called implicit copulas. The last four types of copula, all from the Archimedean copula family. They all have a distribution function with a simple closed form and are therefore called explicit copulas. Each discussed copula has its own distinct properties and strength of dependence in the tails of the bivariate distribution. These distinct properties in strength of dependence in the tails will be described by the so called *upper tail dependence* and the *lower tail dependence*, which was already introduced in Section 1.6.4. However note that there are copulas which are neither elliptical nor Archimedean. Those copulas will however not be discussed in this thesis.

In this chapter we will also present important formulas for each copula, the copula density, the h-function and the inverse h-function. Although it is possible to give an explicit h-function of all six discussed pair-copulas, it is not possible to analytically invert these functions for all six copulas. In a few cases numerical inversion will be necessary. Availability of the h-function and the inverse h-function in an explicit form is important for efficiency. According to Joe (1997) in some cases it might be better to use a different copula with similar properties instead. At this point we have not formally introduced the h-function, however we will still state the h-function and inverse h-function for all six bivariate copulas below. A formal introduction to the h-function will be given in Equation (4.5) in Chapter 4.1.

The copulas presented here are among the most discussed copulas in literature. But keep in mind that these are just a tiny sample of all available bivariate copulas. A good overview of 10 different copulas, including the once described below can be found in Brechmann (2010). In addition to the definitions Brechmann also states the density function, Kendall's τ and lower and upper tail dependence coefficients.

3. Bivariate copula families

The following two sections are based on Brechmann (2010) and A. J. McNeil, Frey, and Embrechts (2006).

3.2. Elliptical copulas

The copulas generated by elliptical distributions are known as elliptical copulas. The most commonly used elliptical copulas are the Gaussian copula and the Student's t-copula. One advantage of elliptical copulas is that one can specify different levels of correlation between marginal distributions. However, as mentioned above it is not possible to have a simple closed form for the distribution function.

3.2.1. Gaussian copula

Let C_R^{Ga} be the copula of a n -dimensional normal distribution with correlation matrix R . We have

$$X \sim N_n(0, R).$$

The copula is given by

$$C_R^{Ga}(u) = \Phi_R(\phi^{-1}(u_1), \dots, \phi^{-1}(u_n)), \quad (3.5)$$

where ϕ^{-1} denotes the inverse cumulative distribution function of a standard normal distribution and Φ_R denotes the joint cumulative distribution function of X with correlation matrix R . The Gaussian copula does not have a simple closed form. We have to express the copula as an integral over the density of X . In the bivariate case using (3.5) we get:

$$C_\rho^{Ga}(u, v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} dx dy,$$

where $\rho \in (-1, 1)$ represents the linear correlation coefficient. By applying (3.2) the density function of the Gaussian copula is given by

$$c_\rho^{Ga}(u, v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{\rho^2(x^2 + y^2) - 2\rho xy}{2(1-\rho^2)}\right\},$$

3. Bivariate copula families

where $x = \phi^{-1}(u)$ and $y = \phi^{-1}(v)$ and ϕ^{-1} denotes the inverse cumulative distribution function of the univariate standard normal distribution. For this copula the h-function is given by (see Appendix A.4.1 for details)

$$h(u, v, \rho) = \phi \left(\frac{\phi^{-1}(u) - \rho \phi^{-1}(v)}{\sqrt{1 - \rho^2}} \right)$$

and the inverse of the h-function is given by

$$h^{-1}(u, v, \rho) = \phi \left(\phi^{-1}(u) \sqrt{1 - \rho^2} + \rho \phi^{-1}(v) \right).$$

According to A. J. McNeil, Frey, and Embrechts (2006) and using (1.7) and (1.6) the rank correlations are:

$$\begin{aligned} \rho_{\tau}(X_1, X_2) &= \frac{2}{\pi} \arcsin \rho \\ \rho_S(X_1, X_2) &= \frac{6}{\pi} \arcsin \frac{1}{2} \rho \end{aligned}$$

In a financial context the multivariate Gaussian distribution should not be used even if it seems very enticing. Empirical studies have shown that using Gaussian copulas in the context of assets that exhibit skewed and fat-tailed distributions, which are often exhibited by financial series, can lead to underestimation of the joint tail probability.

3.2.2. Student's t-copula

To exhibit tail dependence, a frequently used alternative to the Gaussian copula is the *Student's t-copula*. Similarly to how we extracted a copula from the multivariate normal distribution, we can extract an implicit copula from other distributions with continuous marginal distribution functions.

Let $C_{\nu, R}^t$ be the copula of a n -dimensional Student's t-distribution with correlation matrix R . We have

$$X \sim t_n(\nu, 0, R).$$

3. Bivariate copula families

The copula is given by

$$C_{\nu,R}^t(u) = t_{\nu,R}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_n)), \quad (3.6)$$

where t_{ν}^{-1} denotes the inverse cumulative distribution function of a standard univariate Student's t-distribution with ν degrees of freedom and $t_{\nu,R}$ denotes the joint cumulative distribution function of X with correlation matrix R and ν degrees of freedom. The copula can again be expressed as the integral over X . In the bivariate case using (3.6) we get:

$$C_{\rho,\nu}^t(u, v) = \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(v)} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\pi\nu\sqrt{(1-\rho^2)}} \left\{ 1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)} \right\}^{-\frac{\nu+2}{2}} dx dy,$$

where ν represents the degrees of freedom and $\rho \in (-1, 1)$ is the linear correlation coefficient. By applying (3.2) the density function of the Student's t-copula is given by:

$$c_{\rho,\nu}^t(u, v) = \frac{\Gamma(\frac{\nu+2}{2})/\Gamma(\frac{\nu}{2})}{\nu\pi dt(u, \nu) dt(v, \nu)\sqrt{1-\rho^2}} \left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)} \right)^{-\frac{\nu+2}{2}},$$

where $x = t_{\nu}^{-1}(u)$ and $y = t_{\nu}^{-1}(v)$. $dt(\cdot, \nu)$ and $t_{\nu}^{-1}(\cdot)$ denote the probability density and the inverse cumulative distribution function of a standard univariate Student's t-distribution with ν degrees of freedom, expectation 0 and variance $\frac{\nu}{\nu-2}$. For this copula the h-function is given by (see Appendix A.4.2 for details)

$$h(u, v, \rho, \nu) = t_{\nu+1} \left(\frac{t_{\nu}^{-1}(u) - \rho t_{\nu}^{-1}(v)}{\sqrt{\frac{(v + (t_{\nu}^{-1}(v))^2)(1-\rho^2)}{\nu+1}}} \right)$$

and the inverse of the h-function is given by

$$h^{-1}(u, v, \rho, \nu) = t_{\nu} \left(t_{\nu+1}^{-1}(u) \sqrt{\frac{(v(t^{-1}(v))^2)(1-\rho^2)}{\nu+1}} + \rho t_{\nu}^{-1}(v) \right).$$

3. Bivariate copula families

According to A. J. McNeil, Frey, and Embrechts (2006) and using (1.7) and (1.6) the rank correlations are:

$$\rho_{\tau}(X_1, X_2) = \frac{2}{\pi} \arcsin \rho$$

$$\rho_S(X_1, X_2) = \frac{6}{\pi} \arcsin \frac{1}{2}\rho$$

Similarly to the Student's t-distribution approaching normality with increasing degrees of freedom the Student's t-copula is approaching the Gaussian copula as the degrees of freedom increase. Note the similarity in the simulations from a Student's t-copula with $\nu = 30$ degrees of freedom and a Gaussian copula in Figure 3.2.

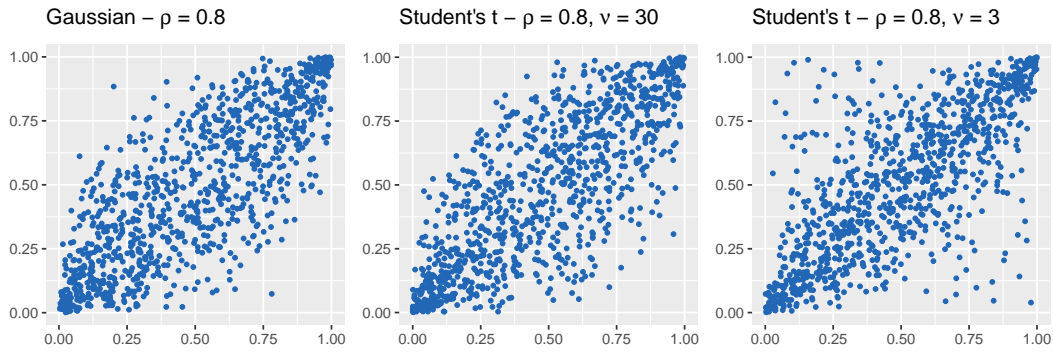


Figure 3.2.: Elliptical copulas: Illustration of dependence structure difference for Gaussian and Student's t-copulas with the same dependence parameter ρ and how the Student's t-copula converges to the Gaussian copula for higher degrees of freedom ν .

In theory all previously defined bivariate copulas can be generalized to higher dimensions. However most of the copulas are designed for bivariate distributions only. Unfortunately, most financial applications of copulas use numerous variables and hence require many marginal distributions to be taken into account.

As mentioned above the Gaussian copula and Student's t-copula do not have simple closed forms and are known as *implicit copulas*. In the next section we will look at *explicit copulas* - copulas which have simple closed forms.

3.3. Archimedean copulas

Definition 17 (Pseudo-inverse function)

Let $\phi : [0, 1] \rightarrow [0, \infty]$ be a continuous, and strictly decreasing function such that $\phi(1) = 0$. The pseudo-inverse function $\phi^{[-1]} : [0, \infty] \rightarrow [0, 1]$ of ϕ is given by:

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & 0 \leq t \leq \phi(0) \\ 0 & \phi(0) \leq t \leq \infty \end{cases}$$

$\phi^{[-1]}$ is continuous and decreasing on $[0, \infty]$, and strictly decreasing on $[0, \phi(0)]$ and we have:

$$\phi^{[-1]}(\phi(u)) = u \text{ for } u \in [0, 1]$$

$$\phi(\phi^{[-1]}(t)) = \begin{cases} t & 0 \leq t \leq \phi(0) \\ \phi(0) & \phi(0) \leq t \leq \infty \end{cases}$$

Finally, if $\phi(0) = \infty$, then $\phi^{[-1]} = \phi^{-1}$.

Theorem 7 (Archimedean copula)

Let $\phi : [0, 1] \rightarrow [0, \infty]$ be a continuous and strictly decreasing function such that $\phi(1) = 0$ and let $\phi^{[-1]}$ denote the pseudo-inverse of ϕ . Let C be the function from $[0, 1]^2$ to $[0, 1]$ given by

$$C(u_1, u_2) = \phi^{[-1]}(\phi(u_1) + \phi(u_2)).$$

Then C is a copula if and only if ϕ is convex. ϕ is called the generator of a copula C . In the case of $\phi(0) = \infty$ we say that ϕ is a strict generator. In this case $\phi^{[-1]} = \phi^{-1}$ and hence $C(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2))$ is called a strict Archimedean copula.

This class of copulas has proved to be very useful for modeling the credit risk of a portfolio. Many families in the class of Archimedean copulas allow for a great variety of different dependence structures. According to Nelsen (2006) it is possible to calculate Kendall's rank correlation for Archimedean copulas directly from the generator using the following theorem.

3. Bivariate copula families

Theorem 8

Let $(X_1, X_2)^T$ be a random vector with continuous marginal distribution and a unique Archimedean copula C generated by ϕ . Then Kendall's τ can be expressed as

$$\rho_\tau(X_1, X_2) = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt.$$

One disadvantage Archimedean copulas face is the lack of free parameter choice in higher dimensions (some entries in the rank correlation matrix are forced to be equal). For this reason, we will only focus on using bivariate Archimedean copulas.

3.3.1. Clayton copula

The Clayton copula has a one-parametric generator $\phi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$, which leads to $\phi^{-1}(t) = (1 + \theta t)^{-1/\theta}$.

$$C_\theta^{Cl}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta},$$

where $0 < \theta < \infty$ is the parameter controlling the dependence. For $\theta \rightarrow 0$ the Clayton copula exhibits independence as $\lim_{\theta \rightarrow 0} C_\theta^{Cl}(u, v) = C^{Ind}(u, v) = uv$, while perfect positive dependence is achieved for $\theta \rightarrow \infty$ as $\lim_{\theta \rightarrow \infty} C_\theta^{Cl}(u, v) = M^2(u, v) = \min\{u, v\}$.

By differentiating $C_\theta^{Cl}(u, v)$ with respect to u and v we obtain the copula

3. Bivariate copula families

density:

$$\begin{aligned}
 c_{\theta}^{Cl}(u, v) &= \frac{\partial^2 C_{\theta}^{Cl}(u, v)}{\partial u \partial v} \\
 &= \frac{\partial}{\partial u} \left[\frac{\partial C_{\theta}^{Cl}(u, v)}{\partial v} \right] \\
 &= \frac{\partial}{\partial u} \left[\frac{\partial (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}}{\partial v} \right] \\
 &= \frac{\partial}{\partial u} \left[-\frac{1}{\theta} (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 1} (-\theta v^{-\theta - 1}) \right] \\
 &= -\frac{1}{\theta} \left(-\frac{1}{\theta} - 1 \right) (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 2} (-\theta u^{-\theta - 1}) (-\theta v^{-\theta - 1}) \\
 &= (-\theta) \left(-\frac{1}{\theta} - 1 \right) (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 2} (uv)^{-\theta - 1} \\
 &= (1 + \theta) (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 2} (uv)^{-\theta - 1}
 \end{aligned}$$

For this copula the h-function is given as follows (see Appendix A.4.3 for details)

$$h(u, v, \theta) = v^{-\theta - 1} (u^{-\theta} + v^{-\theta} - 1)^{-1 - 1/\theta}$$

and the inverse of the h-function is given by

$$h^{-1}(u, v, \theta) = \left[(uv^{\theta + 1})^{-\frac{\theta}{\theta + 1}} + 1 - v^{-\theta} \right]^{-1/\theta}.$$

3. Bivariate copula families

Using Theorem 8 we can calculate Kendall's τ :

$$\begin{aligned}
 \rho_{\tau}(\theta) &= 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \\
 &= 1 + 4 \int_0^1 \frac{\frac{1}{\theta}(t^{-\theta} - 1)}{\frac{1}{\theta}(-\theta)t^{-\theta-1}} dt \\
 &= 1 + 4 \int_0^1 \frac{(t^{\theta+1} - t)}{\theta} dt \\
 &= 1 + \frac{4}{\theta} \left[\frac{t^{\theta+2}}{\theta+2} - \frac{t^2}{2} \right]_0^1 \\
 &= 1 + \frac{4}{\theta} \left(\frac{1}{\theta+2} - \frac{1}{2} \right) = \frac{\theta}{\theta+2}
 \end{aligned}$$

As mentioned in Chapter 1 we can use Theorem 3 to calculate the tail dependence coefficients:

$$\begin{aligned}
 \lambda_U^{\theta, Cl} &= \lim_{u \rightarrow 1^-} \frac{1 - 2u + C_{\theta}^{Cl}(u, u)}{1 - u} = \lim_{u \rightarrow 1^-} \frac{1 - 2u + (2u^{-\theta} - 1)^{-1/\theta}}{1 - u} \\
 &\stackrel{L'Hospital}{=} \lim_{u \rightarrow 1^-} \frac{-2 + (-1/\theta)(2u^{-\theta} - 1)^{-1/\theta-1}(-\theta)2u^{-\theta-1}}{-1} = 0 \\
 \lambda_L^{\theta, Cl} &= \lim_{u \rightarrow 0^+} \frac{C_{\theta}^{Cl}(u, u)}{u} = \lim_{u \rightarrow 0^+} \frac{(2u^{-\theta} - 1)^{-1/\theta}}{u} \\
 &= \lim_{u \rightarrow 0^+} \frac{1}{(2 - u^{\theta})^{1/\theta}} = \frac{1}{2^{1/\theta}} = 2^{-1/\theta}
 \end{aligned}$$

3.3.2. Gumbel copula

The Gumbel copula has a one-parametric generator $\phi(t) = (-\ln(t))^{\theta}$, which leads to $\phi^{-1}(t) = \exp(-t^{1/\theta})$.

$$C_{\theta}^{Gu}(u, v) = \exp \left\{ -[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta} \right\},$$

where $1 \leq \theta < \infty$ is the parameter controlling the dependence. For $\theta = 1$ the Gumbel copula exhibits independence as $C_1^{Gu}(u, v) = C^{Ind}(u, v) = uv$, while

3. Bivariate copula families

perfect positive dependence is achieved for $\theta \rightarrow \infty$ as $\lim_{\theta \rightarrow \infty} C_\theta^{Gu}(u, v) = M^2(u, v) = \min\{u, v\}$.

By differentiating $C_\theta^{Gu}(u, v)$ with respect to u and v we obtain the copula density:

$$\begin{aligned}
 c_\theta^{Gu}(u, v) &= \frac{\partial^2 C_\theta^{Gu}(u, v)}{\partial u \partial v} \\
 &= \frac{\partial}{\partial u} \left[\frac{\partial C_\theta^{Gu}(u, v)}{\partial v} \right] \\
 &= \frac{\partial}{\partial u} \left[\frac{\partial \exp \{ - [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} \}}{\partial v} \right] \\
 &= \frac{\partial}{\partial u} \left[C_\theta^{Gu}(u, v) \left(-\frac{1}{\theta} \right) [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta-1} \theta (-\ln v)^{\theta-1} \left(-\frac{1}{v} \right) \right] \\
 &= \frac{\partial}{\partial u} \left[C_\theta^{Gu}(u, v) [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta-1} (-\ln v)^{\theta-1} \frac{1}{v} \right] \\
 &= \frac{1}{v} (-\ln v)^{\theta-1} \left[C_\theta^{Gu}(u, v) [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta-1} (-\ln u)^{\theta-1} \frac{1}{u} \right. \\
 &\quad \times [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta-1} \\
 &\quad \left. + C_\theta^{Gu}(u, v) (1/\theta - 1) [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta-2} \theta (-\ln u)^{\theta-1} \left(-\frac{1}{u} \right) \right] \\
 &= \frac{1}{v} (-\ln v)^{\theta-1} C_\theta^{Gu}(u, v) \left[[(-\ln u)^\theta + (-\ln v)^\theta]^{2/\theta-2} (-\ln u)^{\theta-1} \frac{1}{u} \right. \\
 &\quad \left. + (1/\theta - 1) [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta-2} \theta (-\ln u)^{\theta-1} \left(-\frac{1}{u} \right) \right] \\
 &= \frac{1}{v} (-\ln v)^{\theta-1} \frac{1}{u} (-\ln u)^{\theta-1} C_\theta^{Gu}(u, v) \left[[(-\ln u)^\theta + (-\ln v)^\theta]^{2/\theta-2} \right. \\
 &\quad \left. + (1 - \theta) [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta-2} \right] \\
 &= \frac{C_\theta^{Gu}(u, v)}{uv} \frac{(\ln v \ln u)^{\theta-1}}{[(-\ln u)^\theta + (-\ln v)^\theta]^{2-1/\theta}} \left[[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} + \theta - 1 \right]
 \end{aligned}$$

For this copula the h-function is given as follows (see Appendix A.4.4 for

3. Bivariate copula families

details)

$$h(u, v, \theta) = v^{-1} \exp \left\{ -[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} \right\} \left(1 + \left(\frac{-\ln u}{-\ln v} \right)^\theta \right)^{-1+1/\theta}$$

and the inverse of the h-function cannot be written in closed form. Therefore it must be obtained numerically using for instance the Newton Raphson method (see Aas et al. (2006)). For large-dimensional problems it might be better to use a different heavy right tail copula like the Clayton survival copula (see e.g. Joe (1997)).

Using Theorem 8 we can easily calculate Kendall's τ :

$$\begin{aligned} \rho_\tau(\theta) &= 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \\ &= 1 + 4 \int_0^1 \frac{-\ln(t)^\theta}{-\frac{1}{t}\theta \ln(t)^{\theta-1}} dt \\ &= 1 + 4 \int_0^1 \frac{t \ln t}{\theta} dt \\ &= 1 + \frac{4}{\theta} \left(\left[\frac{t^2}{2} \ln t \right]_0^1 - \int_0^1 \frac{t}{2} dt \right) \\ &= 1 + \frac{4}{\theta} \left(0 - \frac{1}{4} \right) = 1 - \frac{1}{\theta} \end{aligned}$$

As mentioned in Chapter 1 we can use Theorem 3 to calculate the tail dependence coefficients:

$$\begin{aligned} \lambda_U^{\theta, Gu} &= \lim_{u \rightarrow 1^-} \frac{1 - 2u + C_\theta^{Gu}(u, u)}{1 - u} = \lim_{u \rightarrow 1^-} \frac{1 - 2u + u^{2^{1/\theta}}}{1 - u} \\ &\stackrel{L'Hospital}{=} \lim_{u \rightarrow 1^-} \frac{-2 + 2^{1/\theta} u^{2^{1/\theta}-1}}{-1} = 2 - 2^{1/\theta} \\ \lambda_L^{\theta, Gu} &= \lim_{u \rightarrow 0^+} \frac{C_\theta^{Gu}(u, u)}{u} = \frac{u^{2^{1/\theta}}}{u} \\ &\stackrel{L'Hospital}{=} \lim_{u \rightarrow 0^+} \frac{2^{1/\theta} u^{2^{1/\theta}-1}}{1} = 0 \end{aligned}$$

3. Bivariate copula families

3.3.3. Frank copula

The Frank copula has a one-parametric generator $\phi(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$, which leads to $\phi^{-1}(t) = -\frac{1}{\theta}\ln(1 + e^{-t}(e^{-\theta}-1))$.

$$C_{\theta}^{Fr}(u, v) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right),$$

where $0 < \theta < \infty$ is the parameter controlling the dependence. By differentiating $C_{\theta}^{Fr}(u, v)$ with respect to u and v we obtain the copula density:

$$c_{\theta}^{Fr}(u, v) = \frac{\theta(1 - e^{-\theta})e^{-\theta(u+v)}}{(1 - e^{-\theta} - (1 - e^{\theta u})(1 - e^{\theta v}))^2}$$

For this copula the h-function is given as follows (see Appendix A.4.5 for details)

$$h(u, v, \theta) = \frac{e^{-\theta v}}{\frac{1-e^{-\theta}}{1-e^{-\theta u}} + e^{-\theta v} - 1}$$

and the inverse of the h-function is given by

$$h^{-1}(u, v, \theta) = -\ln\left(1 - \frac{1 - e^{-\theta}}{(u^{-1} - 1)e^{-\theta v} + 1}\right) / \theta.$$

According to Genest (1987) the Kendall's τ is given by:

$$\rho_{\tau}(\theta) = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt = 1 - 4\theta^{-1}(D_1(\theta) - 1),$$

where $D_k(\theta) = k/\theta^{-k} \int_0^{\theta} t^k / (\exp(t) - 1) dt$ is the Debye function.

As mentioned in Chapter 1 we can use Theorem 3 to calculate the tail dependence coefficients:

$$\begin{aligned} \lambda_U^{\theta, Fr} &= \lim_{u \rightarrow 1^-} \frac{1 - 2u + C_{\theta}^{Fr}(u, u)}{1 - u} = \lim_{u \rightarrow 1^-} \frac{1 - 2u - \frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u}-1)^2}{e^{-\theta}-1}\right)}{1 - u} \\ &\stackrel{L'Hospital}{=} \lim_{u \rightarrow 1^-} \frac{-2 - \frac{1}{\theta} \left(\frac{e^{-\theta}-1}{(e^{-\theta}-1)+(e^{-\theta u}-1)^2}\right) \frac{2(e^{-\theta u}-1)(-\theta)e^{-\theta u}}{e^{-\theta}-1}}{-1} = 0 \\ \lambda_L^{\theta, Fr} &= \lim_{u \rightarrow 0^+} \frac{C_{\theta}^{Fr}(u, u)}{u} \stackrel{L'Hospital}{=} \lim_{u \rightarrow 0^+} \frac{-\frac{1}{\theta} \left(\frac{e^{-\theta}-1}{(e^{-\theta}-1)+(e^{-\theta u}-1)^2}\right) \frac{2(e^{-\theta u}-1)(-\theta)e^{-\theta u}}{e^{-\theta}-1}}{1} = 0 \end{aligned}$$

3. Bivariate copula families

3.3.4. Joe copula

The Joe copula has a one-parametric generator $\phi(t) = -\ln(1 - (1 - t)^\theta)$, which leads to $\phi^{-1}(t) = 1 - (1 - \exp(-t))^{1/\theta}$.

$$C_\theta^{Jo}(u, v) = 1 - \left[(1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta(1 - v)^\theta \right]^{1/\theta},$$

where $1 \leq \theta < \infty$ is the parameter controlling the dependence. By differentiating $C_\theta^{Jo}(u, v)$ with respect to u and v we obtain the copula density:

$$c_\theta^{Jo}(u, v) = \left[(1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta(1 - v)^\theta \right]^{1/\theta-2} (1 - u)^{\theta-1} (1 - v)^{\theta-1} \\ \times [\theta - 1(1 + u)^\theta + (1 + v)^\theta - (1 - u)^\theta(1 - v)^\theta]$$

For this copula the h-function is given as follows (see Appendix A.4.6 for details)

$$h(u, v, \theta) = ((1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta(1 - v)^\theta)^{1/\theta-1} (1 - v)^{\theta-1} (1 - (1 - u)^\theta)^\theta$$

and the inverse of the h-function cannot be written in closed form. Therefore it must be obtained numerically.

According to Hofert, Mächler, and A. J. McNeil (2012) the Kendall's τ is given by:

$$\rho_\tau = 1 - 4 \sum_{k=1}^{\infty} \frac{1}{k(\theta k + 2)(\theta(k - 1) + 2)}$$

As mentioned in Chapter 1 we can use Theorem 3 to calculate the tail dependence coefficients:

$$\lambda_U^{\theta, Jo} = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C_\theta^{Jo}(u, u)}{1 - u} = \lim_{u \rightarrow 1^-} \frac{1 - 2u + 1 - [2(1 - u)^\theta - (1 - u)^{2\theta}]^{1/\theta}}{1 - u} \\ \stackrel{L'Hospital}{=} \lim_{u \rightarrow 1^-} \frac{-2 - \frac{1}{\theta}[2(1 - u)^\theta - (1 - u)^{2\theta}]^{1/\theta-1}(-2\theta(1 - u)^{\theta-1} + 2\theta(1 - u)^{2\theta-1})}{-1} \\ = 2 - 2^{1/\theta} \\ \lambda_L^{\theta, Jo} = \lim_{u \rightarrow 0^+} \frac{C_\theta^{Jo}(u, u)}{u} = \lim_{u \rightarrow 0^+} \frac{1 - [2(1 - u)^\theta - (1 - u)^{2\theta}]^{1/\theta}}{u} \\ \stackrel{L'Hospital}{=} \lim_{u \rightarrow 0^+} \frac{-\frac{1}{\theta}[2(1 - u)^\theta - (1 - u)^{2\theta}]^{1/\theta-1}(-2\theta(1 - u)^{\theta-1} + 2\theta(1 - u)^{2\theta-1})}{1} = 0$$

3.4. Simulation of Archimedean and elliptical copulas

In this section the simulation (also known as sampling) from a specific copula family will be described. The common way of simulating copulas is based on specific techniques for each class of copula.

3.4.1. Simulation of elliptical copulas

The following section is based on Chapter 5 of Embrechts, Lindskog, and A. McNeil (2003).

In the sampling for Gaussian and Student's t-copula we use the Cholesky decomposition $LL^T = R$ of a given correlation matrix $R \in \mathbb{R}^{n \times n}$ to obtain a lower triangular matrix $L \in \mathbb{R}^{n \times n}$.

In the case of independence of $Z_1, \dots, Z_n \sim N(0, 1)$ it follows that $X := LZ \sim N_n(0, R)$. Using Theorem 4 (Sklar's Theorem) and the definition of the Gaussian copula given in Section 3.2.1 this leads to

$$\begin{aligned} \mathbb{P}(U_1 \leq u_1, \dots, U_n \leq u_n) &= \mathbb{P}(\phi(X_1) \leq x_1, \dots, \phi(X_n) \leq x_n) \\ &= \mathbb{P}(X_1 \leq \phi^{-1}(x_1), \dots, X_n \leq \phi^{-1}(x_n)) \\ &= \phi_R(\phi^{-1}(u_1), \dots, \phi^{-1}(u_n)) \\ &\stackrel{\substack{\text{Sklar's Theorem} \\ \text{def. Gaussian copula}}}{=} C_R^{Ga}, \end{aligned}$$

where ϕ_R denotes the joint cumulative distribution function of X .

Algorithm 3.4.1 Algorithm for generating a random vector $U = (U_1, \dots, U_n)$ with distribution function of a Gaussian copula C_R^{Ga} .

- 1: Compute the Cholesky decomposition L from R : $R = LL^T$.
 - 2: Sample i.i.d. $Z_1, \dots, Z_n \sim N(0, 1)$.
 - 3: Calculate $X_i = \sum_{j=1}^i l_{ij}Z_j, \forall i \in \{1, \dots, n\}$. Hence $X = LZ \sim N_n(0, R)$.
 - 4: Return (U_1, \dots, U_n) , where $U_i = \Phi(X_i) \sim U(0, 1), \forall i \in \{1, \dots, n\}$.
-

3. Bivariate copula families

In case of the Student's t-copula the sampling algorithm can be derived in a similar way. For the stochastic representation of a Student's t-distribution given the independence of $Z_1, \dots, Z_n \sim N(0, 1)$ and $S \sim \chi_v^2$ it follows that $X := \sqrt{\frac{v}{S}} LZ \sim t_n(v, 0, R)$. Using Theorem 4 (Sklar's Theorem) and the definition of the Student's t-copula given in Section 3.2.2 leads to the desired result.

Algorithm 3.4.2 Algorithm for generating a random vector $U = (U_1, \dots, U_n)$. with distribution function of a Student's t-copula $C_{v,R}^t$.

