DEGREE PROJECT IN MATHEMATICS,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2017

The GARCH-copula model for gaugeing time conditional dependence in the risk management of electricity derivatives

JOHAN VIKTORSSON

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ENGINEERING SCIENCES

The GARCH-copula model for gaugeing time conditional dependence in the risk management of electricity derivatives

JOHAN VIKTORSSON

Degree Projects in Scientific Computing (30 ECTS credits)
Degree Programme in Engineering Physics
KTH Royal Institute of Technology year 2017
Supervisor at Nasdaq: Mikhail Nechaev
Supervisor at KTH: Camilla Johansson Landén
Examiner at KTH: Camilla Johansson Landén

TRITA-MAT-E 2017:49
ISRN-KTH/MAT/E--17/49--SE

Royal Institute of Technology
School of Engineering Sciences
KTH SCI
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci

Sammanfattning

GARCH-copula modellen för att uppskatta tidsbetingat beroende vid riskhanteringen av eletricitetsderivat

Vid riskhantering av elektrictitetsderivat kan, tid till leverans delas upp I ett rutnät med antagandet att volatiliteten kan anses konstant för varje ruta i nätet. Detta upplägg tar emellertid inte hänsyn till beroende mellan de olika rutorna i rutnätet.
Detta examensarbeta försöker att utveckla en metod för att uppskatta detta beroende för eletricitetsderivat som befinner sig i på olika platser i rutnätet och som har olika leveransperioder. Mer specifikt är målet att uppskatta kvoten mellan kvantilen av summerade prisförändringar mot summan av de marginella kvantilerna hos prisförändringar.
Angreppsättet är en kombination av så kallade Generelised Autoregressive Conditional Heteroscedasticity (GARCH) och så kallade copulas. GARCH processen används för att filtrera ut heteroskedicitet i prisdatan. Copulas passas till den filtrerade via pseduo maximum likelihood och ett test av anpassningens kvalitet tillämpas.
GARCH processer allena visar sig vara otillräckliga för att fånga dynamiken i prisdatan. Det visar sig att en kombination av GARCH och autoregressive moving avergae (ARMA) processer ger en bättre anpassning till data. Det resulterande beroendet visar sig fångas bäst av elliptiska copulas.
Den skattade kvoten visar sig vara rätt liten i de studerade fallen. Användingen av ARMA-GARCH visar sig också ge en bättre anpassning till copulas när de används till finansiell data. En tidsbetingning i beroendet kan också observeras.

Abstract

The GARCH-copula model for gaugeing time conditional dependence in the risk management of electricity derivatives

In the risk management of electricity derivatives, time to delivery can be divided into a time grid, with the assumption that within each cell of the grid, volatility is more or less constant. This setup however does not take in to account dependence between the different cells in the time grid.

This thesis tries to develop a way to gauge the dependence between electricity derivatives at the different places in the time grid and different delivery periods. More specifically, the aim is to estimate the size of the ratio of the quantile of the sum of price changes against the sum of the marginal quantiles of the price changes.

The approach used is a combination of Generalised Autoregressive Conditional Heteroscedasticity (GARCH) processes and copulas. The GARCH process is used to filter out heteroscedasticity in the price data. Copulas are fitted to the filtered data using pseudo maximum likelihood and the fitted copulas are evaluated using a goodness of fit test.

GARCH processes alone are found to be insufficient to capture the dynamics of the price data. It is found that combining GARCH with Autoregressive Moving Average processes provides better fit to the data. The resulting dependence is the found to be best captured by elliptical copulas. The estimated ratio is found to be quite small in the cases studied. The use of the ARMA-GARCH filtering gives in general a better fit for copulas when applied to financial data. A time dependency in the dependence can also be observed.

Acknowledgements

I would like to thank Mikhail Nechaev for giving me the opportunity for writing my thesis for Nasdaq Clearing and always taking time to explain and giving input during the processes.

I would also like to thank Camilla Johansson Landén at KTH for supervising me and providing input.

Most of all I would like to thank Maria for all the support and understanding during all of this.

Last, and least, I would like to thank my friends. If it were not for all of you, I would have done this a long time ago.

Stockholm, June 2017

Johan Viktorsson

Contents

Chapter 1

Introduction

The aim of this thesis is to investigate the dependency between price changes of derivatives written on electricity with different time to delivery from a risk management perspective. The goal in mind is to investigate the ratio between the quantile of the sum of price changes versus the the quantile of the marginal price changes. To be mathematically specific, this thesis will try to estimate the ratio

$$\frac{F_{X+Y}^{-1}}{F_X^{-1} + F_Y^{-1}} \tag{1.1}$$

where X and Y are stochastic variables that are not necessarily independent nor necessarily identically distributed. F_X^{-1} is here the quantile function for the random variable X The stochastic variables X and Y will in this thesis be possible price changes of a financial contract from one day t to another $t-1$ e.g.

$$X = P_t - P_{t-1}$$

for some financial contract with the time dependent price P_t.

This problem in itself is in not new. The problem was first formulated by Kolmogorov and a theoretical solution for an upper bound was provided by [13] and [7] expanded upon the solution. However, their methods result in a complex problem to solve for arbitrary marginals and high dimensions, so their approach will not be used. Instead, the approach will be a combination of GARCH models and copulas, with the dynamics of the log returns of the contracts being modelled by GARCH and dependency through a copula.

The disposition of the thesis will be the following. First, the problem will be presented, with a brief background, assumptions made and a presentation of the data used. In the subsequent chapter, the theory of GARCH models and copulas will be briefly discussed, together with a multiplier based goodness of fit test for copulas and further mathematical formulations that will be of aid in the solution of the problem. The third chapter will be devoted to an explanation of the selected model and assumptions made. In the fourth chapter a selection of results will be presented. As the reader will soon be made aware, the problem is quite comprehensive and therefore no complete exposé of the results can be made in finite time (or page numbers). At the end there will