- 1: Compute the Cholesky decomposition L from R : $R = LL^T$.
 - 2: Sample i.i.d. $Z_1, \dots, Z_n \sim N(0, 1)$.
 - 3: Sample a random variable $S \sim \chi_v^2$ independent of Z_1, \dots, Z_n .
 - 4: Calculate $Y_i = \sum_{j=1}^i l_{ij} Z_j, \forall i \in \{1, \dots, n\}$. Hence $Y = LZ \sim N_n(0, R)$.
 - 5: Calculate $X_i = \sqrt{\frac{v}{S}} Y_i, \forall i \in \{1, \dots, n\}$. Hence $X \sim T_{v,n}(0, R)$.
 - 6: Return (U_1, \dots, U_n) , where $U_i = t_v(X_i) \sim U(0, 1), \forall i \in \{1, \dots, n\}$.
-

3.4.2. Simulation of Archimedean copulas

The following section is based on Hofert (2008) and Subsection 5.4.2 of A. J. McNeil, Frey, and Embrechts (2006).

The considered Archimedean copulas fall into the class of so-called *Laplace transform Archimedean copulas* (or LT-Archimedean copulas).

Definition 18 (Complete monotonicity)

A function $g : [0, \infty) \rightarrow [0, \infty)$ is called *completely monotone*, if g is continuous on $[0, \infty)$, infinitely differentiable on $(0, \infty)$ and satisfies

$$(-1)^k \left(\frac{d^k}{ds^k} g(s) \right) \Big|_{s=t} \geq 0, \text{ for } k = 0, 1, 2, \dots \text{ and } \forall t \in (0, \infty).$$

Theorem 9 (Kimberling (1974))

Let $\phi : [0, 1] \rightarrow [0, \infty]$ be continuous and strictly decreasing with $\phi(0) = \infty$ and $\phi(1) = 0$. The function $C : [0, 1]^n \rightarrow [0, 1]$, $C(u) = \phi^{-1}(\phi(u_1) + \phi(u_2) + \dots +$

3. Bivariate copula families

$\phi(u_n)$), is a copula for any $n \geq 2$ if and only if $\phi^{-1} : [0, \infty] \rightarrow [0, 1]$ is completely monotone in $[0, \infty)$.

Theorem 10 (Bernstein)

A function $\psi : [0, \infty) \rightarrow [0, \infty)$ is the Laplace-Stieltjes transform of a distribution function F in $[0, \infty)$ ($\psi(s) = \int_0^\infty e^{-sx} dF(x)$, $s \geq 0$) if and only if ψ is completely monotone and $\psi(0) = 1$.

The class of all completely monotone Archimedean generators is denoted by Ψ_∞ . Bernstein's theorem establishes a link between completely monotone functions and Laplace-Stieltjes transforms. Using Bernstein's theorem it is clear that this class coincides with the class of Laplace-Stieltjes transforms of distribution functions F on the positive real line. The Laplace-Stieltjes transform of F is defined as

$$\mathcal{LS}[F](t) := \int_0^\infty e^{-tx} dF(x), \text{ for } t \geq 0. \quad (3.7)$$

For $\psi \in \Psi_\infty$, we hence have the relation

$$\psi = \mathcal{LS}[F],$$

or equivalently

$$F = \mathcal{LS}^{-1}[\psi].$$

For the sampling of a n -dimensional exchangeable Archimedean copula with generator ψ we use the algorithm described in Marshall and Olkin (1988), where $\mathcal{LS}^{-1}[\psi]$ denotes the inverse Laplace-Stieltjes transform of ψ .

Theorem 11

Let F be a distribution function on $[0, \infty)$ satisfying $F(0) = 0$ with Laplace-Stieltjes transform ψ as in (3.7) and set $\psi(\infty) := 0$. Let V be a random variable with distribution function F and let U_1, \dots, U_n be a sequence of random variables that are conditionally independent given V with conditional distribution function given by

$$F_{U_i|V}(u|v) = \exp(-v\psi^{-1}(u)), \text{ for } u \in [0, 1].$$

Then

$$P(U_1 \leq u_1, \dots, U_n \leq u_n) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_n)),$$

3. Bivariate copula families

so that the distribution function of $U = (U_1, \dots, U_n)$ is a Archimedean copula with generator $\phi = \psi^{-1}$.

Proof. The proof to this theorem can be found in Proposition 5.46 of A. J. McNeil, Frey, and Embrechts (2006). \square

Algorithm 3.4.3 Algorithm for generating a random vector $U = (U_1, \dots, U_n)$. with Archimedean copula $C(u) = \phi^{-1}(\phi(u_1) + \dots + \phi(u_n))$ with generator ϕ as distribution function.

- 1: Sample $V \sim F = \mathcal{L}\mathcal{S}^{-1}[\psi]$ where $\psi = \phi^{-1}$
 - 2: Sample i.i.d $V_1, \dots, V_n \sim U(0, 1)$.
 - 3: Return (U_1, \dots, U_n) , where $U_i = \psi\left(-\frac{\ln(V_i)}{V}\right), \forall i \in \{1, \dots, n\}$.
-

This provides a powerful tool for sampling exchangeable Archimedean copula, given that we know how to sample F . This algorithm is efficient even in larger dimensions. It requires only $n + 1$ random numbers for generating a n -dimensional observation. Regardless of dimensions only one sample $V \sim F$ is required.

Archimedean copula	inv. Laplace-Stiljtjes transf. F^1	inverse generator $\phi^{-1}(t) = \psi(t)$
Clayton	$\Gamma(1/\theta, 1)$	$(1 + t)^{-1/\theta_2}$
Gumbel ³	$St(1/\theta, 1, (\cos(\frac{\pi}{2\theta}))^\theta, 0)$	$\exp(-t^{1/\theta})$
Frank	$y_k = \frac{(1-e^{-\theta})^k}{k\theta}, k \in \mathbb{N}$	$-\frac{1}{\theta} \log(1 + e^{-t}(e^{-\theta} - 1))$
Joe	$y_k = (-1)^{k+1} \binom{1/\theta}{k}, k \in \mathbb{N}$	$1 - (1 - \exp(-t))^{1/\theta}$

Table 3.1.: Overview of selected Archimedean families with corresponding inverse Laplace-Stiljtjes transforms $F = \mathcal{L}\mathcal{S}^{-1}[\psi]$ and inverse generators $\phi^{-1}(t) = \psi(t)$.

¹The Laplace-Stiljtjes transforms are given in Hofert (2008). Further details to the inverse Laplace-Stiljtjes transforms F can be found in Appendix A.5.

²Here we use a alternative but equivalent representation of the generator of the Clayton copula.

³Although the closed form of the density of stable distribution is not known, Nolan (2011) proposed a simulation algorithm for generating stable random variables.

3. Bivariate copula families

For Frank's copula and Joe's copula F is discrete with a given probability mass function $(y_k)_{k \in \mathbb{N}}$. Unfortunately, there is no known way to find F explicitly. In the case where F is a step function we use the following theorem.

Theorem 12

Let $\psi \in \Psi_\infty$ with $F = \mathcal{LS}^{-1}[\psi]$ and let $G(x) = \sum_{k=0}^{\infty} y_k \mathbb{1}_{[x_k, \infty)}(x)$, with $0 < x_0 < x_1 < \dots$ and $y_k > 0, k \in \mathbb{N}_0$, with $\sum_{k=0}^{\infty} y_k = 1$. Then

$$F \equiv G \Leftrightarrow \psi(t) = \sum_{k=0}^{\infty} y_k e^{-x_k t}, t \in [0, \infty].$$

Proof. The proof to this theorem can be found in Hofert (2008). □

For further details on the subject matter of sampling a larger variety of Archimedean copulas see Jaworski et al. (2010).

3.5. Summary of presented copula families and their characteristics

Copula	Parameters	Kendall's τ	Tail-dependence (lower, upper)
Gaussian	$\rho \in (-1, 1)$	$\frac{2}{\pi} \arcsin(\rho)$	$(0, 0)$
Student's t	$\rho \in (-1, 1)$ $\nu > 2$	$\frac{2}{\pi} \arcsin(\rho)$	$\left(2t_{\nu+1} \left(-\sqrt{\nu+1} \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right), 2t_{\nu+1} \left(-\sqrt{\nu+1} \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) \right)$
Clayton	$\theta \in (0, \infty)$	$\frac{\theta}{2+\theta}$	$(2^{-1/\theta}, 0)$
Gumbel	$\theta \in [1, \infty)$	$1 - \frac{1}{\theta}$	$(0, 2 - 2^{1/\theta})$
Frank	$\theta \in (0, \infty)$	$1 - \frac{4}{\theta} (D_1(\theta) - 1)$ with $D_1(\theta) = \theta^{-1} \int_0^\theta t / (\exp(t) - 1) dt$	$(0, 0)$
Joe	$\theta \in [1, \infty)$	$1 - 4 \sum_{k=1}^{\infty} 1 / [k(\theta k + 2)(\theta(k-1) + 2)]$	$(0, 2 - 2^{1/\theta})$

Table 3.2.: Summary of pair-copula families and their characteristics. The first two are from the elliptical copula class, the others from the Archimedean copula class.

3. Bivariate copula families

Archimedean copula	generator $\phi(t)$	inverse generator $\phi^{-1}(t)$
Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$(1 + \theta t)^{-1/\theta}$
Gumbel	$(-\log(t))^\theta$	$\exp(-t^{1/\theta})$
Frank	$-\log\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$	$-\frac{1}{\theta} \log(1 + e^{-t}(e^{-\theta} - 1))$
Joe	$-\log(1 - (1 - t)^\theta)$	$1 - (1 - \exp(-t))^{1/\theta}$

Table 3.3.: Overview of the generators $\phi(t)$ and their inverse of all four Archimedean copulas described in this thesis.

In the following chapter we will introduce a concept called *pair-copula-constructions (PCC)*, which is a general construction method for multivariate copulas using only pair-copulas as building blocs (see Section 4.1). By using PCCs it is possible to capture different types of dependencies and tail behavior by utilizing distinct properties of pair-copulas. Table 3.4 illustrates that the Clayton, Gumbel and Joe copula have tail asymmetry and therefore different properties of upper and lower tail dependence. The Gaussian, Student's t and Frank copula all feature positive and negative dependence. With this set of copulas as we will see in the next chapter it is possible to capture different types of dependence when using PCCs.

	N	t	Cl	Gu	Fr	Jo
Tail asymmetry	-	-	✓	✓	-	✓
Upper tail dependence	-	✓	-	✓	-	✓
Lower tail dependence	-	✓	✓	-	-	-
Positive dependence	✓	✓	✓	✓	✓	✓
Negative dependence	✓	✓	-	-	✓	-

Table 3.4.: Overview of dependencies in the bivariate copula families described in this thesis.

Figure 3.3 gives an overview of all six described copulas using two different dependence parameters for each copula to illustrate the differences in the dependence structure.

3. Bivariate copula families

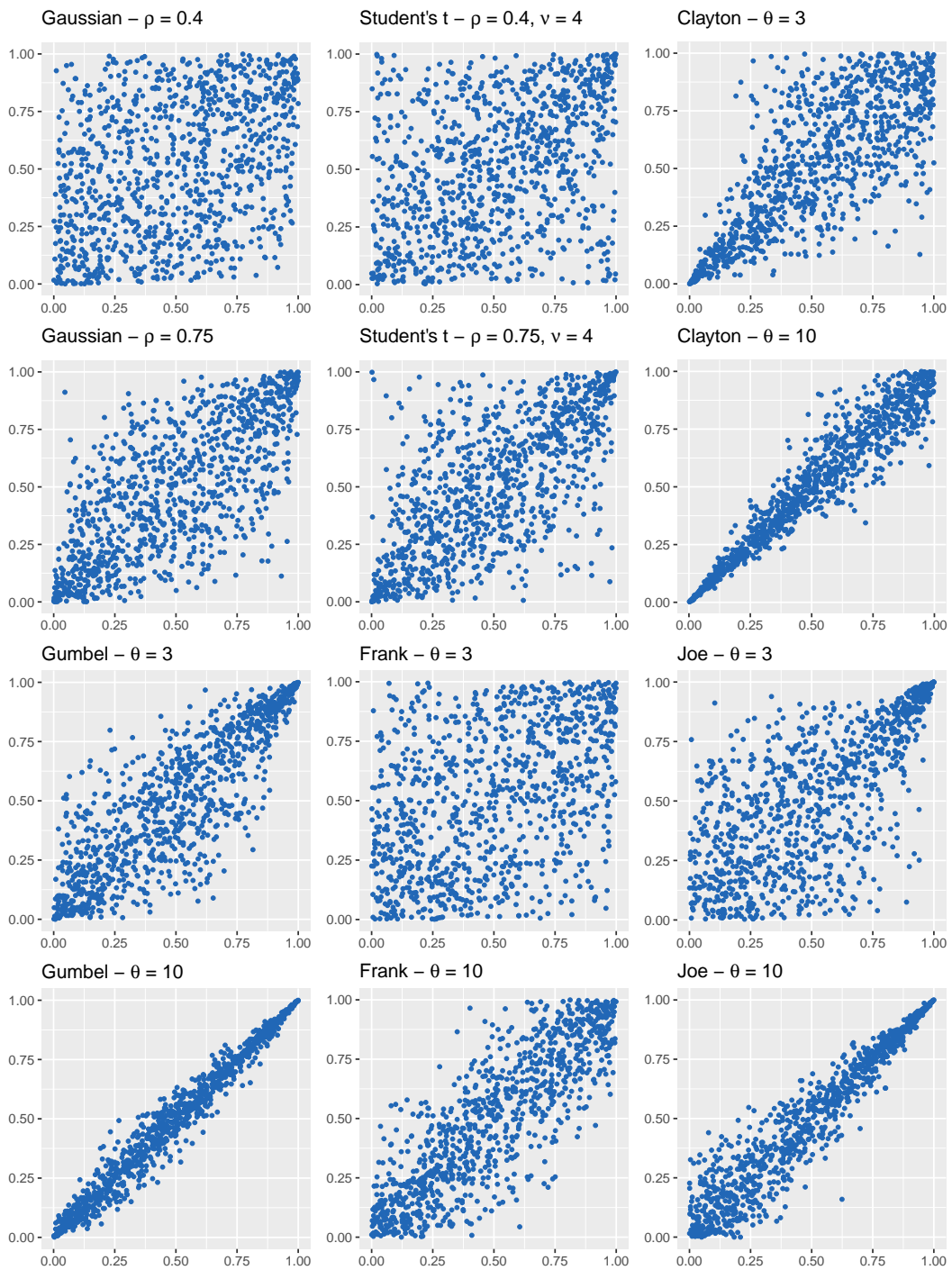


Figure 3.3.: Illustration of differences in dependence structure for all presented copulas. Each copula is depicted with two different dependence parameters to illustrate the change in dependence structure.

4. Vine copula

Even though it is possible to construct multivariate copulas with more than two variables, in practice, when modeling a large financial data set of which not all pairs of risk-factors have the same dependence structure, higher-dimensional copulas are sometimes not sufficient to capture the dependence structure by using just one or two parameters. Standard multivariate copulas do not allow for different dependence structures between different pairs of variables. This chapter will introduce a method called *pair-copula-construction* (PCC), which can be used to model a joint probability distribution function from bivariate copulas and marginal densities. PCCs are in general more flexible than multivariate copulas to capture dependence structure among assets. Bivariate copulas as described in Chapter 3 serve as building blocs for higher-dimensional distributions.

Joe (1996) introduced the first construction of a multivariate copula using bivariate copulas and marginal distributions. Later Bedford and R. Cooke (2002) extended this work to a more general construction method of multivariate densities and subsequently with the introduction of a graphical model called *the regular vine* made it possible to describe the structure of a PCC in a more organized and convenient way. Regular vine is a graphical concept with the goal of capturing the most important dependencies between pairs of variables by using a nested set of trees, where each edge corresponds to a bivariate copula density and the edge labels correspond to the subscript of the bivariate copula density, e.g. edge 12|3 corresponds to the copula density $c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3))$.

There are quite a number of possible PCCs for high-dimensional distributions. According to Morales-Nápoles, R. M. Cooke, and Kurowicka (2010)

4. Vine copula

for a six-dimensional density function there are already 23,040¹ possible regular vines. That's why, *canonical vines* (see Definition 26) and *drawable vines* (see Definition 27), two special cases of *regular vines*, have recently attracted special attention (see Aas et al. (2006)). Both vines impose additional restriction on the structure and therefore limit the number of different models. It can be shown that the number of *canonical vines* on n nodes equals the number of *drawable vines* on n nodes and is $\frac{n!}{2}$ (see Lemma 3.3 in Morales-Nápoles, R. M. Cooke, and Kurowicka (2010)).

Our goal is to find the vine model which depicts the dependence structure most accurately. From now on regular, canonical and drawable vines will be called R-vines, C-vines and D-vines.

4.1. Pair-copula-construction - PCC

Large portions of this section are based on the work of Aas et al. (2006) and Krämer and Schepsmeier (2011). To be able to better understand this concept we will first illustrate a 3-dimensional PCC. Later we will state the generalization for higher dimensions. The concept of PCC is based on the decomposition of an n -dimensional joint density function f

$$\begin{aligned} f(x_1, \dots, x_n) \\ = f_1(x_1) f_{2|1}(x_2|x_1) f_{3|12}(x_3|x_1, x_2) \dots f_{n|1\dots n-1}(x_n|x_1, \dots, x_{n-1}). \end{aligned} \quad (4.1)$$

3-dimensional example of a PCC:

Our goal is to express the probability density function $f(x_1, x_2, x_3)$ as a product of pair-copula densities and marginal densities $f_1(x_1), f_2(x_2), f_3(x_3)$. One possible decomposition of $f(x_1, x_2, x_3)$ would be:

$$f(x_1, x_2, x_3) = \underbrace{f_1(x_1)}_{\text{part 1}} \cdot \underbrace{f_{2|1}(x_2|x_1)}_{\text{part 2}} \cdot \underbrace{f_{3|12}(x_3|x_1, x_2)}_{\text{part 3}} \quad (4.2)$$

¹Morales-Nápoles, R. M. Cooke, and Kurowicka (2010) showed that there are $\frac{n!}{2} \times 2^{\binom{n-2}{2}}$ possible R-vines on n nodes.

4. Vine copula

Furthermore, the conditional density can be expressed as a copula function.

Starting with **part 2** we get:

$$\begin{aligned}
 f_{2|1}(x_2|x_1) &= \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)} \\
 &\stackrel{(3.3)}{=} \frac{f_1(x_1)f_2(x_2)c_{12}(F_1(x_1), F_2(x_2))}{f_1(x_1)} \\
 &= f_2(x_2)c_{12}(F_1(x_1), F_2(x_2))
 \end{aligned} \tag{4.3}$$

For **part 3** one possible decomposition would be:

$$\begin{aligned}
 f_{3|12}(x_3|x_1, x_2) &= \frac{f(x_1, x_2, x_3)}{f_{12}(x_1, x_2)} = \frac{f_{13|2}(x_1, x_3|x_2)f_2(x_2)}{f_{1|2}(x_1|x_2)f_2(x_2)} = \frac{f_{13|2}(x_1, x_3|x_2)}{f_{1|2}(x_1|x_2)} \\
 &= \frac{f_{3|2}(x_3|x_2)f_{1|2}(x_1|x_2)c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)}{f_{1|2}(x_1|x_2)} \\
 &= f_{3|2}(x_3|x_2)c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2) \\
 &\stackrel{(4.3)}{=} f_3(x_3)c_{23}(F_2(x_2), F_3(x_3))c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)
 \end{aligned}$$

An alternative decomposition where we choose $c_{23|1}$ instead of $c_{13|2}$ is:

$$\begin{aligned}
 f_{3|12}(x_3|x_1, x_2) &= f_{3|1}(x_3|x_1)c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1) \\
 &= f_3(x_3)c_{13}(F_1(x_1), F_3(x_3))c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1)
 \end{aligned}$$

As we see in this example the decomposition is not unique. In fact the number of decompositions grows rapidly with the dimension. For a 5-dimensional density we already have as many as 240 different PCCs (see Hobæk Haff, Aas, and Frigessi (2010)).

In general $c_{31|2}$ depends on the conditioning variable x_2 . This dependence consists not only through its arguments $F_{1|2}(x_1|x_2)$ and $F_{3|2}(x_3|x_2)$ but also through x_2 directly. For purely practical reasons Hobæk Haff, Aas, and Frigessi (2010) proposed to leave the *full PCC*, where all pair-copulas can be dependent directly on a conditioning variable, behind and introduced the *simplified PCC* instead, which can be used as a good approximation to the

4. Vine copula

correct decomposition. We therefore assume that pair-copulas are independent of conditioning variables, except through the conditional distribution. Now that we have calculated all three parts in (4.2) using simplified PCC leads to:

$$\begin{aligned}
 f(x_1, x_2, x_3) &= \overbrace{f_1(x_1)}^{\text{part 1}} \overbrace{f_{2|1}(x_2|x_1)}^{\text{part 2}} \overbrace{f_{3|12}(x_3|x_1, x_2)}^{\text{part 3}} \\
 &= f_1(x_1) f_2(x_2) c_{12}(F_1(x_1), F_2(x_2)) \\
 &\quad \times f_3(x_3) c_{23}(F_2(x_2), F_3(x_3)) c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \\
 &= f_1(x_1) f_2(x_2) f_3(x_3) \quad \text{(marginals)} \\
 &\quad \times c_{12}(F_1(x_1), F_2(x_2)) c_{23}(F_2(x_2), F_3(x_3)) \quad \text{(unconditional pairs)} \\
 &\quad \times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \quad \text{(conditional pair)}
 \end{aligned}$$

Now let $\mathbf{v} = (v_1, \dots, v_n)$ be a n -dimensional vector, v_j one component of this vector and $\mathbf{v}_{-j} = (v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n)$ the vector \mathbf{v} excluding this component. According to Aas et al. (2006), we can generalize the factorization from above to

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

This iterative construction can lead to many different possible pair-copula-constructions. Applying this to the factorization of $f(x_1, \dots, x_n)$ from (4.1) we can now express the univariate density using a combination of bivariate copulas marginal densities.

Using $f(x|v) = c_{xv}(F_x(x), F_v(v)) f_x(v)$, for the univariate case where v consists only of one component this leads to

$$\begin{aligned}
 F(x|v) &= \int_{-\infty}^x c_{xv}(F_x(u), F_v(v)) f_x(u) du \\
 &= \int_{-\infty}^x \frac{\partial C_{xv}(F_x(u), F_v(v))}{\partial F_x(u) \partial F_v(v)} f_x(u) du \\
 &= \frac{\partial C_{xv}(F_x(x), F_v(v))}{\partial F_v(v)},
 \end{aligned}$$

Joe (1996) showed that for every j we get

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

4. Vine copula

where $C_{ij|v_{-j}}$ is an arbitrary 2-dimensional copula distribution function.

If x and v are uniformly distributed i.e. $f(x) = f(v) = 1$, $F(x) = x$ and $F(v) = v$ we will use the function $h(x, v, \Theta)$ to represent the conditional distribution function, which proves to be very useful for simulation and construction of vine copulas. This function is often referred to as h-function.

$$h(x, v, \Theta) = F(x|v) = \frac{\partial C_{xv}(x, v, \Theta)}{\partial v} \quad (4.5)$$

where Θ represents the set of parameters for the copula associated with the joint distribution of x and v . The second argument of $h(\cdot)$ denotes the conditioning variable. Let $h^{-1}(x, v, \Theta)$ be the inverse of the conditional distribution function with respect to the first variable x , hence the inverse of the conditional distribution function.

4.2. Introduction to vines

Vines are a graphical model used for the representation of simplified PCC by using graph theory. Hence this graphical model can be used to specify multivariate distributions by specifying its marginal distributions and the way in which the marginal distributions are coupled.

First we will introduce some basic concepts from graph theory needed for the introduction of regular vines, while at the same time limiting graph theory to the bare minimum necessary for this thesis.

Definition 19 (Graph, edge, node)

A pair $G = (N, E)$ of sets such that $E \subseteq \{\{i, j\} : i, j \in N, i \neq j\}$ is called graph. The elements of E are called edges of the graph G whereas the elements of N are called nodes of the graph G .

Definition 20 (Path, cycle)

A graph $P = (N, E)$ with $E = \{\{n_0, n_1\}, \{n_1, n_2\}, \dots, \{n_{k-1}, n_k\}\}$ and $N = \{n_0, n_1, \dots, n_k\}$ is called path. A cycle is a path with $n_0 = n_k$.

Definition 21 (Degree)

Let $G = (N, E)$ be a graph. The degree of a node of a graph is the number edges incident to the node, i.e. the number of its neighbors.

4. Vine copula

Definition 22 (Connected)

Let $G = (N, E)$ be a graph. Two nodes $n_i, n_j \in N$, $n_i \neq n_j$ are connected if there is a path from n_i to n_j . If every node in a graph is connected to every other node we call a graph connected.

Definition 23 (Tree)

A Tree is a graph $T = (N, E)$ that has no cycles and is connected.

Definition 24 (Star, root node)

A tree $T = (N, E)$, with a node n_0 with degree $|N| - 1$ is called a star and n_0 is called the root node. $|N|$ refers to the number of nodes in tree T .

Definition 25 (R-vine)

A n -dimensional regular vine $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$ is a sequence of $n - 1$ trees with nodes N_i and edges E_i with:

1. Tree $T_1 = (N_1, E_1)$ has nodes $N_1 = \{1, \dots, n\}$ and a set of edges E_1
2. For $i = 2, \dots, n - 1$ tree $T_i = (N_i, E_i)$ has nodes $N_i = E_{i-1}$.
3. Proximity condition: For $i = 2, \dots, n - 1$, if $\{a, b\} \in E_i$, with $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$, then it must hold that $|a \cap b| = 1$, i.e. exactly one of the a_j 's equals one of the b_j 's.

The proximity condition assures that if there is an edge in E_i connecting two nodes a and b , then those two nodes are edges in tree T_{i-1} and they share a common node.

The class of R-vines is very general and consists of a very large number of possible pair-copula-decompositions. A way to reduce the number of R-vines is the usage of two special cases of R-vines, the C-vine and the D-vine, which have both been discussed extensively in literature.

Definition 26 (C-vine)

C-vines are regular vines for which each tree T_i , for $i = 1, \dots, n - 1$, is a star. Hence a C-vine has a unique node with degree $n - i$. This unique node is called the root node of tree T_i .

Definition 27 (D-vine)

D-vines are regular vines for which the first tree T_1 is a path. Hence in a D-vine each node in T_1 has a degree of at most 2.

4. Vine copula

Definition 28 (R-vine copula specification - Bedford and R. Cooke (2002))
 (F, \mathcal{V}, B) is called a R-vine copula specification if:

1. $F = (F_1, \dots, F_n)$ is a vector of continuous invertible distribution functions,
2. \mathcal{V} is an n -dimensional R-vine and
3. $B = \{B_e | i = 1, \dots, n-1; e \in E_i\}$ is a set of copulas, with B_e being a pair-copula and E_i the set of edges of tree T_i of the R-vine \mathcal{V} .

On the following pages we will give a short profile including a 5-dimensional example of each of the three types of vines used in this thesis. Bedford and R. Cooke (2001) states the density of an multi-dimensional distribution in terms of a R-vine density, which Aas et al. (2006) specialized to a C-vine and D-vine.

We will use the following abbreviations in order to save some space and to visualize the densities more clearly:

- For a bivariate conditional copula density of X_i and X_j given x_{i_1}, \dots, x_{i_n} for arbitrary distinct indices i, j, i_1, \dots, i_n with $i < j$ and $i_1 < \dots < i_n$ we use:

$$c_{i,j|i_1, \dots, i_n} := c_{i,j|i_1, \dots, i_n}(F(x_i|x_{i_1}, \dots, x_{i_k}), F(x_j|x_{i_1}, \dots, x_{i_k}))$$

- For the marginal density of X_i we use:

$$f_i := f_i(x_i)$$

Definition 29 (Complete union, conditioning set)

Let \mathcal{V} be a n -dimensional regular vine and $e \in \{i(e), j(e)\} \in E_k, i(e), j(e) \in E_{k-1}, k = 1, \dots, n-1$ an arbitrary edge.

(i) The complete union of e is defined by

$$U_e = \{n \in N_1 | \exists e_1 \in E_1, \dots, e_{k-1} \in E_{k-1} \text{ with } n \in e_1 \in \dots, \in e_{k-1} \in e\}$$

U_e is a set of nodes in N_1 that are in some way "connected" by edge e , i.e. all elements of U_e are connected by an edge, that is enclosed in an edge, in e .

(ii) The conditioning set of e is defined by

$$D(e) = U_{i(e)} \cap U_{j(e)}.$$

For more details see Dißmann (2010) and Bedford and R. Cooke (2001).

4. Vine copula

R-vine:

For a n -dimensional density $f(x_1, \dots, x_n)$ the PCC of the **R-vine** is given in Bedford and R. Cooke (2001) as follows:

$$f(\mathbf{x}) = \underbrace{\prod_{k=1}^n f_k(x_k)}_{\text{marginal densities}} \underbrace{\prod_{m=1}^{n-1} \prod_{e \in E_m} c_{i(e),j(e)|D(e)}(F(x_{i(e)}|x_{D(e)}), F(x_{j(e)}|x_{D(e)}))}_{\text{pair-copula densities}},$$

where E_m denotes the edge set of tree T_m , $e = i(e), j(e)|D(e)$ are edges in E_m with conditioning set $D(e)$ and $c_{i(e),j(e)|D(e)}$ is the copula density associated with e . $x_{D(e)}$ stands for the variables in $D(e)$, i.e. $x_{D(e)} = \{x_i | i \in D(e)\}$.

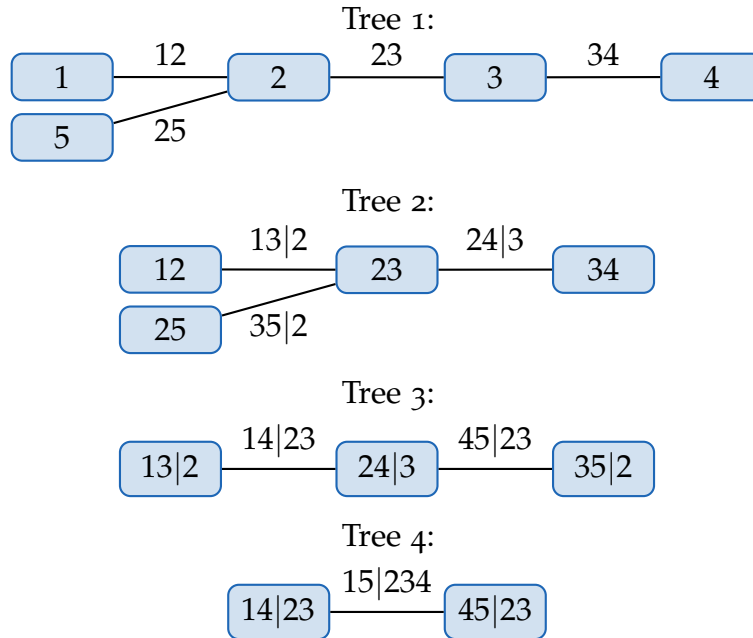


Figure 4.1.: An example R-vine on five variables, four trees and ten pair-copulas.

The density represented in Figure 4.1 can be written as:

$$f_{12345} = \underbrace{f_1 f_2 f_3 f_4 f_5}_{\text{nodes in tree 1}} \underbrace{c_{12} c_{23} c_{34} c_{25}}_{\substack{\text{edges in tree 1} \\ \text{nodes in tree 2}}} \underbrace{c_{24|3} c_{35|1} c_{13|2}}_{\substack{\text{edges in tree 2} \\ \text{nodes in tree 3}}} \underbrace{c_{14|23} c_{45|23}}_{\substack{\text{edges in tree 3} \\ \text{nodes in tree 4}}} \underbrace{c_{15|234}}_{\text{edges in tree 4}}$$

4. Vine copula

D-vine:

For a n -dimensional density $f(x_1, \dots, x_n)$ the PCC of the **D-vine** is given in Aas et al. (2006) as follows:

$$f(\mathbf{x}) = \underbrace{\prod_{k=1}^n f_k(x_k)}_{\text{marginal densities}} \underbrace{\prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|(i+1):(i+j-1)}(F(x_i|x_{(i+1):(i+j-1)}), F(x_{i+j}|x_{(i+1):(i+j-1)}))}_{\text{pair-copula densities}}$$

In a D-vine every node in any Tree T_j is assigned a degree of 1 or 2.

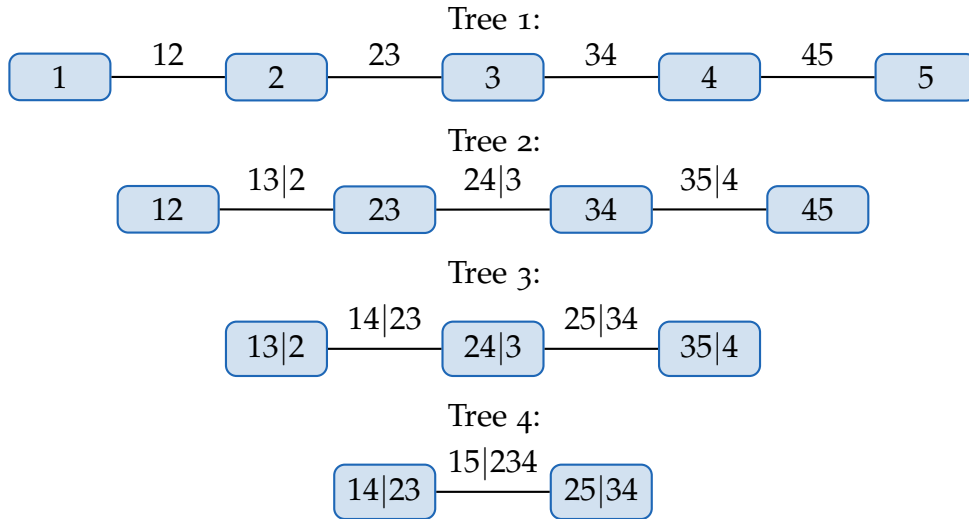


Figure 4.2.: An example D-vine on five variables, four trees and ten pair-copulas.

The density represented in Figure 4.2 can be written as:

$$f_{12345} = \underbrace{f_1 f_2 f_3 f_4 f_5}_{\text{nodes in tree 1}} \underbrace{c_{12} c_{23} c_{34} c_{45}}_{\substack{\text{edges in tree 1} \\ \text{nodes in tree 2}}} \underbrace{c_{13|2} c_{24|3} c_{35|4}}_{\substack{\text{edges in tree 2} \\ \text{nodes in tree 3}}} \underbrace{c_{14|23} c_{25|34}}_{\substack{\text{edges in tree 3} \\ \text{nodes in tree 4}}} \underbrace{c_{15|234}}_{\text{edges in tree 4}}$$

4. Vine copula

C-vine:

For a n -dimensional density $f(x_1, \dots, x_n)$ the PCC of the **C-vine** is given in Aas et al. (2006) as follows:

$$f(\mathbf{x}) = \underbrace{\prod_{k=1}^n f_k(x_k)}_{\text{marginal densities}} \underbrace{\prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j,j+i|1:(j-1)}(F(x_j|x_{1:(j-1)}), F(x_{j+i}|x_{1:(j-1)}))}_{\text{pair-copula densities}}.$$

In a C-vine each tree T_j has one unique node with degree $n - j$. The C-vine might be advantageous in a scenario where one particular variable acts as the key component that governs interactions in the dataset and is associated with all variables in the group. This key variable (variable 1 in Figure 4.3) is then used as root of the C-vine.

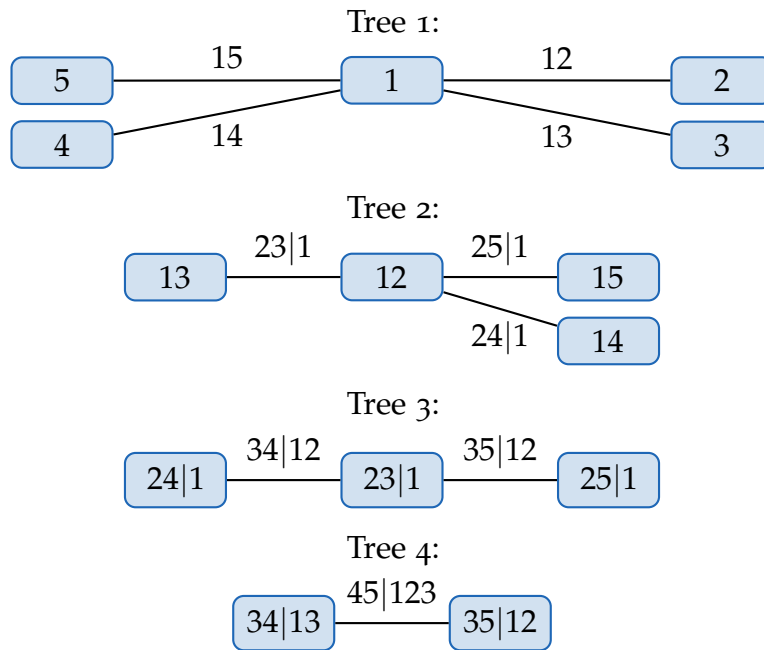


Figure 4.3.: An example C-vine on five variables, four trees and ten pair-copulas.

4. Vine copula

The density represented in Figure 4.3 can be written as:

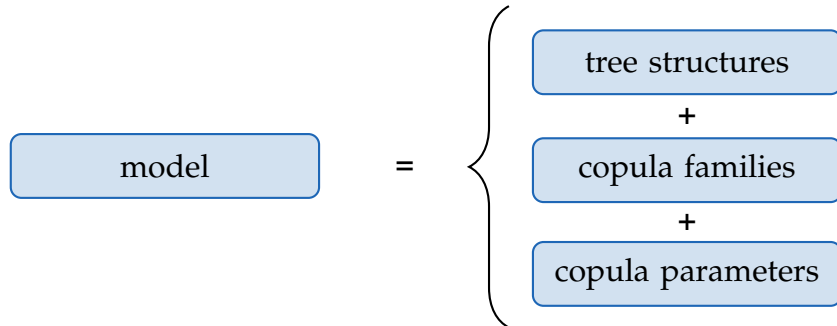
$$f_{12345} = \underbrace{f_1 f_2 f_3 f_4 f_5}_{\substack{\text{nodes in tree 1} \\ \text{nodes in tree 2}}} \underbrace{c_{12} c_{13} c_{14} c_{15}}_{\substack{\text{edges in tree 1} \\ \text{nodes in tree 2}}} \underbrace{c_{23|1} c_{24|1} c_{25|1}}_{\substack{\text{edges in tree 2} \\ \text{nodes in tree 3}}} \underbrace{c_{34|12} c_{35|12}}_{\substack{\text{edges in tree 3} \\ \text{nodes in tree 4}}} \underbrace{c_{45|123}}_{\text{edges in tree 4}}$$

All three vine structures have different properties in representing the structure of a multi-dimensional portfolio.

4.3. Vine copula specification

It takes three steps to fit a R-vine copula specification to a given dataset:

1. **Tree structure selection:** Selection which conditioned and unconditioned pairs to use for the R-vine.
2. **Copula family selection:** Choice of bivariate copula family for each selected pair.
3. **Parameter estimation:** Estimation of parameter(s) for each chosen copula.



One possibility would be to accomplish step 2 and 3 for all possible R-vine structure selections. Obviously, this approach is not very practical since Morales-Nápoles, R. M. Cooke, and Kurowicka (2010) showed that there are $n!/2 \times 2^{\binom{n-2}{2}}$ possible R-vines on n nodes.

4. Vine copula

Nodes [n]	3	4	5	6	7	8
number of different R-vines	3	24	480	23.040	2.580.480	660.602.880

Table 4.1.: Number of regular vines on 3, 4, 5, 6, 7 and 8 nodes.

Due to the enormous number of different R-vines with increasing node count as seen in Table 4.1 we will need to turn to rather heuristic methods used in graph theory to estimate the best structure.

4.3.1. Tree structure selection

For a given dataset it is necessary to determine for which pair of variables we want to specify a copula. Dißmann (2010) described a method called "Sequential method to select an R-vine model based on Kendall's τ " (see Algorithm 4.3.1) which is based on the idea that the copula families specified in the first tree of the R-vine often have the greatest influence on the precision of the model. This means that the model prioritizes strong dependence in the first tree. The method proceeds sequentially tree by tree starting with the first tree $T_1 = (N_1, E_1)$. Since every tree is examined separately, this method can not guarantee to find the global optimum.

There do also exist alternative ways of generating R-vines. Kurowicka (2011) proposed a method that tries to generate the R-vine starting with the last tree, building the tree from the other side, trying to construct a R-vine with the lowest dependencies in the top trees.

Kendall's τ is chosen as dependence measures since it measures dependence independently of the assumed distributions. This is especially useful when combining different copula families.

The method proposed in Dißmann (2010) can be summarized as follows:

4. Vine copula

Algorithm 4.3.1 Sequential method to select an R-vine model based on Kendall's τ .

Require: Uniformly distributed margins of n variables.

Ensure: A fully specified R-vine (C-vine or D-vine).

- 1: Calculate the empirical Kendall's $\hat{\tau}$ for all possible variable pairs.
- 2: Select the spanning tree that maximizes the sum of absolute empirical Kendall's $\hat{\tau}$'s

$$\max_{\text{edges } ij \text{ in spanning tree}} \sum |\hat{\tau}_{ij}|.$$

(For this step we can simply use the algorithm of Prim. Generally speaking we are looking for a tree. If we however modify this step to look for a star or a path instead we will obtain a D-vine or C-vine.)

- 3: Select a copula for all pairs found in Step 2 and fit the corresponding parameter.

(This is explained in detail in Section 4.3.2 and Section 4.3.3)

- 4: Transform the observations by using the h-function (as defined in (4.4)) for the in Step 3 chosen bivariate copula and its parameters.
 - 5: Use these transformed observations to calculate empirical Kendall's $\hat{\tau}$'s for all possible pairs (proximity condition (see Definition 25) must be fulfilled).
 - 6: Proceed with Step 2 until receiving a fully specified R-vine (C-vine or D-vine).
-

If we want to obtain a C-vine we have to determine a unique root node with degree $n - i$ (where n is the dimension of the portfolio and i the number of the considered tree) in each tree. This can be done by finding a *spanning star* which maximizes the edge weights in a complete graph. Summing up the columns in the weight matrix (in our case a matrix of absolute empirical Kendall's τ) and choosing the one with the maximum sum gives us the desired result.

In the case that we want to obtain a D-vine we are only interested in finding the first tree, since all other trees are already uniquely determined by the first tree. Here the construction is much more complex than in the case of C-vines or R-vines. The problem of finding the longest sequence of nodes, by means of absolute edge weights, with each variable occurring

4. Vine copula

exactly once, corresponds to the problem of finding a *longest Hamiltonian path*. Or equivalently the shortest Hamiltonian path in terms of weights $1 - |\hat{\tau}_{ij}|$. Unfortunately this problem is NP-hard, meaning that there is no fast efficient solution, especially in higher dimensions. This problem can be transformed into the *Traveling Salesman Problem* (TSP) i.e where the shortest circle that uses each node exactly has to be found. There are a wide range of heuristics available to solve this problem.

4.3.2. Copula family selection

In addition to a vine structure we need to select a copula family for each pair of variable that fits best. In this thesis we take a look at the following six copula families, all of which have different properties and tail dependence:

- Gaussian/Normal copula (tail-symmetric, no tail dependence)
- Student-t copula (tail-symmetric, tail dependence)
- Clayton copula (tail-asymmetric, lower tail dependence)
- Gumbel copula (tail-asymmetric, upper tail dependence)
- Frank copula (tail-symmetric, no tail dependence)
- Joe copula (tail-asymmetric, upper tail dependence)

For determining which copula family gives us the best fit we use the AIC (Akaike Information Criterion) introduced by Akaike (1974). It is a measure of the relative quality of statistical models for a given set of data. Although the AIC will select the best model from a set, it only measures the quality of a model relative to other models. It won't say anything about absolute quality of a model. Hence AIC gives no indication of a poor fit in a case where all candidate models fit poorly.

Definition 30 (Akaike information criterion)

Let M be a statistical model of some data x . AIC is defined as

$$AIC := -2 \ln [\hat{L}] + 2k,$$

where \hat{L} is the maximized value of the likelihood function for the model and k denotes the number of parameters; i.e. $\hat{L} = f(x|\hat{\theta}, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function.

4. Vine copula

In order to select a suitable copula family we first need to estimate parameters for each copula family using bivariate maximum likelihood estimation. Using the estimated parameters, for a specific bivariate copula density c the AIC is hence given by (see Brechmann (2010))

$$AIC := -2 \sum_{i=1}^d \ln c(u_{i,1}, u_{i,2} | \hat{\theta}) + 2k,$$

where $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)'$, $(u_{i,1}, u_{i,2})$ for $i = 1, \dots, d$ are observations and with $k = 1$ for bivariate copulas with one parameter (e.g. Gaussian, Clayton etc.) and $k = 2$ for bivariate copula with two parameters (Student's t). We calculate the AIC's for every possible family for each pair of variables and then choose the copula family with the smallest AIC. Manner (2007) found that this is overall a relatively reliable criterion. The performance of AIC as selection criterion is not entirely satisfactory for weak dependence (by means of Kendall's τ), nevertheless in those cases where it does not find the best fitting model is usually chooses one that is close to it.

This bivariate estimation method fits nicely into Algorithm 4.3.1 which estimates the copula family and its parameters sequentially one tree after the other. It estimates the copula families and its parameters individually for each bivariate copula in one tree and then computes transformed variables for the following tree using the corresponding h-function, and so on. Using this method only bivariate copula estimation is required, which can be calculated rather quickly.

We now outline the copula family estimation and parameter estimation part of Algorithm 4.3.1 by using an example. For this example we use the 5-dimensional R-vine described in Figure 4.1.

Example: First we need to estimate the tree structure of the first tree by using the algorithm of Prim. By using maximum likelihood estimation and AIC, we estimate the copula families and its parameters, namely C_{12}, C_{23}, C_{34} and C_{25} based on the transformed observations $F_j(x_{k,j})$ for $x_{k,j}$, $k = 1, \dots, d$, $j = 1, \dots, 5$. At this point we have already fully specified the first tree of our R-vine. For the next tree we now consider all edges which do not violate the proximity condition and calculate all transformed observations corresponding to these possible edges. We then calculate Kendall's τ and

4. Vine copula

estimate the tree structure for this tree. In the next step we need to estimate the families and the parameters of all conditional copulas for this tree, which we will illustrate for $C_{13|2}$. For this we have to use transformed observations $F_{1|2}(x_{k,1}|x_{k,2}, \hat{\theta}_{12})$ and $F_{3|2}(x_{k,3}|x_{k,2}, \hat{\theta}_{23})$, $k = 1, \dots, d$ via the h-function using the parameters of copula C_{12} and C_{23} . Let us assume C_{12} was estimated to be a Clayton copula with parameters $\hat{\theta}_{12}$. The transformed observations are now being calculated using a Clayton h-function: $F_{1|2}(x_{k,1}|x_{k,2}, \hat{\theta}_{12}) = h(x_{k,1}, x_{k,2}, \hat{\theta}_{12})$, for $k = 1, \dots, d$. Similarly we calculate $F_{3|2}(x_{k,3}|x_{k,2}, \hat{\theta}_{23})$. By using maximum likelihood estimation and AIC, we estimate the copula family and its parameters for copula $C_{13|2}$. This procedure is continued along the sequence of trees until the last tree in the pair-copula-construction is reached.

Comparison of different vine models:

When constructing different models from the same data, like we do in Chapter 5 when comparing C-vines, D-vines and R-vines, the question of which model is "better" arises. Brechmann (2010) describes a popular way which allows for easy model comparison by using AIC as described in Definition 30. Given observations x_i , $i = 1, \dots, d$, we get

$$AIC := -2 \sum_{i=1}^d \ln f(x_i | \hat{\theta}) + 2k,$$

with $\hat{\theta}$ being the maximum likelihood estimate of θ and k denotes the overall number of parameters $\theta = (\theta_1, \dots, \theta_k)'$ which penalize a high count of parameters in order to avoid over fitting. For example in the case of a R-vine copula density as described in Section 4.2 the AIC is given by

$$\begin{aligned} AIC &= -2 \sum_{i=1}^d \left[\ln \prod_{m=1}^{n-1} \prod_{e \in E_m} c_{j(e), k(e) | D(e)}(F(x_{i, j(e)} | x_{i, D(e)}), F(x_{i, k(e)} | x_{i, D(e)})) \right] + 2k \\ &= -2 \left[\sum_{i=1}^d \sum_{m=1}^{n-1} \sum_{e \in E_m} \ln c_{j(e), k(e) | D(e)}(F(x_{i, j(e)} | x_{i, D(e)}), F(x_{i, k(e)} | x_{i, D(e)})) \right] + 2k, \end{aligned}$$

where k denotes the total number of parameters. In a similar way we can evaluate the AICs for C-vines and D-vines. An alternative to AIC would

4. Vine copula

be using the *Bayesian information criterion* (BIC) which involves a stronger penalty term. In both cases the model selection happens by choosing the model which minimizes the criterion.

4.3.3. Parameter estimation

We have already implicitly estimated parameters for the bivariate copulas in Subsection 4.3.2. These parameters however were calculated by using a sequential estimation approach which only leads to a locally optimal selection for each copula.

A better approach for estimating copula parameters would be using maximum likelihood techniques. Given a already determined suitable tree structure and the appropriate bivariate copula families for each pair of variables we are able to estimate the copula parameters θ for these bivariate copulas for observed data $\mathbf{u} \in \mathbb{R}^{d \times n}$ via maximum likelihood estimation. The number of observations is denoted by d whereas n is the dimension of the PCC. The critical part of this task are the conditional distribution functions $F(x_{j(e)}|x_{D(e)})$, which depend on copulas of previous trees.

For PCCs however the number of parameters θ to be estimated, even in rather low dimensions such as five or six, may be too large and hence computationally too heavy for an overall maximum likelihood estimation. A n -dimensional PCC consists of $n(n-1)/2$ bivariate copulas. Hence, for instance, a six-dimensional PCC consisting of $6(6-1)/2 = 15$ bivariate Student's t-copulas, has already 30 parameters that need to be fitted. This leads to a numerically challenging problem in finding the global maximum in such a high-dimensional space, even using more elaborate optimization schemes. Therefore we need to consider faster and computationally easier estimation procedures.

One way to deal with this problem is to consider appropriate starting values for the maximum likelihood estimation. The already estimated parameters using the previously discussed sequential method can be used as starting values for the joint maximum likelihood estimation, if additional accuracy is desired. However, sequential estimates are usually already quite close to those estimates obtained by full maximum likelihood estimation and

4. Vine copula

therefore sufficient for most applications. For a discussion on this topic see Aas et al. (2006) and Hobæk Haff (2013) who investigate different approaches to parameter estimation for PCC.

5. Empirical studies and analysis

This chapter focuses on the data used for the practical implementation of the vine copula models and its analysis. We are using historical data of a six-dimensional portfolio of assets.

5.1. R implementation and packages

The procedure of data analysis, simulation, estimation and optimization was carried out using R 3.4.0 (2017/04/21) by R Core Team (2017). Here is a short overview of the most important packages used to obtain the results. The list of R packages and functions provided here is by no means complete.

- **VineCopula**¹: This package provides tools for the statistical analysis of vine copula models. For our needs we used `RVineStructureSelect()` to fit either an R- or a C-vine copula model to a n -dimensional data set. This function is used to determine pair-copula families, their parameters and the vine structure. To fit a D-vine copula model we additionally need to use the package *TSP* to manually determine the D-vine structure. `D2RVine()` is used to transform the D-vine structure to the corresponding R-vine structure, for which we then use `RVineStructureSelect()` to determine the appropriate pair-copula families and its parameters. Finally, we use `RVineSim()` to simulate samples from a given vine copula model. The package is also used for the visualization of vine copula trees and contour plots.

¹Schepsmeier, Stoeber, and Brechmann, 2017.

5. Empirical studies and analysis

- **rugarch**²: This package aims to provide a flexible and rich univariate GARCH modeling environment. We use the function `ugarchspec()` to specify a GARCH(1,1) model with different error distributions which is then used by `ugarchfit()` to fit the univariate GARCH(1,1) model, to find model parameters and to calculate other characteristic values. For forecasting we use either the function `ugarchforecast()` or `ugarchroll()`. Finally, we use the function `VaRTest()` and `ESTest()` for backtesting VaR and CVaR. These two functions provide an implementation of the “unconditional and conditional coverage Value at Risk Exceedances Test” of Kupiec and Christoffersen and the “Expected Shortfall Test” of McNeil and Frey. The package also provides a set of very useful plots such as QQ-plots, ACF-plots etc.
- **TSP**³ This package provides tools for solving a traveling salesperson problem (TSP). For the construction of a D-vine copula model we need to find a longest(shortest) Hamiltonian path. With the usage of `insert_dummy()` we insert a dummy node into our object and obtain a TSP problem which is then being solved using `solveTSP()`. To obtain the original problem we use `cut_tour()` to remove the previously added node.
- **Copula**⁴ The copula package provides classes of commonly used elliptical and archimedean copula families, including methods for density, distribution and plots. We use this package mainly to fit a multivariate Student’s t-copula using `tCopula()` and `fitcopula()` to multivariate observations. Furthermore, we use the function `rCopula()` to simulate samples from the previously estimated multivariate Student’s t-copula model.
- **ggplot2**⁵: A system for “declaratively” creating graphics and a tool for visualizing data. In addition to the package `ggplot2`, other packages for visualization and the preparation of data frames are used. e.g. `gridExtra` (arrange multiple grid-based plots), `corrplot` (graphical display of a correlation matrix), `GGally` (extension of `ggplot2`, used to create pairwise plot matrix), `reshape` (manipulation of data frames) and `xtable` (export of tables to latex).

²Ghalanos, 2015.

³Hahsler and Hornik, 2017.

⁴Hofert, Kojadinovic, et al., 2017.

⁵Wickham and Chang, 2016.

5.2. Setting up of fixed vs. refitted rolling window models

We are interested in analyzing and comparing the performance of four multivariate VaR and CVaR model, namely the C-vine, D-vine and R-vine copula model and a multivariate Student's t-copula model. Our goal is the estimation of one-day-ahead VaR and CVaR forecasts using the last 943 daily observations up to this date. For this we are using two different approaches with respect to the data used for the parameter estimation and the frequency of its re-estimation.

5.2.1. Fixed model

In this approach we estimate all parameters for the GARCH(1,1) model and the copula model for VaR and CVaR predictions using the first 943 daily log-returns. These parameters will not be updated, but will stay fixed during the entire backtesting procedure. For this model we expect the performance to decrease as time goes by and dependencies change. However, due to the fact that we are keeping the parameters constant and do neither update the GARCH parameters nor the C-vine, D-vine, R-vine and multivariate Student's t-copula parameters, we are computationally more efficient in comparison to a model that needs to be constantly refitted. This gained benefit can be invested in increasing the number of simulations to 100.000 to get a more accurate estimation of VaR and CVaR.

5.2.2. Refitted rolling window models

In this approach we are refitting both the GARCH(1,1) and vine the copula model respectively multivariate Student's t-copula model parameters periodically by using data from the last 943 days. Hence for re-estimation we are using a rolling window of size 943. We have the option to refit the GARCH parameters and the C-vine, D-vine, R-vine and multivariate Student's t-copula models in every single step. A daily refit of parameters

5. Empirical studies and analysis

is however computationally very demanding. Due to very long estimation time and aiming at the improvement of the overall performance it is often not feasible to refit on every single day but e.g. every 50 days. Additionally we need to reduce the sample size of simulations from the copula models. In the case of a daily refit we are using 30.000 simulations, which is already computationally demanding.

Figure 5.1 illustrates 814 one-day-ahead VaR and CVaR forecasts, for a total of 1757 daily observed log-returns. The last forecast (815) can not be backtested and will therefore be disregarded.

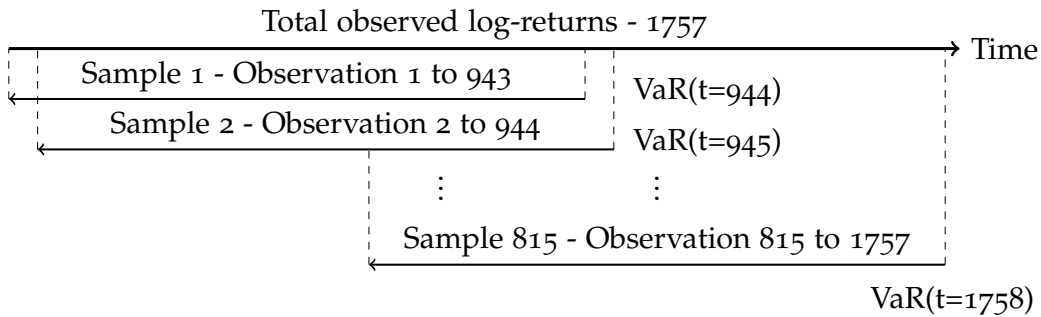


Figure 5.1.: Observations used for one-day-ahead forecasting.

5.3. Data and analysis of daily log-returns

Period	fitting period	testing period	Total
Date	01/01/2004 - 31/12/2007	01/01/2008 - 07/01/2011	
Observations	$N_1 = 943$	$N_2 = 814$	$N = 1757$

Table 5.1.: Initial fitting period and testing period.

For the empirical part of this thesis, we choose a portfolio of six assets with daily observations. We use publicly available data (adjusted closing prices) taken from *Yahoo! Finance* from the periods as described in Table 5.1.

5. Empirical studies and analysis

Allianz SE

Datasource: Yahoo! Finance

Data available since: 12/16/1996

Exchange: XETRA

Symbol: ALV.DE

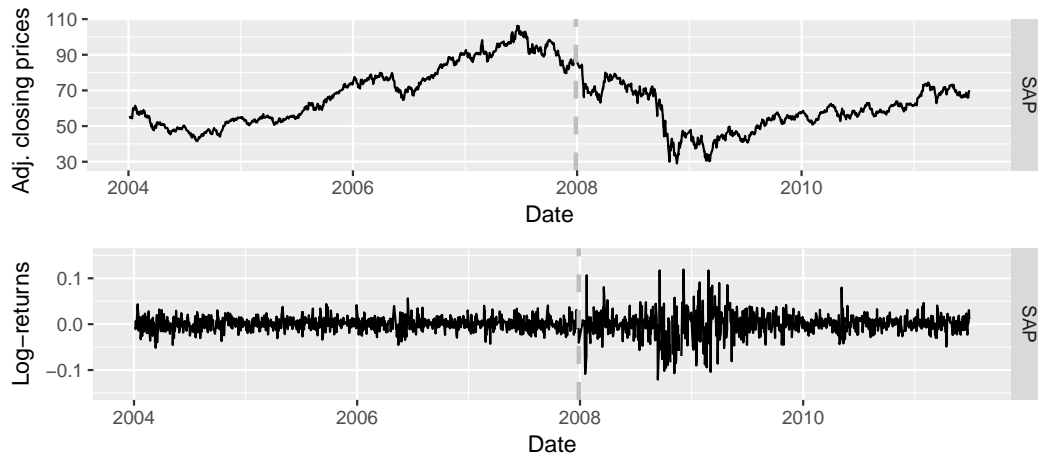


Figure 5.2.: Adjusted closing prices and the corresponding daily log-returns of ALV.DE.

Bayerische Motoren
Werke Aktiengesellschaft

Datasource: Yahoo! Finance

Data available since: 11/08/1996

Exchange: XETRA

Symbol: BMW.DE

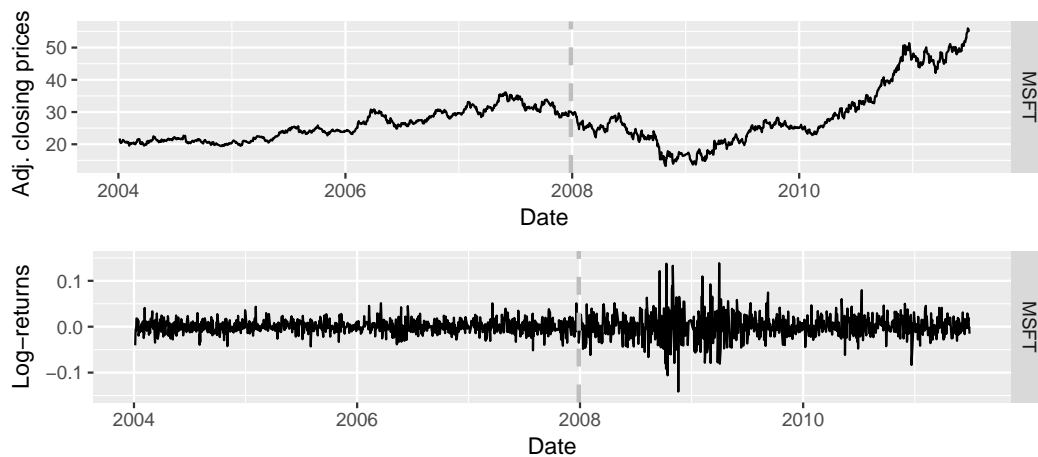


Figure 5.3.: Adjusted closing prices and the corresponding daily log-returns of BMW.DE.

5. Empirical studies and analysis

DAX

Datasource: Yahoo! Finance
Data available since: 12/30/1987
Symbol: ^GDAXI

Exchange: XETRA

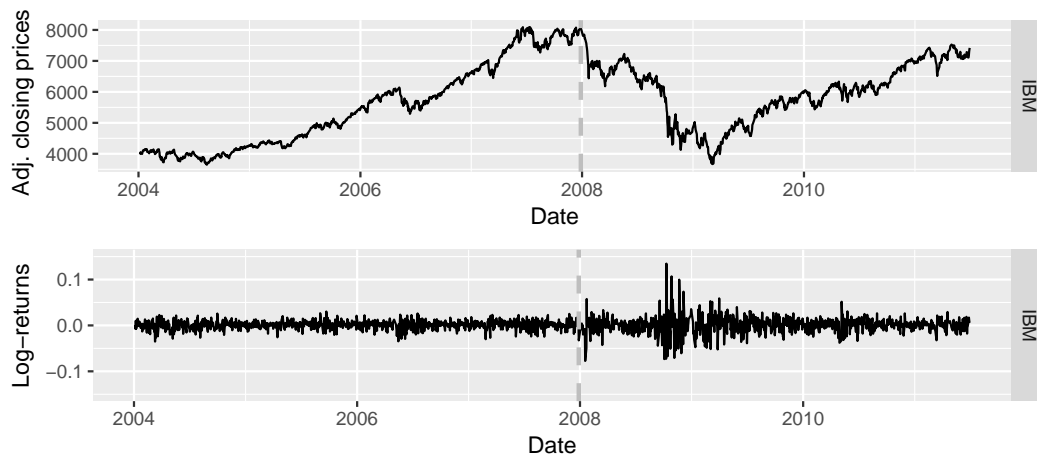


Figure 5.4.: Adjusted closing prices and the corresponding daily log-returns of ^GDAXI.

Treasury Yield Option Index 10-Year Note

Exchange: CBOE

Datasource: Yahoo! Finance
Data available since: 1/2/1962
Symbol: ^TNX

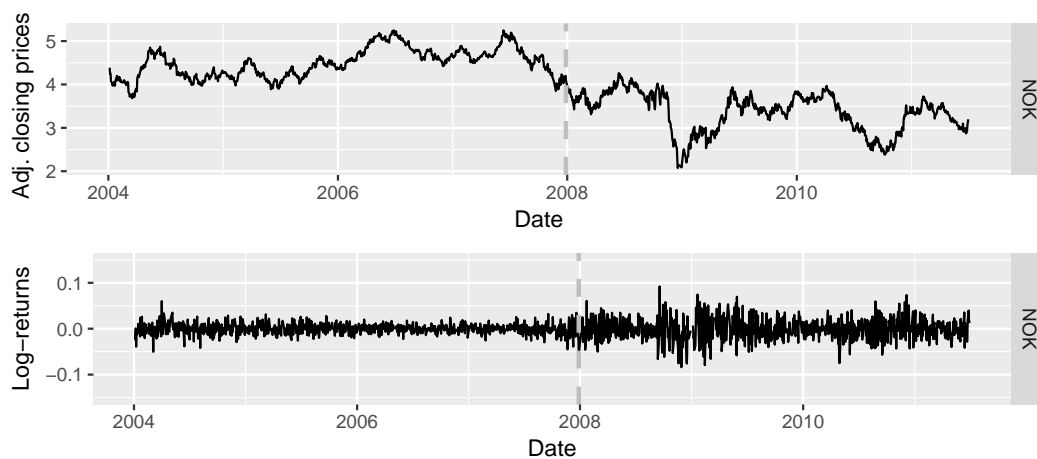


Figure 5.5.: Adjusted closing prices and the corresponding daily log-returns of ^TNX.

5. Empirical studies and analysis

Toyota Motor Corporation
Exchange: NYSE

Datasource: Yahoo! Finance
Data available since: 8/18/1976
Symbol: TM

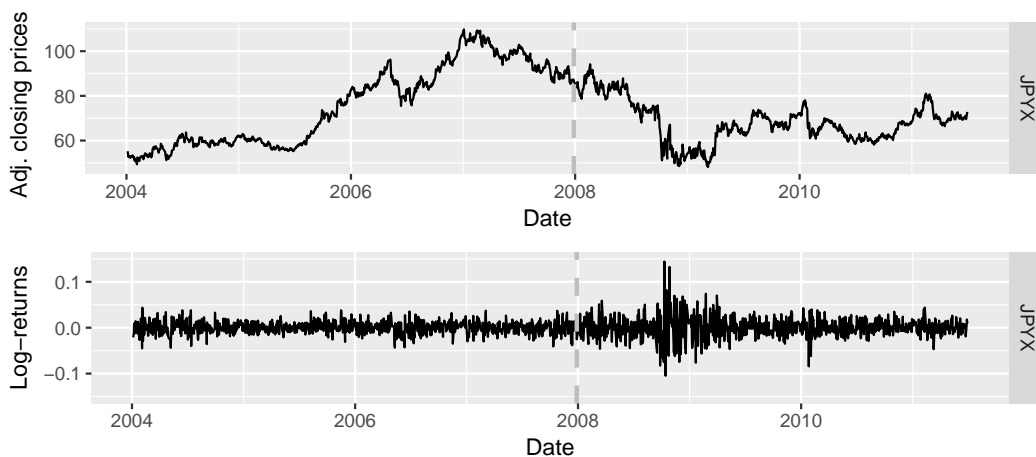


Figure 5.6.: Adjusted closing prices and the corresponding daily log-returns of TM.

Nikkei 225
Exchange: Osaka

Datasource: Yahoo! Finance
Data available since: 1/5/1965
Symbol: ^N255

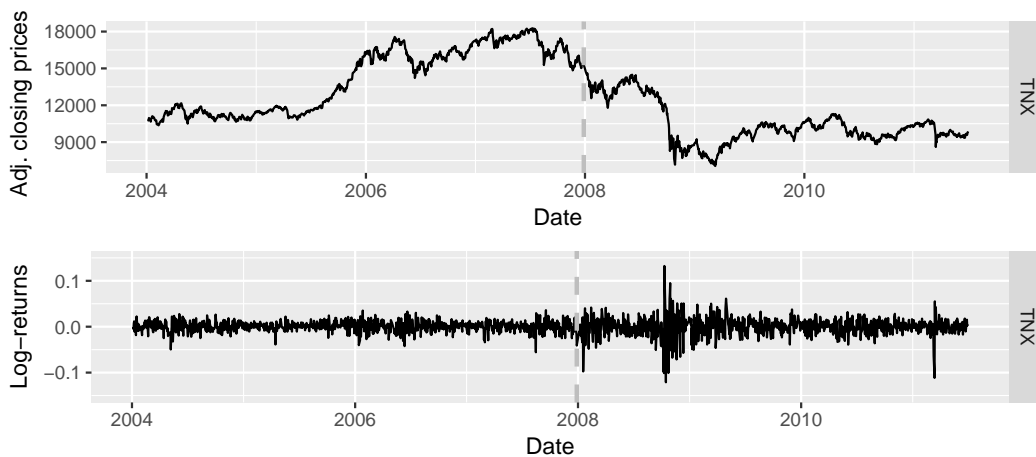


Figure 5.7.: Adjusted closing prices and the corresponding daily log-returns of ^N255.

5. Empirical studies and analysis

Before we go deeper into the construction of suitable vine copula models, we first perform some preliminary analysis of all six log-return series.

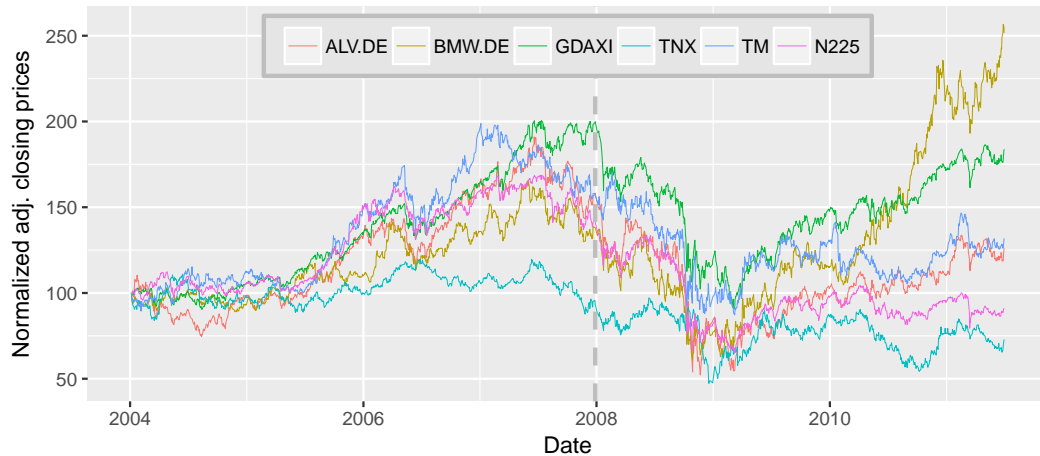


Figure 5.8.: Normalized adjusted closing prizes using 100 as the base value at 1/1/2004.

The vertical dotted line in Figure 5.8 and Figure 5.9 is indicating the separation between days of the fitting period, that are solely used for fitting the vine copula models and the GARCH(1,1) model and days of the testing period.

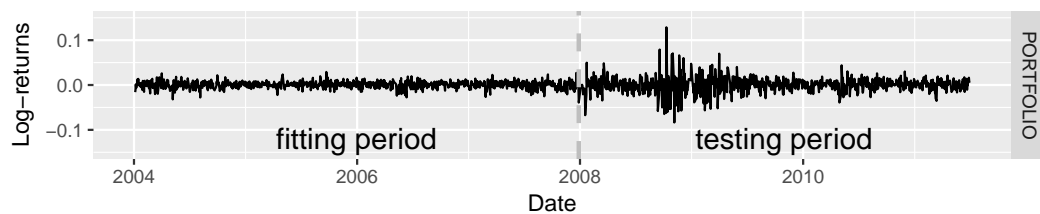


Figure 5.9.: Log-returns of equally weighted portfolio.

Figure 5.9 shows that with the beginning of the global financial crisis from 2008 until 2010 we see a significant increase in volatility of the log-returns. Starting around 2010 the volatility reverts back to almost pre-crisis levels.

5. Empirical studies and analysis

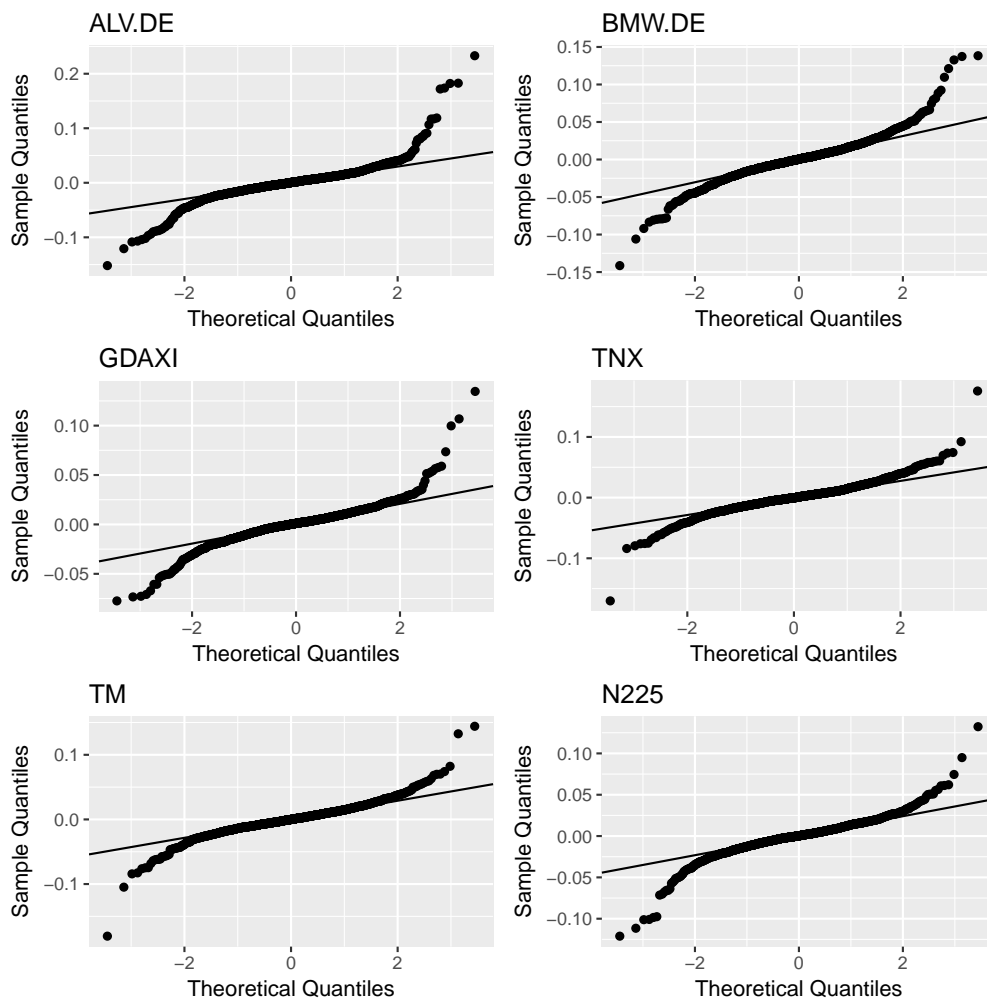


Figure 5.10.: Normal QQ-plots of all daily log-returns.

We carried out a test for normality by examining normal QQ-plots for each of the six log-return series. If the data is from the assumed normal distribution, then the QQ-plot will approximately represent a straight line. As seen in Figure 5.10 our observed data follows by no means a normal distribution. We see a significant deviation from the straight line for all six log-return series, especially in the tails. This is a strong indication that the assumption of a normally distributed dataset has to be discarded for a more heavy-tailed distribution.

5. Empirical studies and analysis

	minimum	maximum	mean	std.dev	skewness	kurtosis
ALV.DE	-0.15187	0.23305	0.00013	0.02368	1.07173	16.32227
BMW.DE	-0.14139	0.13836	0.00053	0.02142	0.28744	6.10397
GDAXI	-0.07739	0.13463	0.00035	0.01445	0.31923	10.19878
TNX	-0.17021	0.17575	-0.00018	0.01901	-0.00446	9.34664
TM	-0.18061	0.14419	0.00016	0.01874	-0.24291	10.21087
N225	-0.12111	0.13235	-0.00005	0.01669	-0.62299	9.25592

Table 5.2.: Results of descriptive statistical analysis of all daily log-returns.

The descriptive statistic of daily log-returns of all assets is given in Table 5.2. As expected we see that the mean of all six time-series is close to zero. Both the kurtosis and skewness of each series deviates from zero, which indicates that all six return series are leptokurtic. This is another indication that the return series distributions are heavy-tailed and do not follow a normal distribution. The *Shapiro-Wilk normality test* also shows that all return series reject normal distribution at the 1% level of significance.

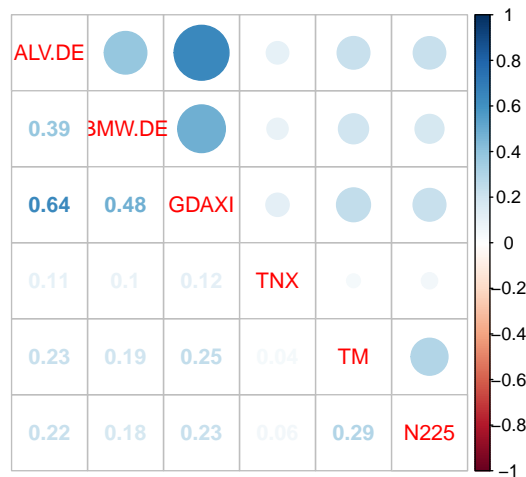


Figure 5.11.: Empirical Kendall's τ of log-returns computed pairwise for the fitting period.

5. Empirical studies and analysis

Figure 5.11 displays a matrix of empirical Kendall's τ of all possible pairs of the log-return series. Negative correlations are displayed in red and positive correlations are displayed in blue. The colors and the size of the circles are proportional to the Kendall's τ correlation coefficients. The legend, on the right side of the correlation matrix, shows the corresponding colors of the correlation coefficients. Our data suggests that the first three assets (*ALV.DE*, *BMW.DE* and *GDAXI*) have a moderate to strong positive correlation to each other. These dependencies can also be seen in the first level of all three vine structures as depicted in Figure 5.31. This also applies similarly, with slightly weaker correlation, to the assets *TM - N255*. All other correlations are rather weak. The correlation of *TNX* to all other assets is very close to non existent.

5.4. Marginal distribution estimates

After the discussion and analysis of the daily log-returns, we will now concentrate on the modeling of the margins. We are going to use methods from time-series analysis to remove serial dependence among observations in order to obtain i.i.d. data, which can then be used as input for copula models. In particular we are going to use a GARCH(1,1) model with appropriate error distributions chosen by using AIC. See Schafzahl (2018) for further details on the choice of error distribution.

Our datasets consist of financial time series, hence the series of log-returns are expected to exhibit some degree of volatility clustering, which is the tendency of large changes in price of financial assets to cluster together. Benoit Mandelbrot described it as *"large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes"*⁶. In Figure 5.12 we see the autocorrelation function of the squared residuals of all six assets. Autocorrelation describes the correlation observations as a function of the time lag between them. If our given series of log-returns are independent over time, the squared value of the log-returns should not be autocorrelated. This gives us a way to detect the presence of volatility clustering. Using the *Ljung-Box test* we have to reject the hypothesis of

⁶Mandelbrot, 1963.

5. Empirical studies and analysis

independence of the squared log-returns for a level of 3%. The same results can be obtained using ACF plots.

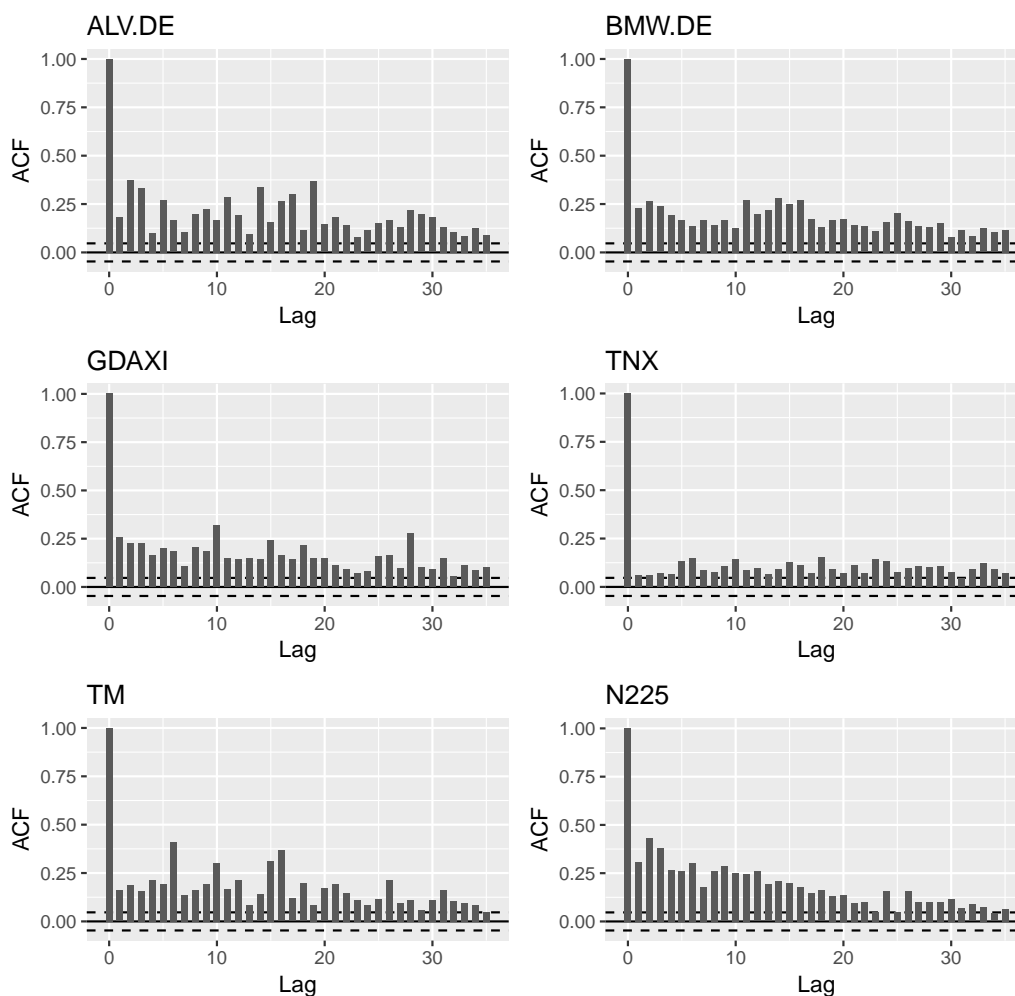


Figure 5.12.: ACF of the squared values of the daily log-returns.

The dotted lines represent the 97% confidence interval and given that there are 30 lags we would expect no more than $(3\% \times 30 = 0.9)$, which can be rounded up to 1 lags, to exceed the boundary. For our six stock the autocorrelation functions of the squared log-returns have a positive value above the dotted line for a relatively large number of lags. Observations

5. Empirical studies and analysis

of this type often indicate volatility clustering and suggests the usage of GARCH models in financial forecasting. In summary we can say that the log-return series do not follow a normal distributions and that the hypothesis of independence of the log-returns over time, has to be rejected for a level of 3%.

We fitted the GARCH(1,1) model to the six return series in order to obtain i.i.d. time series. Table 5.3 presents results from the GARCH(1,1) parameter estimation using data from the first 943 daily log-returns. The sum of the parameters $\alpha_1 + \beta_1$ is approaching 1 from below, which indicates that the GARCH(1,1) models are stationary.

	α_0	ω	α_1	β_1	$\alpha_1 + \beta_1$	shape	Log-Lik
ALV.DE	0.00087	0.00001	0.07516	0.86435	0.93951	10.40133	-2695.69
BMW.DE	0.00038	0.00000	0.01820	0.97545	0.99365	6.86357	-2726.48
GDAXI	0.00088	0.00000	0.07581	0.88201	0.95782	1.48630	-3075.79
TNX	0.00003	0.00000	0.02898	0.96927	0.99825	8.90957	-2939.90
TM	0.00052	0.00000	0.04659	0.93576	0.98235	13.50557	-2776.05
N225	0.00063	0.00000	0.07477	0.91166	0.98643	12.38877	-2940.56

Table 5.3.: Summary of estimated parameters of the GARCH(1,1) model and the log-likelihood value for all six log-return series.

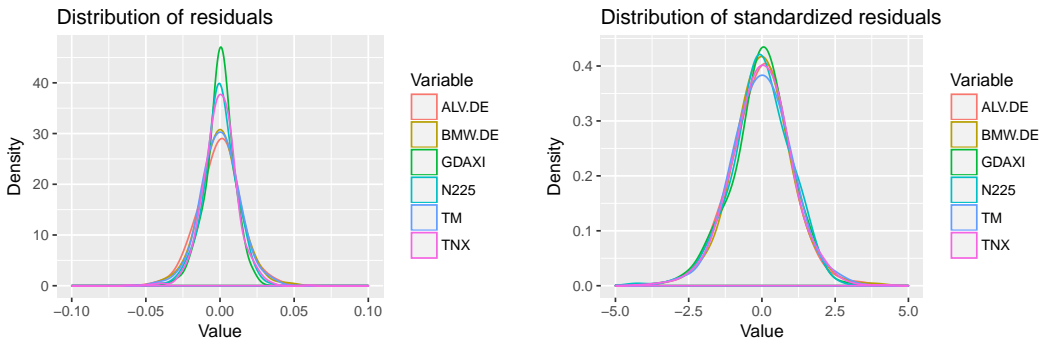


Figure 5.13.: Comparison of GARCH(1,1) residuals vs. GARCH(1,1) standardized residual distribution for each time series for the first 943 daily observations.

Moreover, we have examined the standardized residuals for serial correlation using a series of *Ljung-Box tests*. For a lag of 1 the Ljung-Box test gives us

5. Empirical studies and analysis

p-values all greater than 3%, therefore we cannot reject the null hypothesis that the data is independently distributed and hence any correlation in the population results from randomness. When increasing the number of different time lags, the test shows similar results.

	ALV.DE	BMW.DE	GDAXI	TNX	TM	N225
lag = 1	0.329	0.206	0.563	0.278	0.411	0.547
lag = 15	0.779	0.367	0.851	0.038	0.532	0.926

Table 5.4.: Ljung-Box test of standardized residuals for each time series for first 943 days with lag 1 and lag 15.

The standardized residuals for the first 943 days seen in Figure 5.14 also indicate that there seems to be no visible volatility clustering in our data.

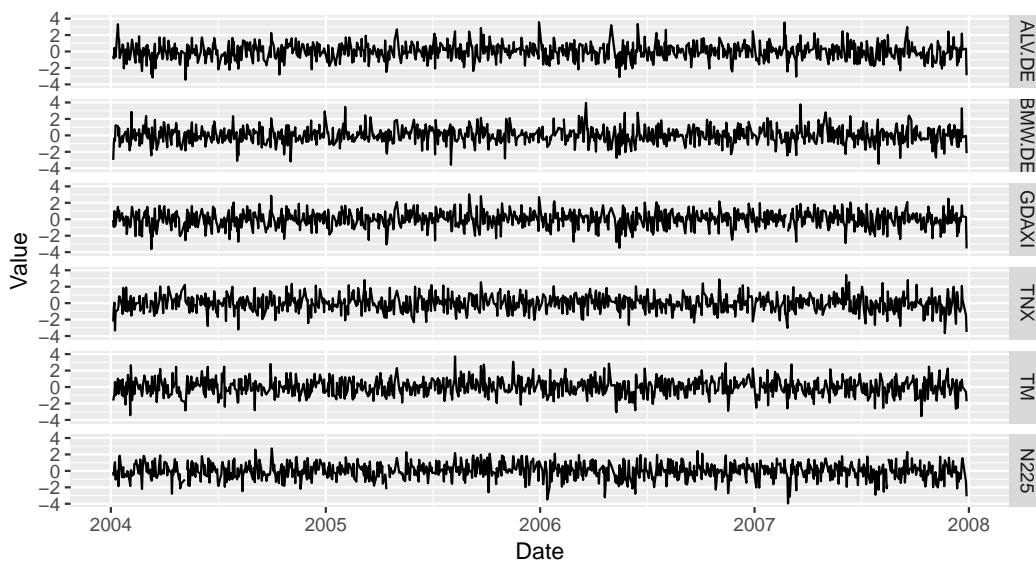


Figure 5.14.: GARCH(1,1) standardized residuals for the first 943 days.

For the same estimation period we also present a more detailed analysis, using a set of four diagnostic plots for each time-series to inspect the goodness-of-fit for the GARCH(1,1) model. The results of this analysis are presented in Figures 5.15 - 5.20.

5. Empirical studies and analysis

ALV.DE

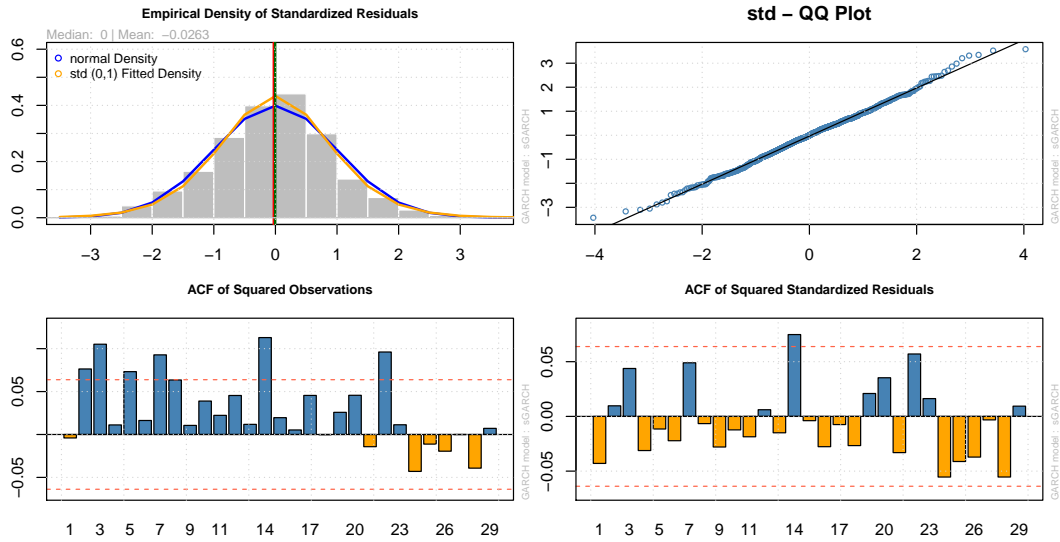


Figure 5.15.: Selected diagnostic plots of a GARCH(1,1) model with Student's t error distribution fitted to ALV.DE data from the first 943 days.

BMW.DE

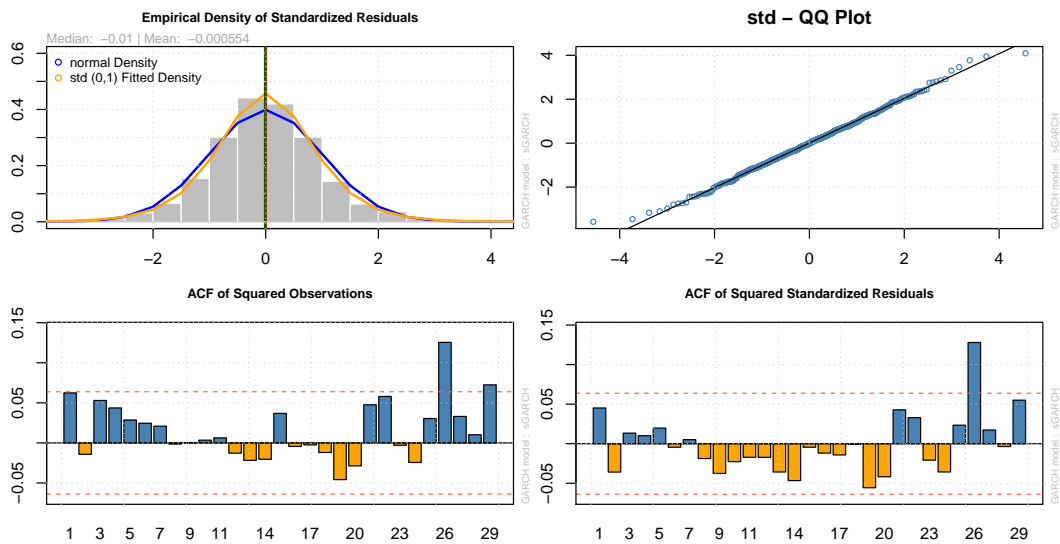


Figure 5.16.: Selected diagnostic plots of a GARCH(1,1) model with Student's t error distribution fitted to BMW.DE data from the first 943 days.

5. Empirical studies and analysis

GDAXI

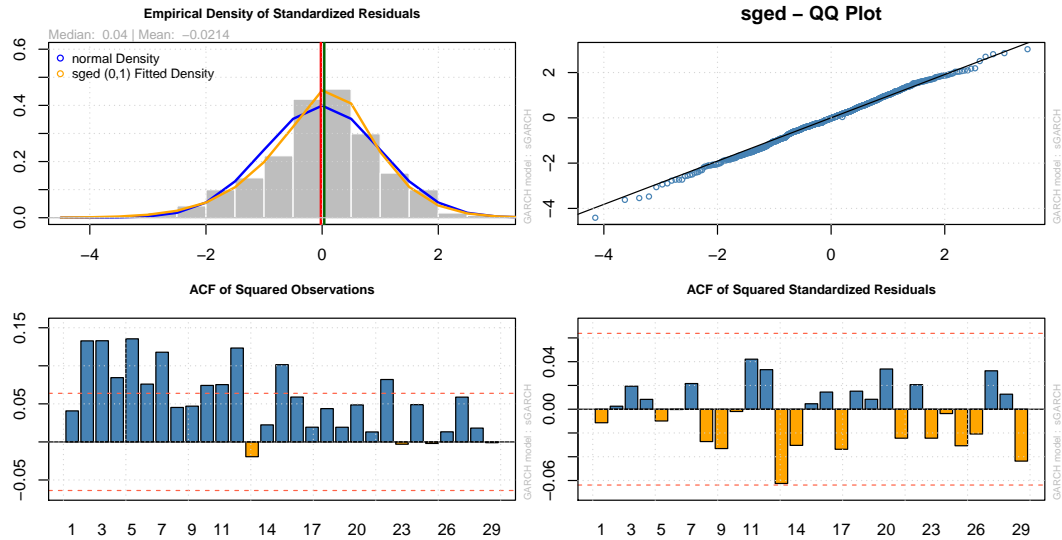


Figure 5.17.: Selected diagnostic plots of a GARCH(1,1) model with skew-generalized error distribution fitted to \hat{GDAXI} data from the first 943 days.

TNX

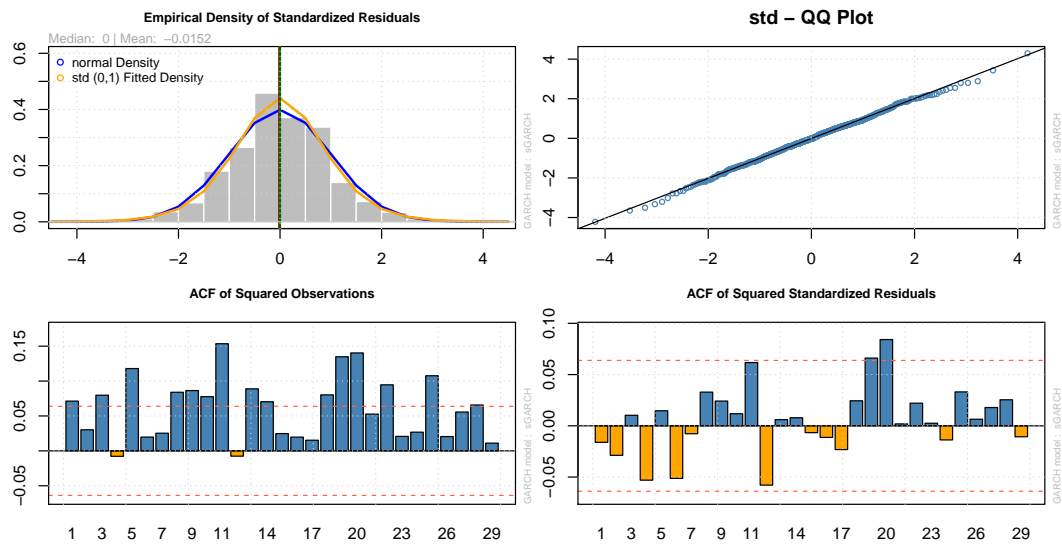


Figure 5.18.: Selected diagnostic plots of a GARCH(1,1) model with Student's t error distribution fitted to \hat{TNX} data from the first 943 days.

5. Empirical studies and analysis

TM

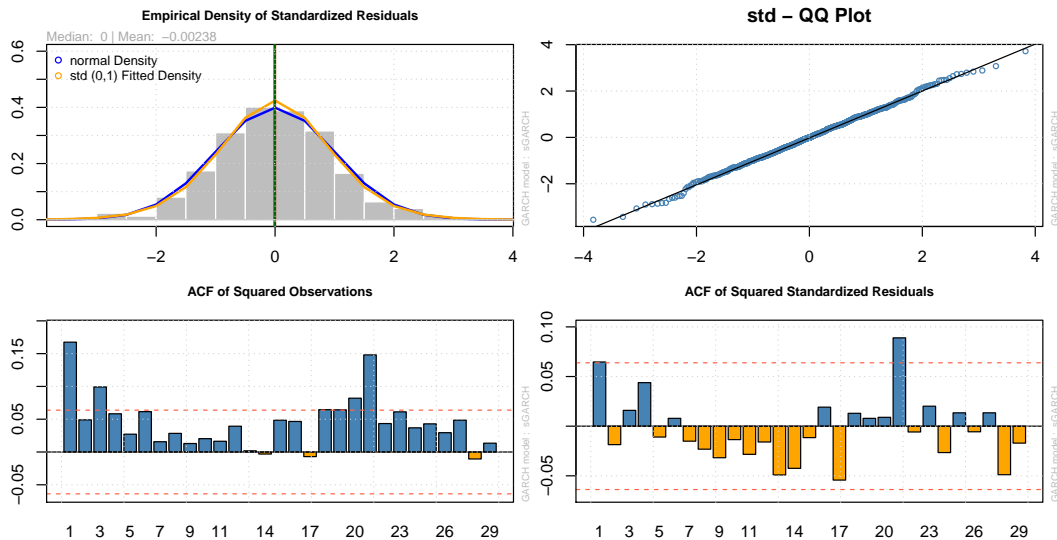


Figure 5.19.: Selected diagnostic plots of a GARCH(1,1) model with Student's t error distribution fitted to TM data from the first 943 days.

N225

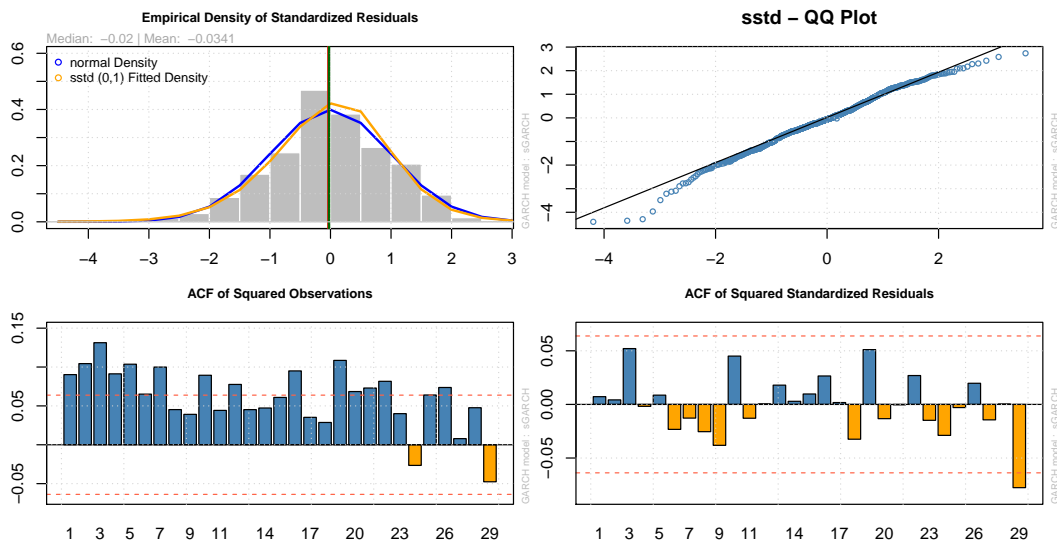


Figure 5.20.: Selected diagnostic plots of a GARCH(1,1) model with skewed Student's t error distribution fitted to $\hat{N}225$ data from the first 943 days.

5. Empirical studies and analysis

Looking at the ACF of the squared standardized residuals

$$\hat{z}_{t,i}^2 = \frac{\hat{\epsilon}_{t,i}^2}{\hat{\sigma}_{t,i}^2}$$

for all six assets $i = 1, \dots, 6$ in Figures 5.15 - 5.20 we see that the squared standardized residuals display no autocorrelation. We see no discernible pattern at any order of lags. This finding has already been confirmed using a series of *Ljung-Box tests*.

In the following, before being able to estimate the parameters of the vine copula models or multivariate Student's t-copula models we will need to transform the margins to uniformly distributed margins, i.e. $u_{t,i} = F_i(x_{t,i})$ for $i = 1, \dots, 6$ and $t = 1, \dots, 943$.

In most cases however, the true distribution functions of the margins F_i 's are unknown. Therefore, according to Brechmann (2010), for the transformation of the standardized residuals we will need to replace the unknown distribution functions F_i 's by their empirical versions

$$\hat{F}_i(\hat{z}) = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{[\hat{z}_{t,i}, \infty)}(\hat{z}). \quad (5.1)$$

where $\mathbb{1}_{[\hat{z}_{t,i}, \infty)}$ is the indicator function. The uniformly distributed margins $u_{t,i}$ for assets $i = 1, \dots, d$ and $t = 1, \dots, n$ can now be calculated by using the so-called *pseudo-observations*

$$u_{t,i} = \frac{\text{rank}(\hat{z}_{t,i})}{n+1} = \frac{n}{n+1} \hat{F}_i(\hat{z}_{t,i}). \quad (5.2)$$

The factor $\frac{n}{n+1}$ is introduced to prevent numerical problems in the boundaries of $[0, 1]^d$.

We now transform the previously calculated standardized residuals $\hat{z}_{t,i}$ for all six return series to uniform variables using the empirical cumulative distribution function according to (5.1) and (5.2). The *Kolmogorov-Smirnov test* fails to reject the null hypothesis H_0 that the distribution of transformed standardized residuals and the uniform distribution are from the same uniform $U(0, 1)$ distribution at a 5% significance level.

5. Empirical studies and analysis

On this basis, copula models such as vine copula or multivariate copula models can now be used to analyze the dependence structure between these six series.

5.5. Finding the vine copula and Student's t-copula models

To model the dependence structure we investigate four models. In particular, a R-vine copula model, selected using the maximum spanning tree algorithm proposed in Dißmann (2010), a C-vine copula model, selected by finding the maximum spanning-star which maximizes the edge weights, a D-vine copula model, selected using a traveling salesman algorithm⁷, and a multivariate Student's t-copula. The procedure of finding a vine copula specification has already been discussed in Section 4.3.

In the vine copula structures shown below the numbers in the boxes represent the stocks (1 = *ALV.DE*, 2 = *BMW.DE*, 3 = *GDAXI*, 4 = *TNX*, 5 = *TM*, 6 = *N255*) whereas the letters on the connecting lines between two boxes represent the corresponding pair-copula chosen by the AIC criteria, which characterizes the relationship between two stocks.

⁷For this we are using the repetitive nearest neighbor algorithm (see Rosenkrantz, Stearns, and Philip M. Lewis (1977)).

5. Empirical studies and analysis

5.5.1. C-vine

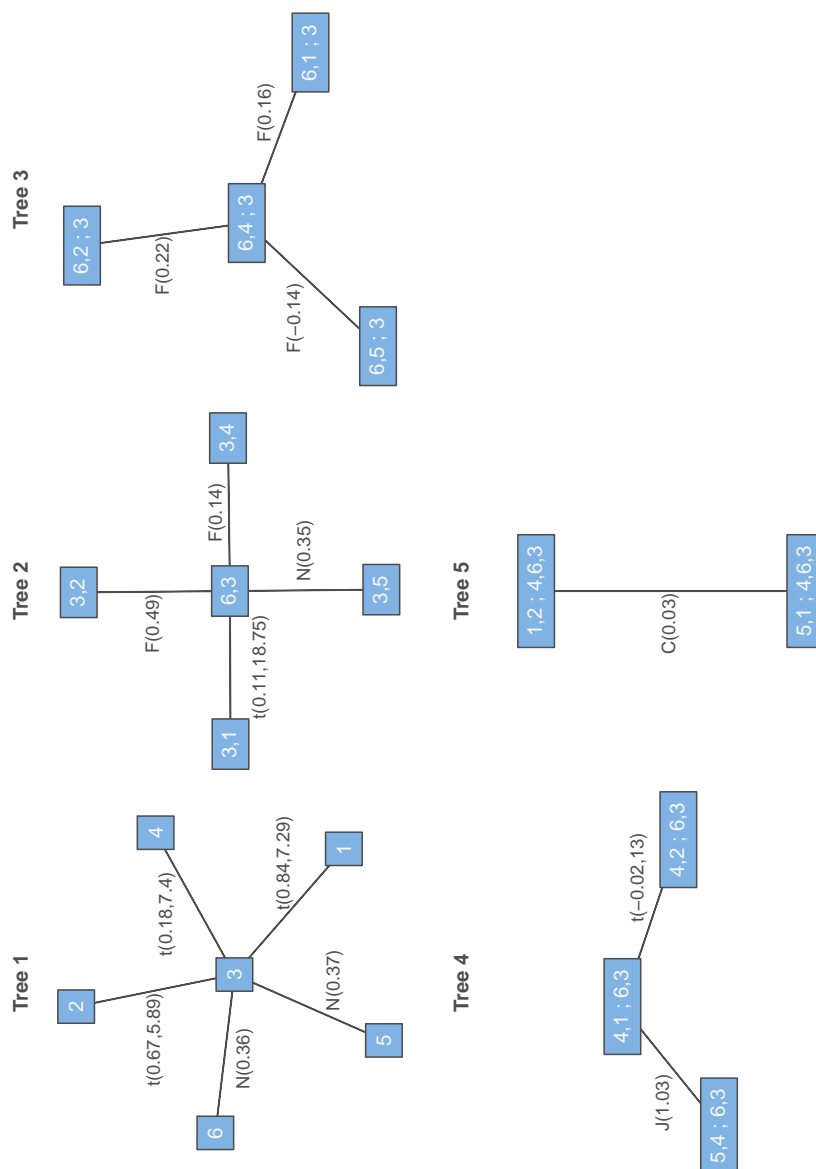


Figure 5.21.: C-vine: Tree structure estimated with the first 943 observations. The edge labels denote the corresponding pair-copula families (N = Gaussian copula, t = Student's t-copula, C = Clayton copula, G = Gumbel copula F = Frank copula, J = Joe copula) and its parameters.

5. Empirical studies and analysis

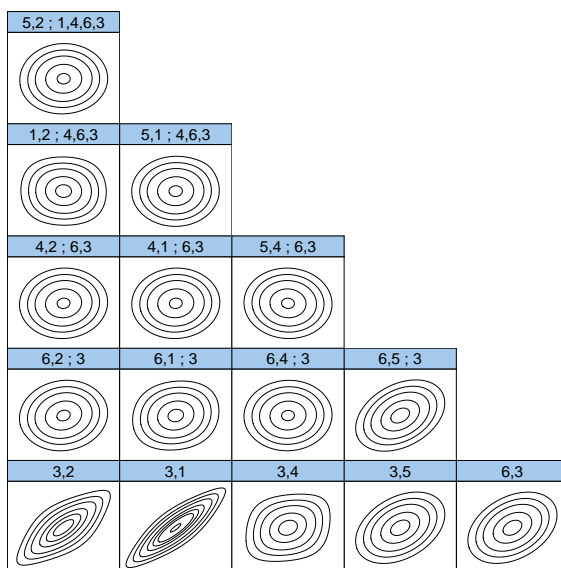


Figure 5.22.: C-vine: Contour plots of all pair-copulas.

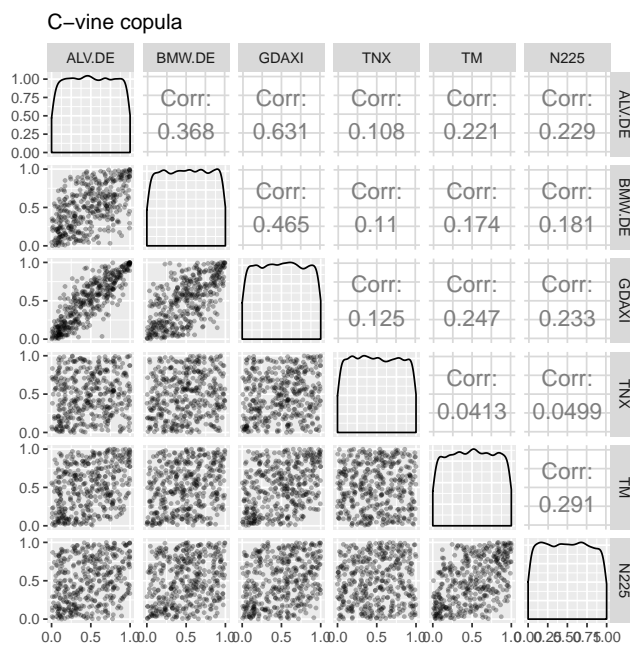


Figure 5.23.: Lower left: Pair plots of C-vine copula simulations.; upper right: Kendall's τ of the transformed data set.

5. Empirical studies and analysis

5.5.2. D-vine

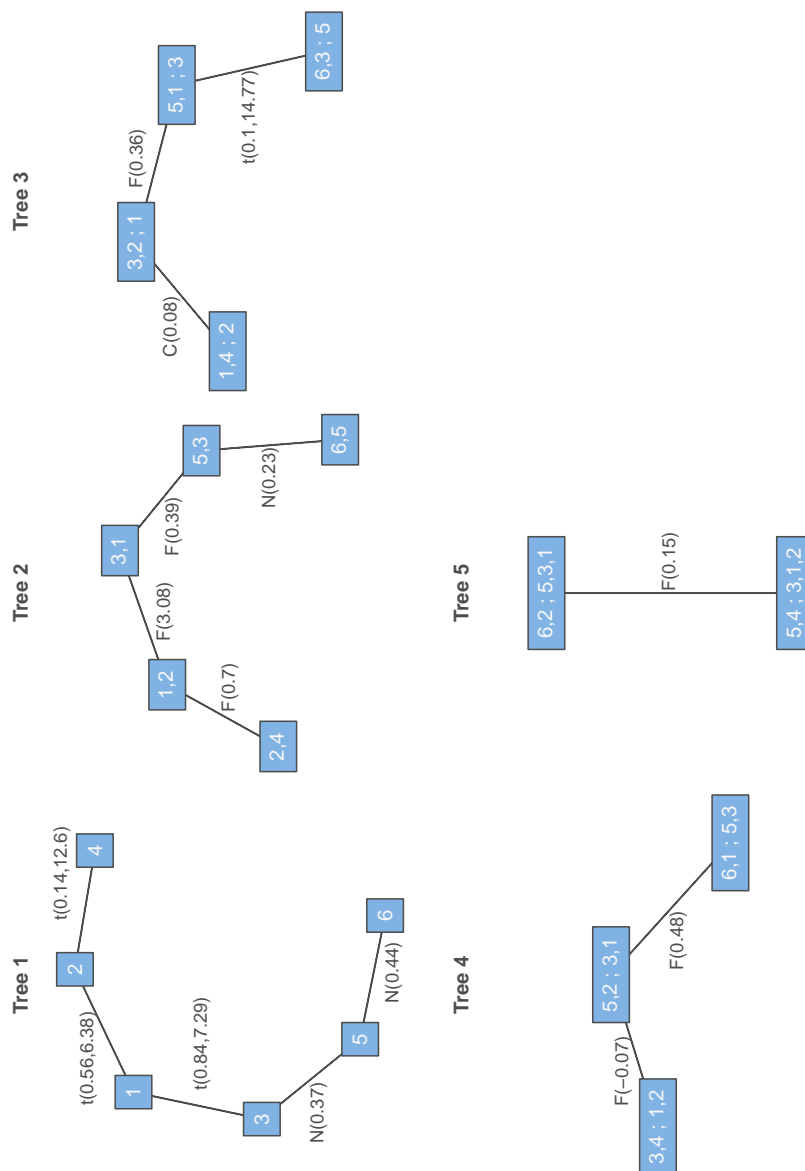


Figure 5.24.: D-vine: Tree structure estimated with the first 943 observations. The edge labels denote the corresponding pair-copula families (N = Gaussian copula, t = Student's t-copula, C = Clayton copula, G = Gumbel copula F = Frank copula, J = Joe copula) and its parameters.

5. Empirical studies and analysis

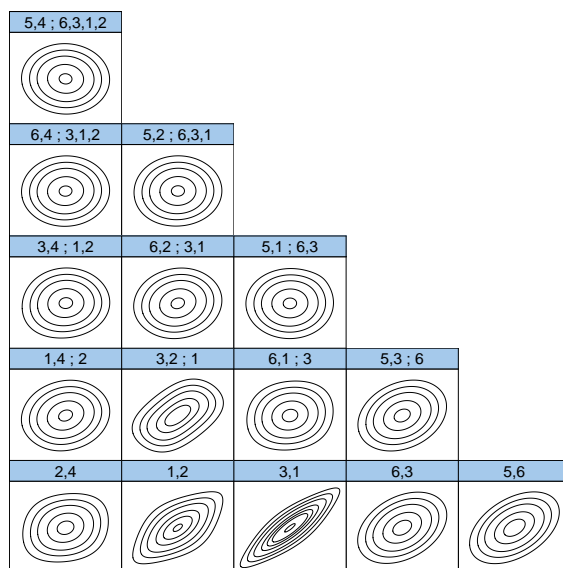


Figure 5.25.: D-vine: Contour plots of all pair-copulas.

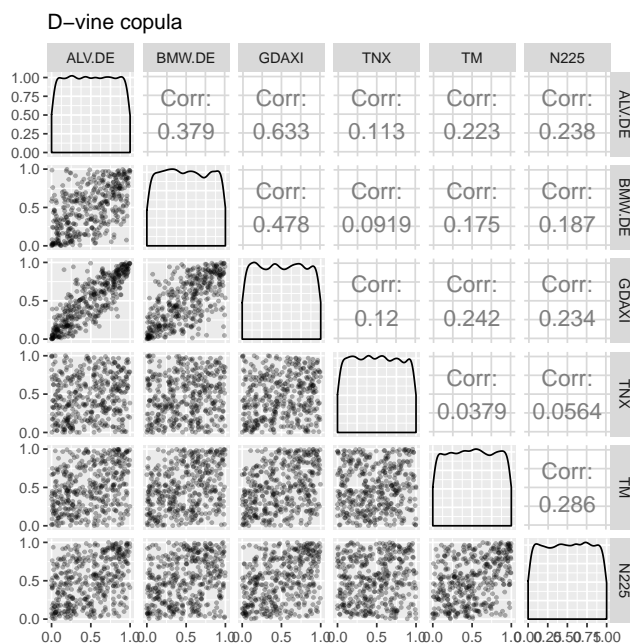


Figure 5.26.: Lower left: Pair plots of D-vine copula simulations.; upper right: Kendall's τ of the transformed data set.

5. Empirical studies and analysis

5.5.3. R-vine

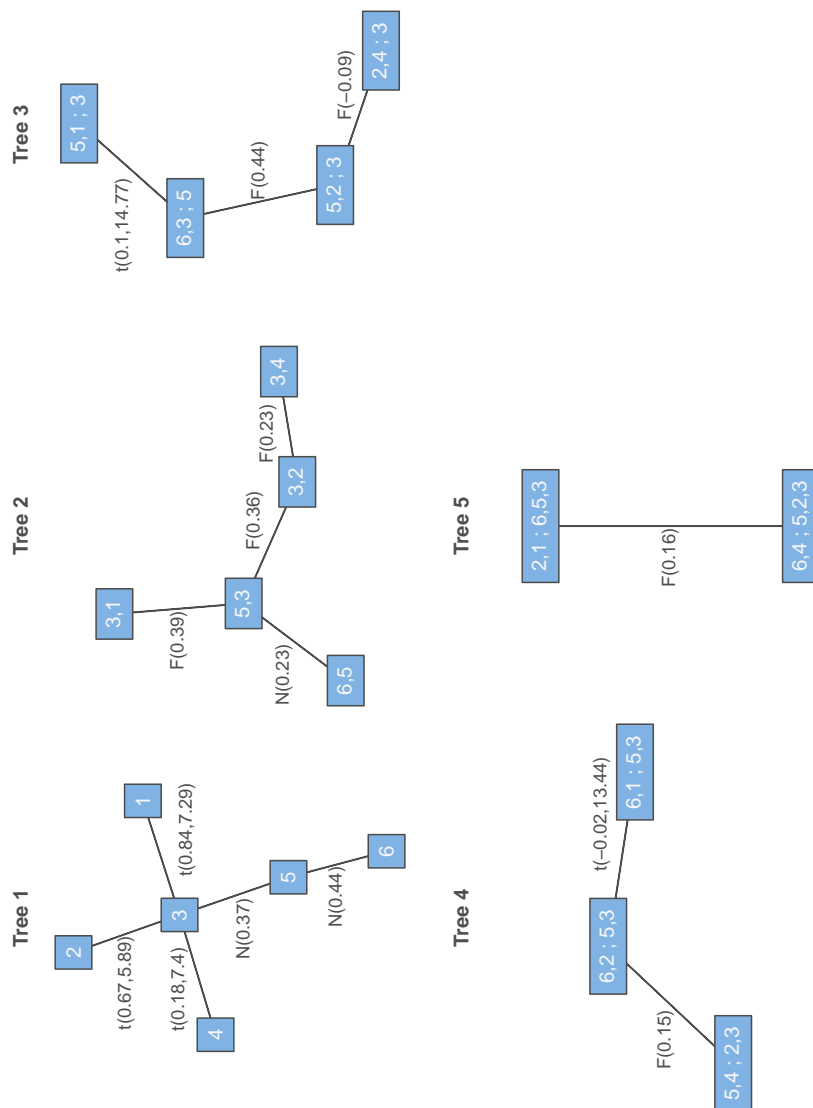


Figure 5.27.: R-vine: Tree structure estimated with the first 943 observations. The edge labels denote the corresponding pair-copula families (N = Gaussian copula, t = Student's t-copula, C = Clayton copula, G = Gumbel copula, F = Frank copula, J = Joe copula) and its parameters.

5. Empirical studies and analysis

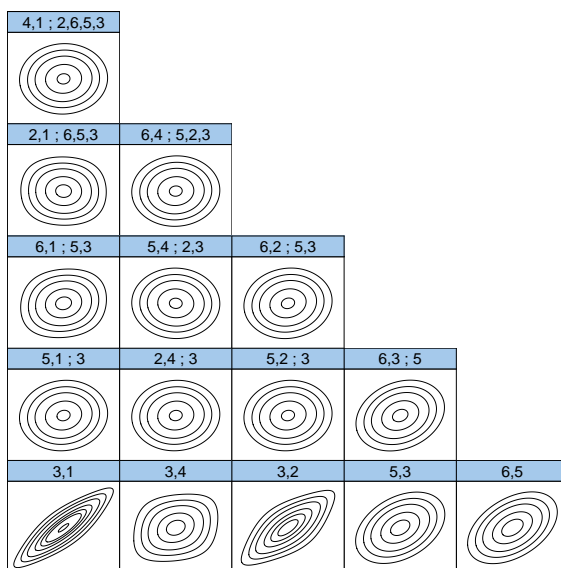


Figure 5.28.: R-vine: Contour plots of all pair-copulas.

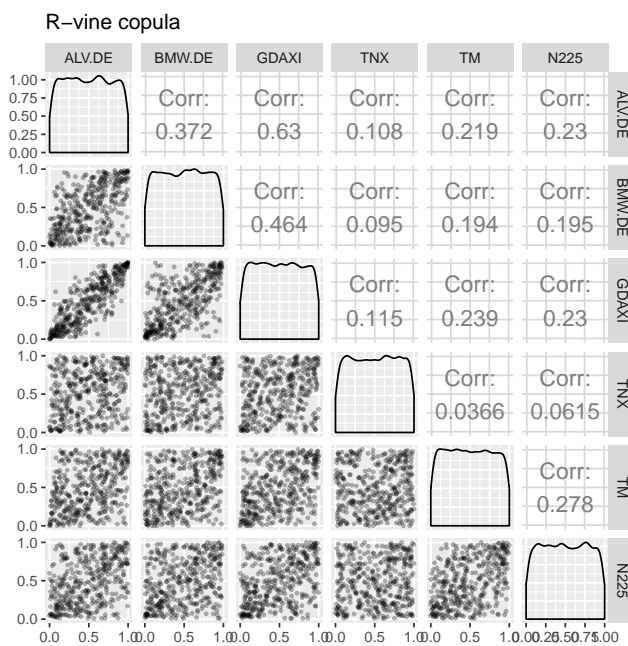


Figure 5.29.: Lower left: Pair plots of R-vine copula simulations.; upper right: Kendall's τ of the transformed data set.

5. Empirical studies and analysis

5.5.4. Student's t-copula

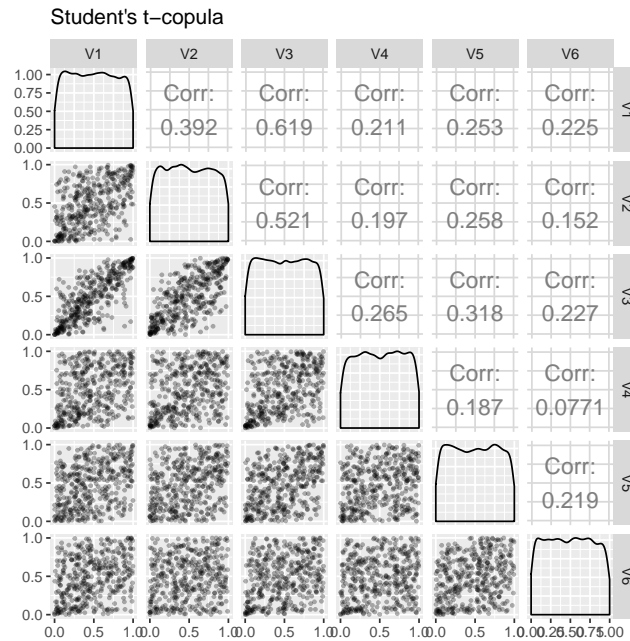


Figure 5.30.: Lower left: Pair plots of Student's t-copula simulations.; upper right: Kendall's τ of the transformed data set.

Moreover, we fitted a multivariate Student's t-copula using the first 943 observations by using the R routine `fitcopula(tCopula(...),...,method="itau.mpl")`⁸. The estimated degree of freedom is $\nu = 16.34$ and the correlation matrix can be seen below:

$$\rho = \begin{bmatrix} 0.56295 & & & & & \\ 0.84533 & 0.16541 & & & & \\ 0.34388 & 0.34473 & 0.68384 & & & \\ 0.14592 & 0.29424 & 0.27350 & 0.18020 & & \\ 0.36866 & 0.35186 & 0.05638 & 0.08615 & 0.43978 & \end{bmatrix}$$

⁸This method for estimating parameters of a Student's t-copula was suggested by Mashal and Zeevi (2002) (see also Demarta and A. McNeil (2005))

5. Empirical studies and analysis

5.5.5. Comparison of models

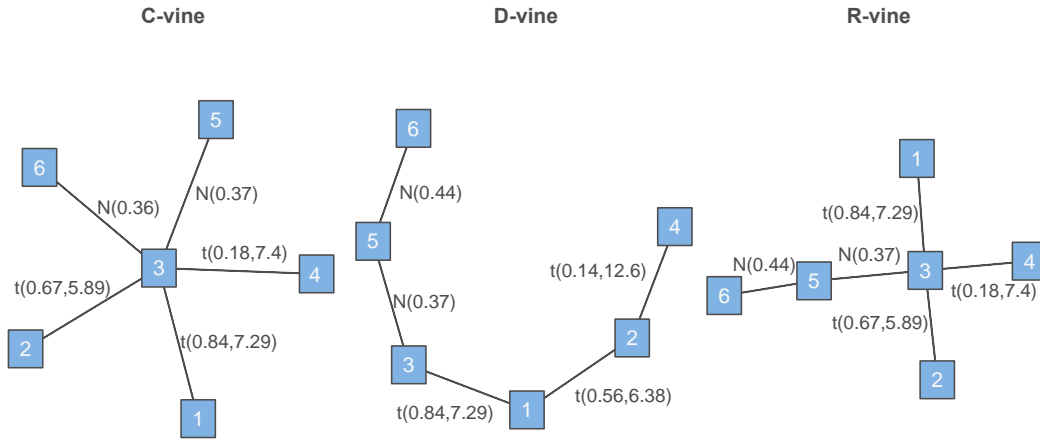


Figure 5.31.: The first level of the C-vine, D-vine and R-vine copula tree structure estimated for the first 943 observations.

The three vine copula structures, which model the relationship of the stocks in our chosen six-dimensional portfolio, all look totally different, but have certainly some bivariate building blocs in common. We will now further analyze the risk measures VaR and $CVaR$ for all three models.

Class	Strategy	No. parameters	Log-likelihood	AIC	BIC
t-copula	ITAU MPL	16	1051.861	-2071.721	-1994.136
C-vine	AIC / MLE	20	1078.935	-2117.870	-2020.888
D-vine	AIC / MLE	19	1073.089	-2108.179	-2016.047
R-vine	AIC / MLE	20	1080.159	-2120.318	-2023.337

Table 5.5.: Comparison of log-likelihood, number of parameters, AIC and BIC of all three vine copula models and the Student's t-copula.

Table 5.5 represents a comparison of the model fit of the three previously described vine copula models and the multivariate Student's t-copula mentioned in Section 5.5.4. In all cases, using either log-likelihood, AIC or BIC,

5. Empirical studies and analysis

we see that the multivariate Student's t-copula performed slightly worse than all three vine copula models.

5.6. Procedure: Forecasting and backtesting

We are using historical data to estimate the parameters of the C-vine, D-vine and R-vine copula models. These models try to capture the dependence structure of our chosen equally weighted six-dimensional portfolio as closely as possible. In this section we explore the use of vine copula models in forecasting one-day-ahead VaR_α and $CVaR_\alpha$ for different levels of α and use statistical tests to determine the accuracy of our predictions. In the first scenario of this backtesting procedure, we determine the dependence structure of our portfolio using C-vine, D-vine and R-vine models and assume that the dependence structure doesn't change for the entire backtesting period. We call this the fixed model approach. This strategy might however be problematic in scenarios where the dependence structure changes, therefore we will also use a 943 days rolling window approach, where we refit the GARCH(1,1) model and the copula models every 50 days and compare the results to the fixed model approach, where the model is estimated on day one and stays constant for all following days.

The main steps of the backtesting procedure follow the procedure given in the master's thesis of Monstvilaite (2016). We have adapted these steps to our application and outlined the procedure below:

1. Convert the data sample to daily log-returns

$$r_{t,j} = \ln \left(\frac{S_{t,j}}{S_{t-1,j}} \right),$$

where $S_{t,j}$ denotes the price of the j -th asset at day t . In our data sample we have a total of 1757 daily log-returns.

5. Empirical studies and analysis

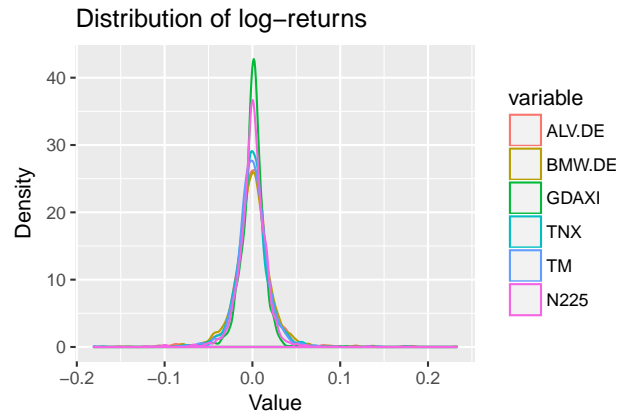


Figure 5.32.: Distribution of daily log-returns of all 6 given assets.

2. The data used for analysis starts on 01/01/2004 and ends on 07/01/2011, which results in a total of 1757 observations per asset. We estimate 814 one-day-ahead VaR and $CVaR$ forecast for the period of 01/01/2008 to 07/01/2011, for which we use data from the last 943 observations as illustrated in Figure 5.1. The separation into fitting period and testing period is in accordance to Table 5.1. The parameters for the GARCH(1,1) models and the vine copula models are either being estimated using only data from the fitting period $[1, \dots, 943]$ in the fixed model (Steps 3.a to 9.a) or re-estimated every 50 days using data from the last 943 observations in the 50 days refitted rolling window model (Steps 3.b to 9.b).

Backtest with fixed models:

In case of a fixed models we will follow Steps 3.a to 9.a and then proceed with Step 10.

- 3.a Fit a univariate GARCH(1,1) model with appropriate innovations separately for $r_{1:943,1}, \dots, r_{1:943,6}$ to convert the daily log-returns into an i.i.d series. Exemplary results for all six time series can be seen in Table 5.3. Using the determined GARCH(1,1) parameters and the remaining daily log-return series we calculate the remaining residuals

5. Empirical studies and analysis

and standard deviations by applying the equations given below. For each of the six assets denoted by j and $t \in [944, \dots, 1757]$ we apply the following:

$$\epsilon_{t,j} = r_{t,j} - \mu_j \quad (5.3)$$

$$\sigma_{t+1,j}^2 = \alpha_{0,j} + \alpha_{1,j}\epsilon_{t,j}^2 + \beta_{1,j}\sigma_{t,j}^2 \quad (5.4)$$

Here μ_j denotes the sample mean and $\epsilon_{t,j}$ denotes the error term of the GARCH(1,1) process.

- 4.a Extract the residuals (5.3) from Step 3.a and standardize them with the standard deviations (5.4) obtained from Step 3.a using:

$$z_{t,i} = \frac{\epsilon_{t,i}}{\sigma_{t,i}}$$

- 5.a Convert the standardized residuals $z_{1:943,1}, \dots, z_{1:943,6}$ to standard uniform residuals $u_{1:943,1}, \dots, u_{1:943,6}$ using the empirical cumulative distribution function according to (5.1) and (5.2). We apply the following equation for all six series:

$$u_{t,j} = \frac{n}{n+1} \hat{F}_j(z_{t,j})$$

- 6.a Fit an C-vine, D-vine or R-vine copula model to the multivariate data $u_{1:943,1}, \dots, u_{1:943,6}$ from the fitting period $[1, \dots, 943]$ using the model selection methods described in Chapter 4. We will predict the distribution for days in the interval $[944, \dots, 1757]$, using the determined vine structure.
- 7.a Generate $sim = 10^5$ simulations $u_1^{sim}, \dots, u_6^{sim}$ using the fitted vine model from the previous step. We will use the same sample for all future dates $t \in [944, \dots, 1757]$.
- 8.a Convert the simulated uniform marginals $u_1^{sim}, \dots, u_6^{sim}$ to standardized residuals using the inverse cumulative distribution function:

$$z_j^{sim} = \hat{F}_j^{-1}(u_j^{sim})$$

- 9.a Calculate returns from the simulated standardized residuals using the previously in Step 3.a calculated GARCH(1,1) coefficients and

5. Empirical studies and analysis

standard deviations. To forecast one-day-ahead for each asset we use:

$$\begin{aligned}\epsilon_{t,j} &= \sigma_{t,j} z_j^{sim} \\ r_{t+1,j}^* &= \mu_j + \epsilon_{t,j},\end{aligned}$$

where z_j^{sim} are the standardized residuals calculated in Step 8.a.

Backtest with 50 days refitted rolling window models:

In case of a 50 days refitted rolling window models we will follow Steps 3.b to 9.b and then proceed with Step 10. Steps 3.b to 9.b have to be repeated every 50 days, hence for all $t \in [943, 993, \dots, 1743]$.

- 3.b Fit a univariate GARCH(1,1) model with appropriate innovations separately for $r_{(t+1-943):t,1}, \dots, r_{(t+1-943):t,6}$ to convert the daily log-returns into an i.i.d series. The previously for each asset individually chosen error distributions of the GARCH(1,1) model is kept the same in all iterations to maintain uniformity in the method. This approach also makes the method less computationally intensive. Using the determined GARCH(1,1) parameters and the daily log-return series $r_{t:(t+49),1}, \dots, r_{t:(t+49),6}$ we calculate the following 50 residuals and standard deviations by applying the equations given below.

$$\epsilon_{t,j} = r_{t,j} - \mu_j \quad (5.5)$$

$$\sigma_{t+1,j}^2 = \alpha_{0,j} + \alpha_{1,j} \epsilon_{t,j}^2 + \beta_{1,j} \sigma_{t,j}^2 \quad (5.6)$$

Here μ_j denotes the sample mean and $\epsilon_{t,j}$ denotes the error term of the GARCH(1,1) process.

- 4.b Extract the residuals (5.5) from Step 3.b and standardize them with the standard deviations (5.6) obtained from Step 3.b using:

$$z_{t,i} = \frac{\epsilon_{t,i}}{\sigma_{t,i}}$$

- 5.b Convert the standardized residuals $z_{(t+1-943):t,1}, \dots, z_{(t+1-943):t,6}$ to standard uniform residuals $u_{(t+1-943):t,1}, \dots, u_{(t+1-943):t,6}$ using the empirical cumulative distribution function according to (5.1) and (5.2). We apply the following equation for all six series:

$$u_{t,j} = \frac{n}{n+1} \hat{F}_j(z_{t,j})$$

5. Empirical studies and analysis

- 6.b Fit an C-vine, D-vine or R-vine copula model to the multivariate data $\mathbf{u}_{(t+1-943):t,1}, \dots, \mathbf{u}_{(t+1-943):t,6}$ using the model selection methods described in Chapter 4. We will predict the distribution for days in the interval $[t+1, \dots, t+50]$, using the determined vine structure.
- 7.b Generate $sim = 30.000$ simulations $\mathbf{u}_1^{sim}, \dots, \mathbf{u}_6^{sim}$ using the fitted vine model from the previous step. We will use the same sample for the following 50 days.
- 8.b Convert the simulated uniform marginals $\mathbf{u}_1^{sim}, \dots, \mathbf{u}_6^{sim}$ to standardized residuals using the inverse cumulative distribution function:

$$\mathbf{z}_j^{sim} = \hat{F}_j^{-1}(\mathbf{u}_j^{sim})$$

- 9.b Calculate returns from the simulated standardized residuals using the previously in Step 3.b calculated GARCH(1,1) coefficients and standard deviations. To forecast one-day-ahead for each asset we use:

$$\begin{aligned}\epsilon_{t,j} &= \sigma_{t,j} \mathbf{z}_j^{sim} \\ r_{t+1,j}^* &= \mu_j + \epsilon_{t,j},\end{aligned}$$

where \mathbf{z}_j^{sim} are the standardized residuals calculated in Step 8.b.

10. Generate a series of simulated daily portfolio returns at time $t+1$:

$$\mathbf{X}_{t+1} = \frac{1}{6} \sum_{j=1}^6 r_{t+1,j}^*$$

The length of the vector $\mathbf{X}_{t+1} = (X_{t+1}^1, \dots, X_{t+1}^{sim})$ of simulated portfolio returns corresponds to the number of copula simulations sim .

11. Calculate

$$VaR_\alpha^{t+1}(\mathbf{X}_{t+1}) = q_\alpha(X_{t+1}^1, \dots, X_{t+1}^{sim})$$

and

$$CVaR_\alpha^{t+1}(\mathbf{X}_{t+1}) = \frac{1}{[sim \cdot \alpha]} \sum_{k=1}^{[sim \cdot \alpha]} X_{t+1, sim}^k$$

for $\alpha \in \{1\%, 3\%, 5\%\}$ using $X_{t+1, sim}^1 \leq X_{t+1, sim}^2 \leq \dots \leq X_{t+1, sim}^{sim}$ which is generated by sorting $X_{t+1}^1, \dots, X_{t+1}^{sim}$ and $[y] = \sup\{n \in \mathbb{N} : n \leq y\}$ for every $y \in \mathbb{R}$.

5. Empirical studies and analysis

This strategy results in the forecasting of VaR and CVaR of a portfolio of six co-dependent stocks in a portfolio. This is used to demonstrate the usefulness and the flexibility captured by the vine copula approach in an practical application to construct VaR and CVaR forecasts.

In a very similar way we can calculate VaR and CVaR for a multivariate Student's t-copula model, which is compared to the C-vine, D-vine and R-vine models too.

6. Results

This chapter features the practical results using the sample data as described in Chapter 5. The marginal distribution of the different assets are modeled by means of a GARCH(1,1) model as described in Section 5.6. Moreover the standardized uniform variables, which were obtained by converting the standardized residuals, are used to estimate C-vine, D-vine, R-vine and Student's t-copula models and to find the appropriate bivariate copula families and their parameters, as described in Chapter 3. Both a *fixed model approach* and a *50 days refitted rolling window approach* in the fitting procedure is used. As may be seen later, both approaches have advantages and disadvantages.

The estimated models and their parameters are then used to forecasting VaR_α and $CVaR_\alpha$ for $\alpha \in \{1\%, 3\%, 5\%\}$ as described in Section 5.6.

In this chapter the focus lies on the results and problems that emerge in the practical implementation of these kind of models. Special attention is turned to the computational performance and the accuracy of the estimated one-day-ahead VaR and CVaR using the backtesting procedures of Kupiec and Christoffersen for VaR and McNeil and Frey for CVaR for all considered copula models and refitting approaches.

For the sake of completeness traditional methods, especially the historical simulation (HS) and mean variance (MV) are also considered, regardless of their expected inferiority in forecasting VaR and CVaR of high-dimensional portfolios in comparison to vine copula models .

6. Results

6.1. Backtesting: Traditional methods - VaR and CVaR

As in the vine copula and multivariate Student's t-copula cases, we are using data from the last 943 days for the historical simulation (HS) and mean variance (MV) VaR and CVaR estimation for the following day (see Tabel 5.1). For details on the used HS and MV methods have been described at the beginning of this thesis, see Section 1.3.

Table 6.1.: VaR test - historical simulation vs. mean variance.

Method 1: Historical simulation					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	81	71	$p = 0.21518$ "Fail to Reject H_0 "	$p = 1.12 \times 10^{-06}$ "Reject H_0 "	
$VaR_{0.03}$	40	52	$p = 0.08086$ "Fail to Reject H_0 "	$p = 4.79 \times 10^{-05}$ "Reject H_0 "	
$VaR_{0.01}$	8	22	$p = 5.50 \times 10^{-05}$ "Reject H_0 "	$p = 9.63 \times 10^{-05}$ "Reject H_0 "	
Method 2: Mean variance					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	81	64	$p = 0.03537$ "Reject H_0 "	$p = 2.10 \times 10^{-07}$ "Reject H_0 "	
$VaR_{0.03}$	40	53	$p = 0.05822$ "Fail to Reject H_0 "	$p = 5.51 \times 10^{-05}$ "Reject H_0 "	
$VaR_{0.01}$	8	38	$p = 2.03 \times 10^{-14}$ "Reject H_0 "	$p = 1.11 \times 10^{-15}$ "Reject H_0 "	

Table 6.1 show the results for both the unconditional coverage (UC) rest and the conditional coverage (CC) test as described in Section 1.4.1 for all three tested levels of alpha.

In the case of a HS model we cannot reject the hypothesis that exceedance is correct for two of three α -levels. However the hypothesis H_0 is rejected in all overall conditional coverage (CC) tests and hence HS fails to properly forecast VaR for all three levels of alpha. MV failed to pass backtesting under all unconditional coverage (UC) tests but one and all conditional coverage (CC) tests.

Overall both traditional models have failed all (CC) tests for correct exceedance and independence and are therefore not suitable for accurately forecasting VaR for our six-dimensional portfolio of assets.

6. Results

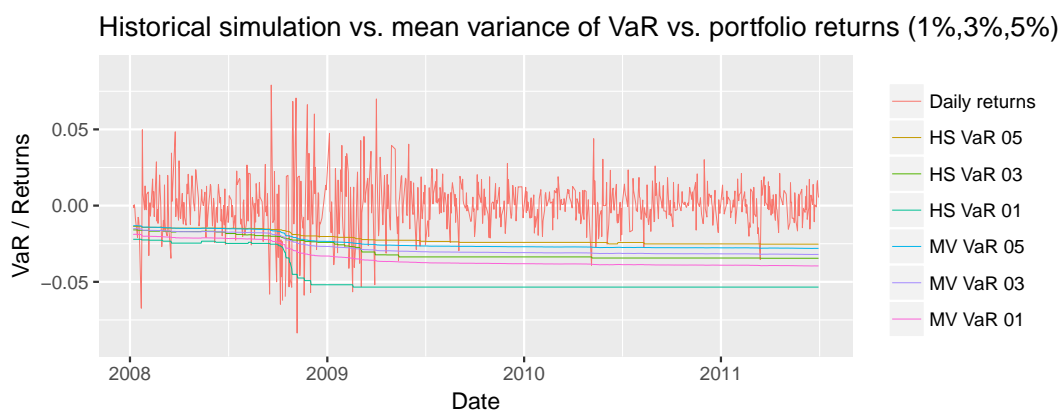


Figure 6.1.: VaR backtest: Historical simulation and mean variance.

In Figure 6.1 we can clearly see why both methods are failing miserably. Both approaches suffer from the so-called ghosting effect starting in late 2008 as larger loss events occur. HS and MV regard all observations used for the estimation of risk measures as equally important and therefore large loss observations from far back (as long as still part of the used data) still have importance for today's VaR and CVaR forecasts.

Using an age-weighted model, which gives larger weight to recent returns and at the same time smaller weight to distant returns, would make the VaR estimates more responsive to large loss observations and reduces the ghosting effect. Figure 6.2 illustrates VaR estimates using the age-weighted historical simulation (AWHS) method. For details on the used weights and the AWHS method in general see Boudoukh, Richardson, and Whitelaw (1997).

6. Results

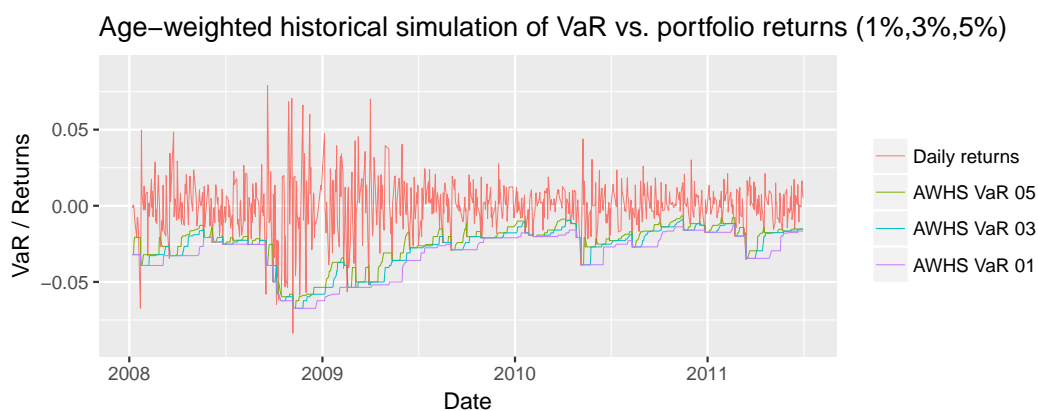


Figure 6.2.: VaR backtest: Age-weighted historical simulation.

CVaR is a risk measure for forecasting the average excess loss with respect to the VaR. Therefore we cannot discuss the CVaR forecast of the traditional models HS and MV which are not even suitable for accurately forecast VaR of a stock portfolio.

6.2. Backtesting: Fixed models - VaR and CVaR

We chose an estimation period of 943 days to fit each of the three vine copula models, the multivariate Student's t-copula model and GARCH(1,1) models which we all do not update, but keep fixed. It is to be expected that the model decreases in accuracy as time goes by. Especially with the beginning of the global economic crisis we expect a sudden drop in the quality of the predictions. Here we use 100.000 copula simulations per day.

6. Results

6.2.1. Fixed model: Value-at-Risk:

Table 6.3.: VaR test of fixed models - Student's t-copula vs. vine copulas.

Method 3: Student's t-copula					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	40	69	$p = 3.20 \times 10^{-05}$ "Reject H_0 "	$p = 2.50 \times 10^{-05}$ "Reject H_0 "	
$VaR_{0.03}$	24	44	$p = 0.00029$ "Reject H_0 "	$p = 0.00080$ "Reject H_0 "	
$VaR_{0.01}$	8	14	$p = 0.06114$ "Fail to Reject H_0 "	$p = 0.08545$ "Fail to Reject H_0 "	

Method 4: C-vine					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	40	69	$p = 3.20 \times 10^{-05}$ "Reject H_0 "	$p = 2.50 \times 10^{-05}$ "Reject H_0 "	
$VaR_{0.03}$	24	43	$p = 0.00055$ "Reject H_0 "	$p = 0.00156$ "Reject H_0 "	
$VaR_{0.01}$	8	13	$p = 0.11519$ "Fail to Reject H_0 "	$p = 0.12619$ "Fail to Reject H_0 "	

Method 5: D-vine					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	40	69	$p = 3.20 \times 10^{-05}$ "Reject H_0 "	$p = 2.50 \times 10^{-05}$ "Reject H_0 "	
$VaR_{0.03}$	24	43	$p = 0.00055$ "Reject H_0 "	$p = 0.00156$ "Reject H_0 "	
$VaR_{0.01}$	8	14	$p = 0.06114$ "Fail to Reject H_0 "	$p = 0.0855$ "Fail to Reject H_0 "	

Method 6: R-vine					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	40	69	$p = 3.20 \times 10^{-05}$ "Reject H_0 "	$p = 2.50 \times 10^{-05}$ "Reject H_0 "	
$VaR_{0.03}$	24	43	$p = 0.00055$ "Reject H_0 "	$p = 0.00156$ "Reject H_0 "	
$VaR_{0.01}$	8	13	$p = 0.11520$ "Fail to Reject H_0 "	$p = 0.12619$ "Fail to Reject H_0 "	

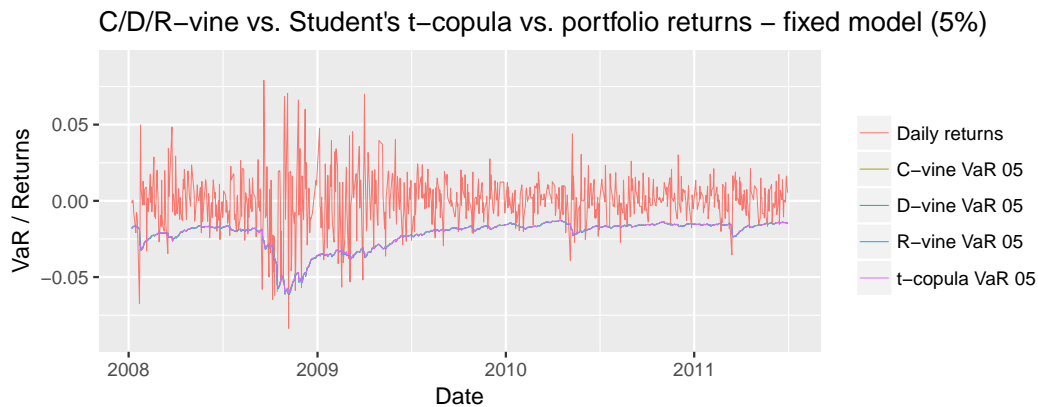


Figure 6.3.: VaR backtest: Fixed model - C/D/R-vine copula and Student's t-copula.

6. Results

Using a fixed setting we can observe that neither model seems to perform well. We found that all four models on one hand fail at accurately forecasting VaR for $\alpha \in \{5\%, 3\%\}$, but on the other hand still were able to pass the Kupiec and Christoffersen coverage test for $\alpha = 1\%$.

All four models have very similar VaR estimates, this can be also observed in Figure 6.3, which illustrates one-day-ahead VaR estimations with multivariate Student's t-copula and C/D/R-vine copula models using a level of $\alpha = 5\%$. From this data we see no obvious best model, we also don't see a clear improvement when using vine copula models over multivariate Student's t-copula models.

6.2.2. Fixed model: Conditional-Value-at-Risk:

We would also like to explore the ability of multivariate Student's t-copula models and vine copula models in estimating one-day-ahead CVaR. CVaR however is a risk measure for forecasting the average excess loss with respect to the VaR. Therefore we can only discuss the CVaR forecast in cases where we were able to properly forecast VaR.

In this specific case we were only able to forecast VaR at $\alpha = 1\%$. As seen in the summary of the McNeil and Frey zero mean test in Table 6.5, the CVaR predictions for $\alpha = 1\%$ were also sufficiently accurate.

6. Results

Table 6.5.: CVaR test of fixed models - Student's t-copula vs. vine copulas.

Method 3: Student's t-copula

Risk measure	Expected exceed.	Actual exceed.	H_1 "Mean of Excess Violations of VaR is greater than zero"
$CVaR_{0.05}$	40	69	$p = 0.22451$ "Fail to Reject H_0 " - VaR backtest failed!
$CVaR_{0.03}$	24	44	$p = 0.21922$ "Fail to Reject H_0 " - VaR backtest failed!
$CVaR_{0.01}$	8	14	$p = 0.09557$ "Fail to Reject H_0 "

Method 4: C-vine

Risk measure	Expected exceed.	Actual exceed.	H_1 "Mean of Excess Violations of VaR is greater than zero"
$CVaR_{0.05}$	40	69	$p = 0.26394$ "Fail to Reject H_0 " - VaR backtest failed!
$CVaR_{0.03}$	24	43	$p = 0.22301$ "Fail to Reject H_0 " - VaR backtest failed!
$CVaR_{0.01}$	8	13	$p = 0.06297$ "Fail to Reject H_0 "

Method 5: D-vine

Risk measure	Expected exceed.	Actual exceed.	H_1 "Mean of Excess Violations of VaR is greater than zero"
$CVaR_{0.05}$	40	69	$p = 0.20216$ "Fail to Reject H_0 " - VaR backtest failed!
$CVaR_{0.03}$	24	43	$p = 0.16849$ "Fail to Reject H_0 " - VaR backtest failed!
$CVaR_{0.01}$	8	14	$p = 0.07100$ "Fail to Reject H_0 "

Method 6: R-vine

Risk measure	Expected exceed.	Actual exceed.	H_1 "Mean of Excess Violations of VaR is greater than zero"
$CVaR_{0.05}$	40	69	$p = 0.23669$ "Fail to Reject H_0 " - VaR backtest failed!
$CVaR_{0.03}$	40	43	$p = 0.20933$ "Fail to Reject H_0 " - VaR backtest failed!
$CVaR_{0.01}$	8	13	$p = 0.06952$ "Fail to Reject H_0 "

C/D/R-vine vs. Student's t-copula vs. portfolio returns – fixed model (5%)

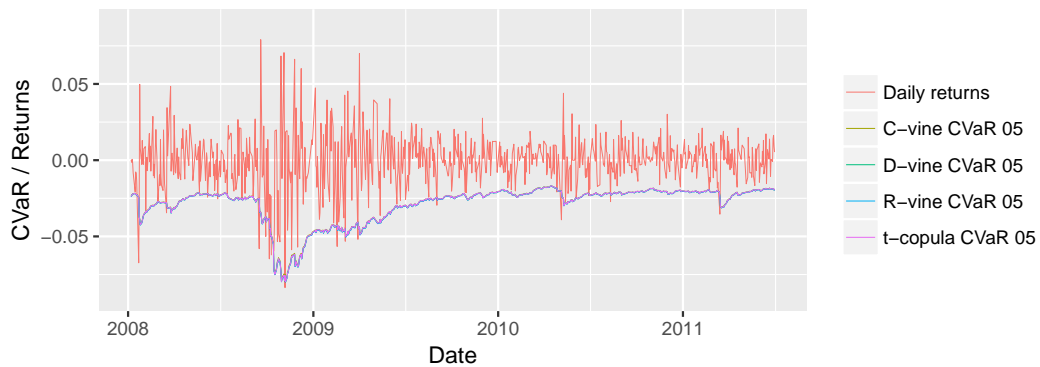


Figure 6.4.: CVaR backtest: Fixed model - C/D/R-vine copula and Student's t-copula.

6. Results

6.3. Backtesting: 50 days refitted rolling window models - VaR and CVaR

We chose a rolling window of 943 days to fit each of the three vine copula models and the multivariate Student's t-copula model. In the forecasting procedure we are refitting all models every 50 days. The process of calculating a new model takes a lot of computational resources, hence we are decreasing the number of simulations from 100.000 simulations per day used for the fixed model to 30.000 simulation per day. We can also observe that the vine copula models are computationally more intensive than the Student's t-copula model (for further details see Section 6.5).

6.3.1. 50 days refitted rolling window model: Value-at-Risk:

Table 6.7.: VaR test of rolling window models - Student's t-copula vs. vine copulas.

Method 3: Student's t-copula					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	40	44	$p = 0.60018$ "Fail to Reject H_0 "	$p = 0.49774$ "Fail to Reject H_0 "	
$VaR_{0.03}$	24	29	$p = 0.36042$ "Fail to Reject H_0 "	$p = 0.65781$ "Fail to Reject H_0 "	
$VaR_{0.01}$	8	9	$p = 0.76579$ "Fail to Reject H_0 "	$p = 0.21329$ "Fail to Reject H_0 "	

Method 4: C-vine					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	40	40	$p = 0.91012$ "Fail to Reject H_0 "	$p = 0.72438$ "Fail to Reject H_0 "	
$VaR_{0.03}$	24	23	$p = 0.76832$ "Fail to Reject H_0 "	$p = 0.87865$ "Fail to Reject H_0 "	
$VaR_{0.01}$	8	8	$p = 0.96055$ "Fail to Reject H_0 "	$p = 0.17660$ "Fail to Reject H_0 "	

Method 5: D-vine					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	40	41	$p = 0.96157$ "Fail to Reject H_0 "	$p = 0.68922$ "Fail to Reject H_0 "	
$VaR_{0.03}$	24	26	$p = 0.74796$ "Fail to Reject H_0 "	$p = 0.93351$ "Fail to Reject H_0 "	
$VaR_{0.01}$	8	8	$p = 0.96055$ "Fail to Reject H_0 "	$p = 0.17660$ "Fail to Reject H_0 "	

Method 6: R-vine					
Risk measure	Expected exceed.	Actual exceed.	H_0 "Correct Exceed."	H_0 "Correct Exceed. & Indep."	
$VaR_{0.05}$	40	46	$p = 0.40331$ "Fail to Reject H_0 "	$p = 0.24798$ "Fail to Reject H_0 "	
$VaR_{0.03}$	24	26	$p = 0.74796$ "Fail to Reject H_0 "	$p = 0.93351$ "Fail to Reject H_0 "	
$VaR_{0.01}$	8	8	$p = 0.96055$ "Fail to Reject H_0 "	$p = 0.17660$ "Fail to Reject H_0 "	

6. Results

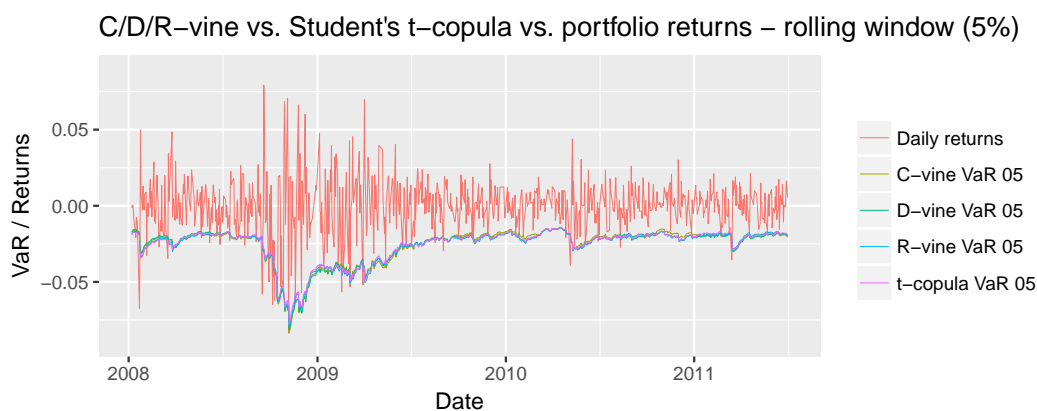


Figure 6.5.: VaR backtest: Rolling window - C/D/R-vine copula and Student's t-copula.

When using a 50 days refitted rolling window approach we can observe that all four model seems to perform well in forecasting VaR for $\alpha \in \{1\%, 3\%, 5\%\}$. All four models passed the Kupiec and Christoffersen coverage test.

All four models have very similar VaR estimates, this can also we observed in Figure 6.5, which illustrates one-day-ahead VaR estimations with multivariate Student's t-copula and C/D/R-vine copula models using a level of $\alpha = 5\%$. Similar to the fixed model approach we see no obvious best model, we also don't see a clear improvement when using vine copula models over multivariate Student's t-copula models.

6.3.2. 50 days refitted rolling window model: Conditional-Value-at-Risk:

We are now in the advantageous situation that we succeeded in forecasting suitable VaR estimates for all four copula models and all levels of α . Therefore we are able to meaningfully backtest all CVaR forecasts.

All four copula models were able to forecast CVaR for all levels of $\alpha \in \{1\%, 3\%, 5\%\}$, as seen in the summary of the McNeil and Frey zero mean test in Table 6.9.

6. Results

Table 6.9.: CVaR test of rolling window models - Student's t-copula vs. vine copulas.

Method 3: Student's t-copula

Risk measure	Expected exceed.	Actual exceed.	H_1 "Mean of Excess Violations of VaR is greater than zero"
$CVaR_{0.05}$	40	44	$p = 0.42913$ "Fail to Reject H_0 "
$CVaR_{0.03}$	24	29	$p = 0.41063$ "Fail to Reject H_0 "
$CVaR_{0.01}$	8	9	$p = 0.19516$ "Fail to Reject H_0 "

Method 4: C-vine

Risk measure	Expected exceed.	Actual exceed.	H_1 "Mean of Excess Violations of VaR is greater than zero"
$CVaR_{0.05}$	40	40	$p = 0.47909$ "Fail to Reject H_0 "
$CVaR_{0.03}$	24	23	$p = 0.35018$ "Fail to Reject H_0 "
$CVaR_{0.01}$	8	7	$p = 0.15917$ "Fail to Reject H_0 "

Method 5: D-vine

Risk measure	Expected exceed.	Actual exceed.	H_1 "Mean of Excess Violations of VaR is greater than zero"
$CVaR_{0.05}$	40	41	$p = 0.25024$ "Fail to Reject H_0 "
$CVaR_{0.03}$	24	26	$p = 0.35318$ "Fail to Reject H_0 "
$CVaR_{0.01}$	8	8	$p = 0.20777$ "Fail to Reject H_0 "

Method 6: R-vine

Risk measure	Expected exceed.	Actual exceed.	H_1 "Mean of Excess Violations of VaR is greater than zero"
$CVaR_{0.05}$	40	46	$p = 0.64325$ "Fail to Reject H_0 "
$CVaR_{0.03}$	24	26	$p = 0.34175$ "Fail to Reject H_0 "
$CVaR_{0.01}$	8	8	$p = 0.17987$ "Fail to Reject H_0 "

C/D/R-vine vs. Student's t-copula vs. portfolio returns – rolling window (5%)

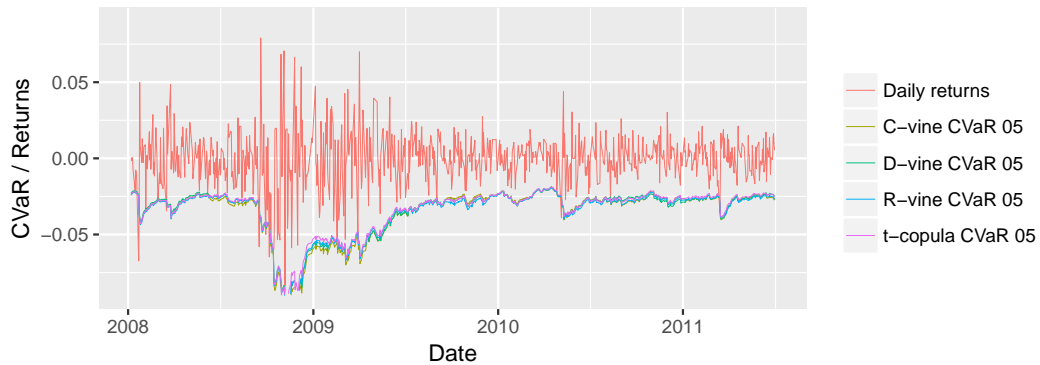


Figure 6.6.: VaR backtest: Rolling window - C/D/R-vine copula and Student's t-copula.

6. Results

6.4. Graphical comparison of 50 days refitted rolling window model vs. fixed model

Over time we expect the dependence structure of our portfolio to change. As we can see below this can lead to a significant deviation between the regularly refitted rolling window model and the fixed model approach, in both the vine copula models and the multivariate Student's t-copula model.

Student's t-copula vs. portfolio returns – rolling window vs. fixed (5%)



Figure 6.7.: VaR backtest: Rolling window vs. fixed - Student's t-copula.

C-vine copula vs. portfolio returns – rolling window vs. fixed (5%)



Figure 6.8.: VaR backtest: Rolling window vs. fixed - C-vine copula.

6. Results

In Figure 6.7 and 6.8 we see that both, the rolling window and the fixed model approaches, lead to similar results in the first half of 2008. However with the beginning of the global economic crisis the results, due to a probable change in the dependence structure, start to deviate significantly. After the financial crisis the results start to approach each other again.

Schafzahl (2018) investigated this problem for the fixed setting by separating the backtesting period into a *crisis period* and a *post crisis period* for this exact portfolio.

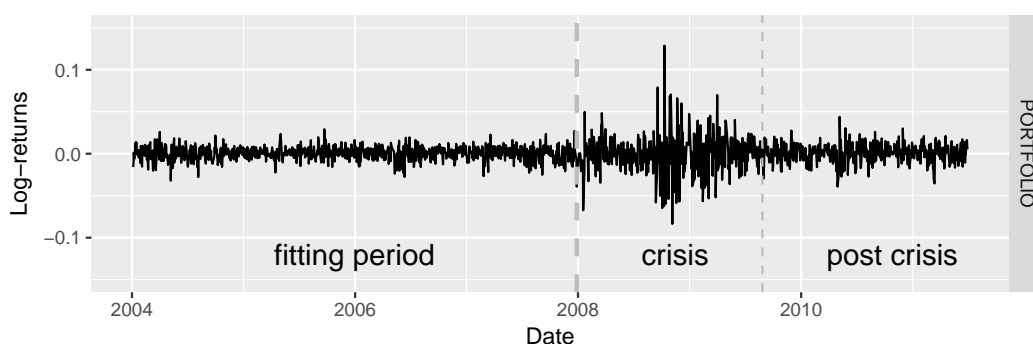


Figure 6.9.: Log-returns of equally weighted portfolio.

The estimated VaR and CVaR values did not lead to sufficiently good results in the *crisis period*, however the VaR and CVaR values in the *post crisis period* did. This lets us believe that the dependence structure of the examined six-dimensional portfolio in the fitting period (*pre crisis*) and the *post crisis period* pose similar characteristics, whereas the *crisis period* exhibits a different dependence structure.

The inability to adapt to changes in the data is one major drawback of the fixed model. The 50 days refitted rolling window model on the other hand was able to sufficiently adapt to the changes introduced by the global economic crisis of 2008 and returned adequate results throughout both the *crisis* and *post crisis* period.

6. Results

6.5. Runtime comparison

After comparing the forecasting quality of the discussed models, we will take a look at the runtime. As we can see below the performance of the approaches differs quite significantly.

All numerical experiments were run on a Intel(R) Core(TM) i7-3635QM CPU @2.4GHz 4 real cores + 4 virtual, 16 GB RAM. The algorithms were implemented in R using a 64 bit Windows 10 operating system. Keep in mind that R is not utilizing multi-threading by default, hence parallelization of the code might improve the performance significantly.

	runtime[sec]	simulations	refit[days]
HS	11.95	0	0
MV	3.97	0	0
t-copula fixed	36.51	100000	0
C-vine copula fixed	56.01	100000	0
D-vine copula fixed	60.75	100000	0
R-vine copula fixed	60.18	100000	0
t-copula rolling	69.89	30000	50
C-vine copula rolling	264.71	30000	50
D-vine copula rolling	263.31	30000	50
R-vine copula rolling	268.41	30000	50

Table 6.11.: Comparison of runtime for all approaches and different simulation count.

Table 6.11 gives an overview of the runtime to calculate one-day-ahead VaR_α and $CVaR_\alpha$ for $\alpha \in \{1\%, 3\%, 5\%\}$ for a backtesting period of 814 days for all models with different simulation count. Especially the fixed models achieve a good runtime while however struggling to maintain consistent performance in forecasting. Best precision was archived by the C-vine, D-vine and R-vine copula models with a rolling window using 30.000 simulations. Schafzahl (2018) compared the results using 3.000, 30.000 and 100.000 and found out that 30.000 simulations presents a good trade-off between precision and runtime. However the increase in runtime when

6. Results

increasing the number of simulations is significant. The increase from 3.000 simulations to 30.000 simulations leads to a 30% increase in runtime, whereas the increase from 3.000 simulations to 100.000 simulations leads to an increase in runtime of around 100%.

Overall, the runtime of the 50 days refitted rolling window approach was around 4 – 7 times higher than the runtime of the fixed models. A reduction of the days between two refits would additionally increase the runtime drastically.

Conclusions & outlook

In this master's thesis, we investigated the risk measures VaR_α and $CVaR_\alpha$ for $\alpha \in \{1\%, 3\%, 5\%\}$ of a six-dimensional portfolio by using three different vine copula models, a multivariate Student's t-copula model and two commonly used traditional methods historical simulation and mean variance. We then backtested the calculated risk measures with the Kupiec and Christoffersen coverage test and the McNeil and Frey zero mean test. In the first step we modeled the marginal distribution of all six assets using a GARCH(1,1) model. Using C-vine, D-vine and R-vine copula models we were able to model the interdependence of these six assets. Taking the corresponding vine copula specification we simulated returns for each asset using the Monte Carlo method and thereof forecasted both risk measures (VaR_α and $CVaR_\alpha$) for the six-dimensional portfolio.

In order to compare the vine copula forecasted risk measures VaR_α and $CVaR_\alpha$, we also did a series of forecasts using historical simulation (HS), mean variance (MV) and multivariate Student's t-copula. For the modeling with vine copulas and Student's t-copula we used both a fixed setting and a rolling window setting of 943 days which was refitted every 50 days.

The main findings we obtained through calculating the VaR_α and $CVaR_\alpha$ of the six-dimensional portfolio with all six different models can be summarized as follows:

Traditional methods (historical simulation and mean variance):

- VaR measured by using the traditional methods HS and MV is not accurate according to the conditional coverage (CC) test. Therefore these models are not suitable to forecast VaR for our six-dimensional portfolio of assets.

6. Results

Fixed models:

- When looking at the entire *backtesting period* the C-vine, D-vine and R-vine copula models and the multivariate Student's t-copula model were only able to effectively forecast VaR_α and $CVaR_\alpha$ for our six-dimensional portfolio according to the conditional coverage test (CC) and the zero mean test for $\alpha = 1\%$.
- The separation of the backtesting period into a *crisis period* and a *post crisis period* indicated that the fixed models have major difficulties with respect to the adaption to changes in the data. The models failed to predict VaR and CVaR in the *crisis period*. The inability of these model to adapt to changed dependence structure is a major drawback of the fixed models.¹
- Nevertheless, the fixed models yield quite good results in periods of similar characteristics and dependence structure in the fitting and backtesting period. We were able to successfully forecast VaR and CVaR in the *post crisis period* using a model which was fitted just before the crisis.

50 days refitted rolling window models:

- The refitted rolling window approach yields good results in predicting VaR and CVaR in all four copula models and all levels of α .
- We witness a 5 to 8 fold increase in runtime for the 50 days rolling window approach in comparison to the fixed model setting depending on the number of used simulations. When only comparing rolling window approaches, using a baseline of 3.000 simulations, an increase to 30.000 simulations lead to an increase in runtime of around 30% whereas 100.000 simulations lead to an increase of around 100% increase.

Other observations that were made while writing this thesis:

- In all cases, using either log-likelihood, AIC or BIC for testing the model-fit, the multivariate Student's t-copula performed slightly worse

¹For further details and finding on the separation of a fixed model into two periods with different characteristics and interdependencies see the Master's Project Schafzahl (2018).

6. Results

than all three vine copula models. The C-vine and R-vine copula models performed slightly better than the D-vine copula model. However, we saw no significant improvement in using vine copula models over Student's t-copula models in forecasting risk measures of our six-dimensional portfolio.

- Copulas allow to model marginal distributions and the dependence structure separately.
- Vine copulas are easily constructed: any bivariate copula can be used as building blocs.
- Pair-copula-construction models such as C-vine, D-vine and R-vine copula models allow for a very flexible and intuitive way of constructing higher-dimensional copulas.
- Simulation and inference are straight-forward, however it can be very time-consuming in higher dimensions.
- Copulas are a very promising and flexible tool that still need a lot of research and development.
- Vine copulas can be very useful in financial risk management.
- There are a few very useful R package available, e.g. *VineCopula* and *CDVine*, which allows for sequential and maximum likelihood parameter estimation for C-vine, D-vine and R-vine copulas. This package also includes tools for model selection, simulation, goodness-of-fit tests, and visualization.

The main shortcoming of VaR in risk measurement is the inability to meet the requirement of sub-additivity, which means that the VaR of a portfolio may be higher than the VaR of the sum of the individual assets. CVaR, which is based on VaR was introduced to overcome this shortcoming. Despite the importance of CVaR in risk management, existing literature rarely discusses CVaR forecasting with C-vine, D-vine or R-vine models. With this thesis we have tried to combine CVaR forecasting with various vine copula models on one hand and also with a multivariate Student's t-copula model on the other hand. Overall, in our setting, we can say that vine copula models and multivariate Student's t-copula models are more accurate and reliable in predicting VaR and CVaR than traditional models, which failed entirely.

With increasing dimensions we expect the vine copula models, due to its increased flexibility, to perform superior to the multivariate Student's

6. Results

t-copula model. Albeit, some further research into this topic has to be conducted.

The number of parameters in a vine structure grows exponentially with the dimension. One way of reducing the computation time of such a model might be the reduction of the complexity of the vine structure. Brechmann (2010) studied the problem of determining whether a vine structure may be truncated. The truncation of a vine structure at level T refers to a vine structure where all bivariate copulas with conditioning set larger than or equal to T are set to independence copulas. This procedure might lead to a reduction in time needed to fit and simulate from a vine copula.

Overall vine copula models constitute a powerful and flexible class of high-dimensional dependency models, which is not only useful in financial risk management but also in a wide range of other applications such as medicine, biology, imaging or actuarial science.

Appendix

Appendix A.

Proofs, calculations etc.

A.1. Likelihood ratio test

This section is based on the lecture "Mathematische Statistik" held by Friedl (2015).

Definition 31 (Likelihood function)

Let $f(X_1, \dots, X_n | \theta)$ be the joint density or probability function of a data sample X_1, \dots, X_n . Given $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ is observed, the function of θ defined by

$$L(\theta | x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta)$$

is called likelihood function.

If comparing the likelihood function at two parameter points θ_1 and θ_2 finds that $L(\theta_1 | x) > L(\theta_2 | x)$ then the observed sample is more likely to have occurred under $\theta = \theta_1$ than under $\theta = \theta_2$. Hence, θ_1 is a more plausible value for the real parameter θ than θ_2 .

Let X_1, \dots, X_n be i.i.d. random variables from $f(x | \theta)$, then the likelihood function is given as:

$$L(\theta | x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta)$$

Appendix A. Proofs, calculations etc.

Definition 32 (Likelihood ratio test)

The likelihood ratio test statistic for the test

$$H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \in \Theta_0^C,$$

where Θ_0 is a subset of the parameter space Θ and Θ_0^C is the compliment of Θ_0 i.e. $\Theta \setminus \Theta_0$, is given as

$$\lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta | x_1, \dots, x_n)}{\sup_{\theta \in \Theta} L(\theta | x_1, \dots, x_n)} \in [0, 1].$$

The likelihood ratio test is a test with a rejection region R of the form

$$R = \{x_1, \dots, x_n | \lambda(x_1, \dots, x_n) \leq c\}$$

for any $c \in [0, 1]$.

A small $\lambda(x_1, \dots, x_n)$ is a indication against the null hypothesis H_0 , whereas for a large $\lambda(x_1, \dots, x_n)$ the null hypothesis cannot be rejected.

Now let $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ be the maximum likelihood estimate of θ and let $\hat{\theta}_0 = \hat{\theta}_0(x_1, \dots, x_n)$ be the maximum likelihood estimate of θ under the restricted parameter space Θ_0 . Using this setting the likelihood ratio test statistic is

$$\lambda(x_1, \dots, x_n) = \frac{L(\hat{\theta}_0 | x_1, \dots, x_n)}{L(\hat{\theta} | x_1, \dots, x_n)}.$$

A.2. Calculations for likelihood ratio test of unconditional coverage (UC)

A.2.1. Maximum likelihood estimate of π for the likelihood function of the alternative hypothesis

We can solve for the maximum likelihood estimates by taking the first derivative with respect to π and setting the derivative to zero. The likelihood function is given by:

$$L(\pi|\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) = (1 - \pi)^{n_0} \pi^{n_1}$$

Taking the first derivative with respect to π yields:

$$\begin{aligned} \frac{\partial}{\partial \pi} L(\pi|\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) \\ = -(1 - \pi)^{n_0-1} n_0 \pi^{n_1} + (1 - \pi)^{n_0} \pi^{n_1-1} n_1 \end{aligned}$$

Setting the derivative to zero gives us an estimate for $\hat{\pi}$:

$$\begin{aligned} \frac{\partial}{\partial \pi} L(\pi|\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) &\stackrel{!}{=} 0 \\ -(1 - \pi)^{n_0-1} n_0 \pi^{n_1} + (1 - \pi)^{n_0} \pi^{n_1-1} n_1 &\stackrel{!}{=} 0 \\ \pi^{n_1-1} (1 - \pi)^{n_0-1} (\pi n_0 - (1 - \pi) n_1) &\stackrel{!}{=} 0 \end{aligned}$$

This equation holds if one of the three factors on the left side is equal to 0. By setting the third factor to 0 and excluding $\pi \in \{0, 1\}$ we get the desired result.

$$\begin{aligned} \pi n_0 - (1 - \pi) n_1 &\stackrel{!}{=} 0 \\ \pi(n_0 + n_1) - n_1 &\stackrel{!}{=} 0 \\ \hat{\pi} = \frac{n_1}{n_0 + n_1} &= \frac{n_1}{n} \end{aligned}$$

A.3. Calculations for likelihood ratio test of independence (IND)

A.3.1. Maximum likelihood estimate of π_{01} and π_{11} for the likelihood function of the alternative hypothesis

We can solve for the maximum likelihood estimates by taking the first derivatives with respect to π_{01} and π_{11} and setting these derivatives to zero. The likelihood function for the first-order Markov process is given by:

$$L(\Pi_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{10}} (1 - \pi_{11})^{n_{01}} \pi_{11}^{n_{11}}$$

Taking the first derivative with respect to π_{01} yields:

$$\begin{aligned} \frac{\partial}{\partial \pi_{01}} L(\Pi_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) \\ = \left[-(1 - \pi_{01})^{n_{00}-1} n_{00} \pi_{01}^{n_{10}} + (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{10}-1} n_{10} \right] (1 - \pi_{11})^{n_{01}} \pi_{11}^{n_{11}} \end{aligned}$$

Setting the derivative to zero gives us an estimate for $\hat{\pi}_{01}$:

$$\begin{aligned} \frac{\partial}{\partial \pi_{01}} L(\Pi_1 | \mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) &\stackrel{!}{=} 0 \\ \left[-(1 - \pi_{01})^{n_{00}-1} n_{00} \pi_{01}^{n_{10}} + (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{10}-1} n_{10} \right] (1 - \pi_{11})^{n_{01}} \pi_{11}^{n_{11}} &\stackrel{!}{=} 0 \\ -(1 - \pi_{01})^{n_{00}-1} n_{00} \pi_{01}^{n_{10}} + (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{10}-1} n_{10} &\stackrel{!}{=} 0 \\ \pi_{01}^{n_{10}-1} (1 - \pi_{01})^{n_{00}-1} (\pi_{01} n_{00} - (1 - \pi_{01}) n_{10}) &\stackrel{!}{=} 0 \end{aligned}$$

This equation holds if one of the three factors on the left side is equal to 0. By setting the third factor to 0 and excluding $\pi_{01} \in \{0, 1\}$ we get the desired result.

$$\begin{aligned} \pi_{01} n_{00} - (1 - \pi_{01}) n_{10} &\stackrel{!}{=} 0 \\ \pi_{01} (n_{00} + n_{10}) - n_{10} &\stackrel{!}{=} 0 \\ \hat{\pi}_{01} &= \frac{n_{10}}{n_{00} + n_{10}} \end{aligned}$$

In the same way it is easily possible to calculate

$$\hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}.$$

A.3.2. Maximum likelihood estimate of π for the likelihood function of the zero hypothesis

We can solve for the maximum likelihood estimates by taking the first derivative with respect to π and setting the derivative to zero. The likelihood function is given by:

$$L(\Pi|\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) = (1 - \pi)^{n_{00} + n_{10}} \pi^{n_{01} + n_{11}}$$

Taking the first derivative with respect to π yields:

$$\begin{aligned} \frac{\partial}{\partial \pi} L(\Pi|\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) \\ = -(1 - \pi)^{n_{00} + n_{10} - 1} (n_{00} + n_{10}) \pi^{n_{01} + n_{11}} + (1 - \pi)^{n_{00} + n_{10}} \pi^{n_{01} + n_{11} - 1} (n_{01} + n_{11}) \end{aligned}$$

Setting the derivative to zero gives us an estimate for $\hat{\pi}$:

$$\begin{aligned} \frac{\partial}{\partial \pi} L(\Pi|\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n) &\stackrel{!}{=} 0 \\ -(1 - \pi)^{n_{00} + n_{10} - 1} (n_{00} + n_{10}) \pi^{n_{01} + n_{11}} + (1 - \pi)^{n_{00} + n_{10}} \pi^{n_{01} + n_{11} - 1} (n_{01} + n_{11}) &\stackrel{!}{=} 0 \\ \pi^{n_{01} + n_{11} - 1} (1 - \pi)^{n_{00} + n_{10} - 1} (\pi(n_{00} + n_{10}) - (1 - \pi)(n_{01} + n_{11})) &\stackrel{!}{=} 0 \end{aligned}$$

This equation holds if one of the three factors on the left side is equal to 0. By setting the third factor to 0 and excluding $\pi \in \{0, 1\}$ we get the desired result.

$$\begin{aligned} \pi(n_{00} + n_{10}) - (1 - \pi)(n_{01} + n_{11}) &\stackrel{!}{=} 0 \\ \pi(n_{00} + n_{01} + n_{10} + n_{11}) - (n_{01} + n_{11}) &\stackrel{!}{=} 0 \\ \hat{\pi} &= \frac{n_{01} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}} \end{aligned}$$

A.4. Derivation of h-functions for different copulas

The derivations of the h-functions of most copulas discussed in this thesis can be found in Aas et al. (2006). For the sake of completeness we include the computations of the h-functions of the six copulas involved in this thesis. The h-function is defined as

$$h(x, v, \Theta) = F(x|v) = \frac{\partial C_{xv}(x, v, \Theta)}{\partial v},$$

where Θ represents the set of parameters for the copula associated with the joint distribution function of x and v .

A.4.1. Gaussian copula

The distribution function of the bivariate Gaussian copula is given as

$$C_{\rho}^{Ga}(u, v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \underbrace{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\}}_{=g(x,y)} dx dy.$$

To save some notational work we use $g(x, y)$ and for the bounds we use

$$b_1 = \phi^{-1}(u) \text{ and } b_2 = \phi^{-1}(v),$$

where $\phi(\cdot)$ is the standard normal distribution function.

For this copula we have:

$$\begin{aligned} h(u, v, \rho) &= F_{1|2}(u|v) \\ &= \frac{\partial}{\partial v} C_{\rho}^{Ga}(u, v) \\ &= \frac{\partial}{\partial v} \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} g(x, y) dy dx \\ &= \frac{\partial b_2}{\partial v} \frac{\partial}{\partial b_2} \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} g(x, y) dy dx \end{aligned}$$

Appendix A. Proofs, calculations etc.

Using

$$\frac{\partial b_2}{\partial v} = \frac{\partial}{\partial v} \phi^{-1}(v) = \frac{1}{f(b_2)},$$

where

$$f(z) = \frac{1}{2\pi} \exp\left(-\frac{z^2}{2}\right)$$

is the density of the standard normal distribution and

$$\phi(h) = \int_{-\infty}^h f(z) dz,$$

we get:

$$\begin{aligned} h(u, v, \rho) &= \frac{1}{f(b_2)} \frac{\partial}{\partial b_2} \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} g(x, y) dy dx \\ &= \frac{1}{f(b_2)} \int_{-\infty}^{b_1} \left[\frac{\partial}{\partial b_2} \int_{-\infty}^{b_2} g(x, y) dy \right] dx \\ &= \frac{1}{f(b_2)} \int_{-\infty}^{b_1} g(x, b_2) dx \\ &= \frac{1}{f(b_2)} \int_{-\infty}^{b_1} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho x b_2 + b_2^2}{2(1-\rho^2)}\right\} dx \\ &= \frac{1}{f(b_2)} \int_{-\infty}^{b_1} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{(x - \rho b_2)^2 + (b_2^2 - \rho^2 b_2^2)}{2(1-\rho^2)}\right\} dx \\ &= \frac{1}{f(b_2)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{b_2^2}{2}\right\} \int_{-\infty}^{b_1} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{(x - \rho b_2)^2}{2(1-\rho^2)}\right\} dx \\ &= \frac{1}{f(b_2)} \frac{f(b_2)}{1} \phi\left(\frac{b_1 - \rho b_2}{\sqrt{1-\rho^2}}\right) \\ &\stackrel{\substack{b_1 = \phi^{-1}(u) \\ b_2 = \phi^{-1}(v)}}{=} \phi\left(\frac{\phi^{-1}(u) - \rho\phi^{-1}(v)}{\sqrt{1-\rho^2}}\right) \end{aligned}$$

A.4.2. Student's t-copula

Using the some techniques as previously used for the Gaussian copula we can derive the h-function for the Student's t-copula.

The distribution function of the bivariate Student's t-copula is given as

$$C_{\rho, \nu}^t(u, v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \underbrace{\frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^2(1-\rho^2)}} \left\{ 1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)} \right\}^{-\frac{\nu+2}{2}}}_{=g(x,y)} dx dy,$$

where ν represents the degrees of freedom and $\rho \in (-1, 1)$ is the linear correlation coefficient.

To save some notational work we use $g(x, y)$ and for the bounds we use

$$b_1 = t_v^{-1}(u)$$

and

$$b_2 = t_v^{-1}(v).$$

Using

$$\frac{\partial b_2}{\partial v} = \frac{\partial}{\partial v} t_v^{-1}(v) = \frac{1}{f_\nu(b_2)},$$

where

$$f_\nu(z) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \left(1 + \frac{z^2}{\nu} \right)^{-\frac{\nu+1}{2}}$$

for the density distribution function of a Student's t-distribution with ν degrees of freedom and

$$t_\nu(h) = \int_{-\infty}^h f_\nu(z) dz.$$

Appendix A. Proofs, calculations etc.

we get:

$$\begin{aligned}
h(u, v, \rho, \nu) &= F_{1|2}(u|v) \\
&= \frac{\partial}{\partial v} C_{\rho, \nu}^t(u, v) \\
&= \frac{\partial}{\partial v} \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} g(x, y) dy dx \\
&= \frac{\partial b_2}{\partial v} \frac{\partial}{\partial b_2} \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} g(x, y) dy dx \\
&= \frac{1}{f_\nu(b_2)} \frac{\partial}{\partial b_2} \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} g(x, y) dy dx \\
&= \frac{1}{f_\nu(b_2)} \int_{-\infty}^{b_1} \left[\frac{\partial}{\partial b_2} \int_{-\infty}^{b_2} g(x, y) dy \right] dx \\
&= \frac{1}{f_\nu(b_2)} \int_{-\infty}^{b_1} g(x, b_2) dx \\
&= \frac{1}{f_\nu(b_2)} \int_{-\infty}^{b_1} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{(\pi\nu)^2(1-\rho^2)}} \left[1 + \frac{x^2 - 2\rho x b_2 + b_2^2}{\nu(1-\rho^2)} \right]^{-\frac{\nu+2}{2}} dx \\
&= \frac{1}{f_\nu(b_2)} \int_{-\infty}^{b_1} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{(\pi\nu)^2(1-\rho^2)}} \left[1 + \frac{(x - \rho b_2)^2}{(\nu + b_2^2)(1-\rho^2)} \right]^{-\frac{\nu+2}{2}} \left[1 + \frac{b_2^2}{\nu} \right]^{-\frac{\nu+2}{2}} dx \\
&= \frac{1}{f_\nu(b_2)} \frac{\Gamma(\frac{\nu+1}{2}) \sqrt{\pi(\nu + b_2^2)(1-\rho^2)}}{\Gamma(\frac{\nu}{2}) \sqrt{(\pi\nu)^2(1-\rho^2)}} \left[1 + \frac{b_2^2}{\nu} \right]^{-\frac{\nu+1}{2}} \left[1 + \frac{b_2^2}{\nu} \right]^{-\frac{1}{2}} \\
&\quad \times \int_{-\infty}^{b_1} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2}) \sqrt{\pi(\nu + b_2^2)(1-\rho^2)}} \left[1 + \frac{(x - \rho b_2)^2}{(\nu + b_2^2)^2(1-\rho^2)} \right]^{-\frac{\nu+2}{2}} dx \\
&= \frac{1}{f_\nu(b_2)} \underbrace{\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\pi\nu}} \left[1 + \frac{b_2^2}{\nu} \right]^{-\frac{\nu+1}{2}}}_{=f_\nu(b_2)} \underbrace{\frac{\sqrt{\pi(\nu + b_2^2)(1-\rho^2)}}{\sqrt{(\pi\nu)(1-\rho^2)}} \left[1 + \frac{b_2^2}{\nu} \right]^{-\frac{1}{2}}}_{=1} \\
&\quad \times \int_{-\infty}^{b_1} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2}) \sqrt{\pi(\nu + b_2^2)(1-\rho^2)}} \left[1 + \frac{(x - \rho b_2)^2}{(\nu + b_2^2)^2(1-\rho^2)} \right]^{-\frac{\nu+2}{2}} dx \\
&= \int_{-\infty}^{b_1} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2}) \sqrt{\pi(\nu + b_2^2)(1-\rho^2)}} \left[1 + \frac{(x - \rho b_2)^2}{(\nu + b_2^2)^2(1-\rho^2)} \right]^{-\frac{\nu+2}{2}} dx
\end{aligned}$$

Appendix A. Proofs, calculations etc.

For notational purposes we set

$$\begin{aligned}\hat{\nu} &= \nu + 1 \\ \mu &= \rho b_2 \\ \sigma &= \sqrt{\frac{(\nu + b_2^2)(1 - \rho^2)}{\nu + 1}}.\end{aligned}$$

$$\begin{aligned}h(u, v, \rho, \nu) &= \int_{-\infty}^{b_1} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2})\sqrt{\pi(\nu + b_2^2)(1 - \rho^2)}} \left[1 + \frac{(x - \rho b_2)^2}{(\nu + b_2^2)(1 - \rho^2)} \right]^{-\frac{\nu+2}{2}} dx \\ &= \int_{-\infty}^{b_1} \frac{\Gamma(\frac{\hat{\nu}+1}{2})}{\Gamma(\frac{\hat{\nu}}{2})\sqrt{\pi\hat{\nu}\sigma}} \left[1 + \frac{1}{\hat{\nu}} \left(\frac{x - \mu}{\sigma} \right)^2 \right]^{-\frac{\hat{\nu}+1}{2}} dx \\ &= \int_{-\infty}^{b_1} \frac{1}{\sigma} f_{\hat{\nu}} \left(\frac{x - \mu}{\sigma} \right) dx \\ &= \int_{-\infty}^{(b_1 - \mu)/\sigma} f_{\hat{\nu}}(z) dz \\ &= t_{\hat{\nu}} \left(\frac{b_1 - \mu}{\sigma} \right)\end{aligned}$$

After inserting the expressions for $b_1, \mu, \hat{\nu}$ and σ we finally have

$$h(u, v, \rho, \nu) = t_{\nu+1} \left(\frac{t_{\nu}^{-1}(u) - \rho t_{\nu}^{-1}(v)}{\sqrt{\frac{(\nu + (t_{\nu}^{-1}(v))^2)(1 - \rho^2)}{\nu + 1}}} \right).$$

A.4.3. Clayton copula

The distribution function of the Clayton copula is given as

$$C_{\theta}^{Cl}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \theta \in (0, \infty).$$

For this copula we have:

$$\begin{aligned} h(u, v, \theta) &= F_{1|2}(u|v) \\ &= \frac{\partial}{\partial v} C_{\theta}^{Cl}(u, v) \\ &= \frac{\partial}{\partial v} (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \\ &= v^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-1-1/\theta} \end{aligned}$$

A.4.4. Gumbel copula

The distribution function of the Gumbel copula is given as

$$C_{\theta}^{Gu}(u, v) = \exp \left\{ -[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta} \right\}, \theta \in [1, \infty).$$

For this copula we have:

$$\begin{aligned} h(u, v, \theta) &= F_{1|2}(u|v) \\ &= \frac{\partial}{\partial v} C_{\theta}^{Gu}(u, v) \\ &= \frac{\partial}{\partial v} \exp \left\{ -[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta} \right\} \\ &= v^{-1} \exp \left\{ -[(-\log u)^{\theta} + (-\log v)^{\theta}]^{1/\theta} \right\} \left(1 + \left(\frac{-\log u}{-\log v} \right)^{\theta} \right)^{-1+1/\theta} \end{aligned}$$

A.4.5. Frank copula

The distribution function of the Frank copula is given as

$$C_{\theta}^{Fr}(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right), \theta \in (0, \infty).$$

For this copula we have:

$$\begin{aligned} h(u, v, \theta) &= F_{1|2}(u|v) \\ &= \frac{\partial}{\partial v} C_{\theta}^{Fr}(u, v) \\ &= \frac{\partial}{\partial v} \left[-\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \right] \\ &= \frac{\partial}{\partial v} \left[-\frac{1}{\theta} \ln \left(\frac{(1 - e^{-\theta}) - (1 - e^{-\theta u})(1 - e^{-\theta v})}{1 - e^{-\theta}} \right) \right] \\ &= -\frac{1}{\theta} \left(\frac{1 - e^{-\theta}}{(1 - e^{-\theta}) - (1 - e^{-\theta u})(1 - e^{-\theta v})} \right) \frac{(-\theta)(1 - e^{-\theta u})e^{-\theta v}}{1 - e^{-\theta}} \\ &= \frac{e^{-\theta v}}{\frac{1 - e^{-\theta}}{1 - e^{-\theta u}} + e^{-\theta v} - 1} \end{aligned}$$

A.4.6. Joe copula

The distribution function of the Joe copula is given as

$$C_{\theta}^{Jo}(u, v) = 1 - \left[(1 - u)^{\theta} + (1 - v)^{\theta} - (1 - u)^{\theta}(1 - v)^{\theta} \right]^{1/\theta}, \theta \in [1, \infty).$$

For this copula we have:

$$\begin{aligned} h(u, v, \theta) &= F_{1|2}(u|v) \\ &= \frac{\partial}{\partial v} C_{\theta}^{Jo}(u, v) \\ &= \frac{\partial}{\partial v} \left[1 - \left[(1 - u)^{\theta} + (1 - v)^{\theta} - (1 - u)^{\theta}(1 - v)^{\theta} \right]^{1/\theta} \right] \\ &= \left((1 - u)^{\theta} + (1 - v)^{\theta} - (1 - u)^{\theta}(1 - v)^{\theta} \right)^{1/\theta - 1} (1 - v)^{\theta - 1} (1 - (1 - u)^{\theta}) \end{aligned}$$

A.5. Inverse Laplace-Stiltjes transforms F

Here we will state some thoughts and explanations to the inverse Laplace-Stiltjes transforms F listed in Table 3.1.

A.5.1. Clayton copula

We have given two different generators which both generate a Clayton copula C_θ^{Cl} . For this application we use the generator $\phi(t) = t^{-\theta} - 1$ which corresponds to $\phi(t)^{-1} = (t + 1)^{-\frac{1}{\theta}}$.

Definition 33 (Gamma distribution)

A random variable X that follows a Gamma distribution with shape parameter $\alpha \in (0, \infty)$ and scale parameter $\beta \in (0, \infty)$ is denoted by $X \sim \Gamma(\alpha, \beta)$. The probability density function is defined as

$$f_X(x|\alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, x \in [0, \infty).$$

For $X \sim \Gamma(1/\theta, 1)$ this leads to $f_X(x|1/\theta, 1) = \frac{1}{\Gamma(1/\theta)} x^{\frac{1}{\theta}-1} e^{-x}$. Using the Laplace-Stiltjes transform we can verify the stated inverse Laplace-Stiltjes transform F for the Clayton copula:

$$\begin{aligned} \mathbb{E}[e^{-tX}] &= \int_0^\infty e^{-tx} \frac{1}{\Gamma(1/\theta)} x^{\frac{1}{\theta}-1} e^{-x} dx \\ &= \int_0^\infty \frac{e^{-(t+1)x} x^{\frac{1}{\theta}-1}}{\Gamma(1/\theta)} dx \\ &\stackrel{(t+1)x=s}{=} \frac{1}{\Gamma(1/\theta)} \int_0^\infty e^{-s} \frac{s^{\frac{1}{\theta}-1}}{(t+1)^{\frac{1}{\theta}-1} (t+1)} ds \\ &= \frac{(t+1)^{-\frac{1}{\theta}}}{\Gamma(1/\theta)} \underbrace{\int_0^\infty e^{-s} s^{\frac{1}{\theta}-1} ds}_{=\Gamma(1/\theta)} \\ &= (t+1)^{-\frac{1}{\theta}} = \phi^{-1}(t) \end{aligned}$$

A.5.2. Gumbel copula

Definition 34 (Stable distribution)

Let $St(\alpha, \beta, \gamma, \delta)$ denote a Stable distribution with characteristic exponent $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, scale $\gamma \in [0, \infty)$ and location parameter $\delta \in \mathbb{R}$ defined by the characteristic function

$$\begin{aligned} \phi(t) &= \exp(i\delta t - \gamma^\alpha |t|^\alpha (1 - i\beta \operatorname{sgn}(t)\Phi(t, \alpha))), \quad t \in \mathbb{R}, \\ \text{with } \Phi(t, \alpha) &= \begin{cases} \tan(\frac{\alpha\pi}{2}), & \alpha \neq 1, \\ -\frac{2}{\pi} \log(|t|), & \alpha = 1, \end{cases} \text{ and} \\ \operatorname{sgn}(t) &= \mathbb{1}_{[0, \infty)}(t) - \mathbb{1}_{(-\infty, 0]}(t), \quad t \in \mathbb{R}. \end{aligned}$$

For $X \sim St(1/\theta, 1, (\cos(\frac{\pi}{2\theta}))^\theta, 0)$, $t \in [0, \infty)$ and $\alpha \neq 1$ the characteristic function simplifies to:

$$\begin{aligned} \phi_X(t) &= \mathbb{E}[\exp(itX)] = \exp(-\cos(\frac{\pi}{2\theta})|t|^{1/\theta}(1 - i \operatorname{sgn}(t)\Phi(t, 1/\theta))) \\ &\stackrel{\alpha \neq 1}{=} \exp(-\cos(\frac{\pi}{2\theta})|t|^{1/\theta}(1 - i \operatorname{sgn}(t) \tan(\frac{\pi}{2\theta}))) \\ &\stackrel{t \in [0, \infty)}{=} \exp(-t^{1/\theta} (\underbrace{\cos(\frac{\pi}{2\theta}) - i \cos(\frac{\pi}{2\theta}) \tan(\frac{\pi}{2\theta})}_{=\sin(\frac{\pi}{2\theta})})) \\ &\quad \underbrace{= \exp(-i\frac{\pi}{2\theta}) = (\exp(-\frac{i\pi}{2}))^{1/\theta} = (-i)^{1/\theta}}_{=} \\ &= \exp((-t)^{1/\theta} (-i)^{1/\theta}) = \exp((it)^{1/\theta}) \quad (\text{A.1}) \end{aligned}$$

Formally $\mathcal{LS}[F](t) = \phi_X(it)$, hence

$$\begin{aligned} \mathcal{LS}[F](t) &= \phi_X(it) \\ &\stackrel{(\text{A.1})}{=} \exp((iit)^{1/\theta}) = \exp(-t^{1/\theta}), \end{aligned}$$

which corresponds to the inverse generator of the Gumbel copula.

Although the closed form of the density of stable distribution is not known, Nolan (2011) proposed a simulation algorithm for generating stable random variables.

A.5.3. Frank copula

For Frank's copula and Joe's copula F is discrete with a given probability mass function $(y_k)_{k \in \mathbb{N}}$ for $k \in \mathbb{N}$.

Definition 35 (Logarithmic series distribution)

A Random variable K follows a logarithmic series distribution $\text{Log}(\alpha)$ with shape parameter $\alpha \in (0, 1)$ if K has a discrete distribution on \mathbb{N}^+ with probability mass function

$$y_k = \frac{1}{-\ln(1-\alpha)} \frac{\alpha^k}{k}, k \in \mathbb{N}^+.$$

The inverse Laplace-Stiltjes transform F for Frank's copula corresponds to a logarithmic series distribution with shape parameter $\alpha = (1 - e^{-\theta})$.

A algorithm for generating logarithmic series variables can be found in Kemp (1981).

A.5.4. Joe copula

Definition 36 (Discrete Sibuya distribution)

The Sibuya distribution $\text{Sibuya}(\alpha)$ for $\alpha \in (0, 1]$ can be defined either by its Laplace-Stiltjes transform

$$1 - (1 - \exp(-t))^\alpha, t \in [0, \infty),$$

its distribution function

$$F(k) = 1 - (-1)^k \binom{\alpha - 1}{k}, k \in \mathbb{N},$$

or its probability mass function

$$y_k = \binom{\alpha}{k} (-1)^{k-1}, k \in \mathbb{N},$$

where $\binom{\alpha}{k} = \alpha(\alpha - 1) \dots (\alpha - k + 1) / k!$ denotes the (generalized) binomial coefficient.

Appendix A. Proofs, calculations etc.

The inverse Laplace-Stiltjes transform F for Joe's copula corresponds to a Sibuya distribution with parameter $\alpha = 1/\theta$.

For further details on the *Sibuya*(α) distribution see Kozubowski and Podgorski (2016).

Bibliography

- Aas, Kjersti et al. (2006). *Pair-copula constructions of multiple dependence*. 487. Discussion paper // Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München (cit. on pp. 1, 54, 65, 67, 70, 72, 73, 81, 139).
- Akaike, Hirotugu (1974). "A New Look at the Statistical Model Identification." In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723 (cit. on p. 77).
- Ali, Ghulam (2013). "EGARCH, GJR-GARCH, TGARCH, AVGARCH, NGARCH, IGARCH and APARCH Models for Pathogens at Marine Recreational Sites." In: *Journal of Statistical and Econometric Methods* 2.3, pp. 57–73 (cit. on p. 39).
- Andersson, Jonas (2001). "On the Normal Inverse Gaussian Stochastic Volatility Model." In: *Journal of Business and Economic Statistics* 19.1, pp. 44–54 (cit. on p. 38).
- Artzner, Philippe, Freddy Delbaen, and Jean-Marc Ebner (1999). "Coherent Measure of Risk." In: *Mathematical Finance* 9.3, pp. 203–228 (cit. on pp. 4, 6, 8, 21).
- Basel Committee on Banking Supervision (1996). *Amendment to the Capital Accord to Incorporate market Risk* (cit. on p. 8).
- Basel Committee on Banking Supervision (2009). *Revisions to the Basel II market risk framework*. ISBN: 92-9131-774-8 (cit. on p. 6).
- Basel Committee on Banking Supervision (2013). *Fundamental review of the trading book: A revised market risk framework* (cit. on p. 8).
- Baumol, William J. (1963). "An Expected Gain-Confidence Limit Criterion for Portfolio Selection." In: *Management Science* 10.1, pp. 174–182 (cit. on p. 5).

Bibliography

- Bedford, Tim and Roger Cooke (2001). "Probability density decomposition for conditional dependent random variables modeled by vines." In: *Annals of Mathematics and Artificial Intelligence* 32.1, pp. 245–268 (cit. on pp. 70, 71).
- Bedford, Tim and Roger Cooke (2002). "Vines - A new graphical model for dependent random variables." In: *The Annals of Statistics* 30.4, pp. 1031–1068 (cit. on pp. 64, 70).
- Berkowitz, Jeremy, Peter Christoffersen, and Denis Pelletier (2011). "Evaluating Value-at-Risk Models with Desk-Level Data." In: *Management Science* 57.12, pp. 2213–2227 (cit. on p. 13).
- Bollerslev, Tim (1986). "Generalized Autoregressive Conditional Heteroscedasticity." In: *Journal of Econometrics* 31.3, pp. 307–327 (cit. on pp. 27, 33).
- Bollerslev, Tim (1987). "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." In: *The Review of Economics and Statistics* 69.3, pp. 542–547 (cit. on p. 37).
- Boudoukh, Jacob, Matthew Richardson, and Robert F. Whitelaw (1997). *The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk*. Working paper series (New York University. Salomon Center). New York University, Leonard N. Stern School of Business (cit. on pp. 10, 117).
- Brechmann, Eike Christian (2010). "Truncated and simplified regular vines and their applications." MA thesis. Center for Mathematical Science, Technische Universität München (cit. on pp. 27, 44, 45, 78, 79, 99, 132).
- Campbell, Sean D. (2005). *A review of backtesting and backtesting procedures*. Finance and Economics Discussion Series 2005-21. Board of Governors of the Federal Reserve System (U.S.) (cit. on p. 13).
- Christoffersen, Peter F. (1998). "Evaluating Interval Forecasts." In: *International Economic Review* 39.4, pp. 841–862 (cit. on p. 13).
- Cont, Rama (2001). "Empirical properties of asset returns: stylized facts and statistical issues." In: *Quantitative Finance* 1, pp. 223–236 (cit. on p. 29).
- Demarta, Stefano and Alexander McNeil (2005). "The t Copula and Related Copulas." In: *International Statistical Review* 73.1, pp. 111–129 (cit. on p. 107).
- Dißmann, Jeffrey F. (2010). "Statistical Inference for Regular Vines and Application." MA thesis. Center for Mathematical Science, Technische Universität München (cit. on pp. 27, 70, 75, 100).
- Dragoti-Čela, Eranda (2009). *Risikotheorie und -management VO*. Department of Discrete Mathematics - Graz University of Technology (cit. on p. 26).

Bibliography

- Embrechts, Paul, Filip Lindskog, and Alexander McNeil (2003). *Modelling Dependence with Copulas and Applications to Risk Management*. Ed. by Elsevier Rachev. *Handbook of Heavy Tailed Distributions in Finance*. Chap. 8, pp. 329–384 (cit. on pp. 23, 57).
- Engel, Robert F. (1982). “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.” In: *Econometrica* 50.4, pp. 987–1007 (cit. on p. 27).
- Fernandez, Carmen and F. J. Mark Steel (1998). “On Bayesian Modeling of Fat Tails and Skewness.” In: *Journal of the American Statistical Association* 93.441, pp. 359–371 (cit. on p. 37).
- Friedl, Herwig (2015). *Lecture notes in Mathematische Statistik VO*. Institute of Statistics - Graz University of Technology (cit. on p. 134).
- Genest, Christian (1987). “Frank’s family of bivariate distributions.” In: *Biometrika* 74.3, pp. 549–555 (cit. on p. 55).
- Ghalanos, Alexios (2015). *rugarch: Univariate GARCH Models*. URL: <https://CRAN.R-project.org/package=rugarch> (cit. on p. 83).
- Hahsler, Michael and Kurt Hornik (2017). *TSP: Traveling Salesperson Problem (TSP)*. URL: <https://CRAN.R-project.org/package=TSP> (cit. on p. 83).
- Harper, David (2017). *An Introduction to Value at Risk (VAR)*. [Online; accessed Juli 12, 2017]. URL: <http://www.investopedia.com/articles/04/092904.asp> (cit. on pp. 10, 11).
- Hobæk Haff, Ingrid (2013). “Parameter estimation for pair-copula constructions.” In: *Bernoulli* 19.2, pp. 462–491 (cit. on p. 81).
- Hobæk Haff, Ingrid, Kjersti Aas, and Arnaldo Frigessi (2010). “On the simplified pair-copula construction — Simply useful or too simplistic?” In: *Journal of Multivariate Analysis* 101.5, pp. 1296–1310 (cit. on p. 66).
- Hofert, Marius (2008). “Sampling Archimedean Copulas.” In: *Computational Statistics & Data Analysis* 52.12, pp. 5163–5174 (cit. on pp. 58, 60, 61).
- Hofert, Marius, Ivan Kojadinovic, et al. (2017). *copula: Multivariate Dependence with Copulas*. URL: <https://CRAN.R-project.org/package=copula> (cit. on p. 83).
- Hofert, Marius, Martin Mächler, and Alexander J. McNeil (2012). “Likelihood inference for Archimedean copulas in high dimensions under known margins.” In: *Journal of Multivariate Analysis* 110, pp. 133–150 (cit. on p. 56).

Bibliography

- Jaworski, P. et al. (2010). *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*. Lecture Notes in Statistics. Springer Berlin Heidelberg. ISBN: 9783642124655 (cit. on p. 61).
- Joe, Harry (1996). "Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters." In: *Lecture Notes-Monograph Series* 28 (cit. on pp. 64, 67).
- Joe, Harry (1997). *Multivariate Models and Multivariate Dependence Concepts*. Vol. 73. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC. ISBN: 9780412073311 (cit. on pp. 1, 23, 44, 54).
- Kemp, A. W. (1981). "Efficient Generation of Logarithmically Distributed Pseudo-Random Variables." In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 30.3, pp. 249–253. ISSN: 00359254, 14679876. URL: <http://www.jstor.org/stable/2346348> (cit. on p. 148).
- Kozubowski, J. Tomasz and Krzysztof Podgorski (2016). "A generalized Sibuya distribution." In: *Working Papers in Statistics No 2016:8 8* (cit. on p. 149).
- Krämer, Nicole and Ulf Schepsmeier (2011). *Introduction to vine copulas*. NIPS Workshop, Granada (cit. on p. 65).
- Kupiec, Paul H. (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models." In: *The Journal of Derivatives* 3.2, pp. 73–84 (cit. on p. 13).
- Kurowicka, Dorota (2011). "Optimal truncation of vines." In: *Dependence Modeling: Vine Copula Handbook*. Ed. by D. Kurowicka and H. Joe. World Scientific Publishing Co. Pte. Ltd. Chap. 11, pp. 233–247. ISBN: 978-981-4299-87-9 (cit. on p. 75).
- Ljung, G. M. and G. E. Box (1978). "On a measure of lack of fit in time series models." In: *Biometrika* 65.2, pp. 297–303 (cit. on p. 38).
- Mandelbrot, Benoît (1963). "The Variation of Certain Speculative Prices." In: *The Journal of Business* 36.4, pp. 394–419 (cit. on p. 92).
- Manner, Hans (2007). *Estimation and Model Selection of Copulas with an Application to Exchange Rates*. Working paper: Research Memorandum 056. Maastricht University, Maastricht Research School of Economics of Technology and Organization (METEOR) (cit. on p. 78).
- Marshall, Albert W. and Ingram Olkin (1988). "Families of Multivariate Distributions." In: *Journal of the American Statistical Association* 83.403, pp. 834–841 (cit. on p. 59).

Bibliography

- Mashal, Roy and Assaf Zeevi (2002). "Beyond Correlation: Extreme Comovements Between Financial Assets." In: *SSRN Electronic Journal* (cit. on p. 107).
- McNeil, Alexander J. and Rüdiger Frey (2000). *Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach*. Vol. 7. 3-4. *Journal of Empirical Finance*, pp. 271–300 (cit. on pp. 13, 21).
- McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts (2006). *Quantitative Risk Management: Concepts, Techniques and Tools*. Vol. Revised Edition. Princeton University Press. ISBN: 9780691166278 (cit. on pp. 3, 25, 35, 40, 41, 45, 46, 48, 58, 60).
- Monstvilaite, Monika (2016). "Portfolio Value-at-Risk Using Regular Vine Copulas." MA thesis. Mathematical Statistics, Stockholm University (cit. on p. 109).
- Morales-Nápoles, Oswaldo, Rodger M. Cooke, and Dorota Kurowicka (2010). "About The Number of Vines and Regular Vines on n Nodes." In: *Submitted to Applied Discrete Mathematics* (cit. on pp. 64, 65, 74).
- Nelsen, Rodger B. (2006). *An Introduction to Copulas*. Vol. 2. Springer-Verlag New York. ISBN: 978-0-387-28659-4 (cit. on pp. 40, 49).
- Nolan, John P (2011). *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser. Chap. 1. URL: <http://academic2.american.edu/~jpnolan/stable/chap1.pdf> (cit. on pp. 60, 147).
- Orskaug, Elisabeth (2009). "Multivariate DCC-GARCH Model - With Various Error Distributions." MA thesis. Norwegian University of Science and Technology (cit. on p. 27).
- Pflug, Georg Ch. (2000). "Some Remarks on the Value-at-Risk and the Conditional-Value-at-Risk." In: *Probabilistic Constrained Optimization: Methodology and Applications*. Ed. by Stanislav P. Uryasev. Dordrecht: Kluwer Academic Publishers. ISBN: 0-7923-6644-1 (cit. on pp. 7, 8).
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/> (cit. on p. 82).
- Rosenkrantz, Daniel J., Richard E. Stearns, and II Philip M. Lewis (1977). "An Analysis of Several Heuristics for the Traveling Salesman Problem." In: *SIAM Journal on Computing* 6.3, pp. 563–581 (cit. on p. 100).
- Rossi, Eduardo (2004). *Lecture notes on GARCH Models*. University of Pavia (cit. on p. 39).

Bibliography

- Schafzahl, Nikolaus (2018). "Measuring portfolio risk: Elected issues using vine copula models and multivariate t-copula models." MA Project. Technische Universität Graz (cit. on pp. 92, 126, 127, 130).
- Schepsmeier, Ulf, Jakob Stoeber, and Eike Christian Brechmann (2017). *VineCopula: Statistical Inference of Vine Copulas*. URL: <http://CRAN.R-project.org/package=VineCopula> (cit. on p. 82).
- Sklar, Abe (1959). "Fonctions de répartition à n dimensions et leurs marges." In: *Publication de l'Institut de Statistique de L'Université de Paris* 8, pp. 229–231 (cit. on p. 40).
- Wickham, Hadley and Winston Chang (2016). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. URL: <http://CRAN.R-project.org/package=ggplot2> (cit. on p. 83).
- Wilks, Samuel S. (1938). "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses." In: *The Annals of Mathematical Statistics* 9.1, pp. 60–62 (cit. on p. 16).
- Xu, Chao and Huigeng Chen (2012). "Measuring Portfolio Value at Risk." MA thesis. Department of Economics - School of Economics and management, Lund University (cit. on p. 8).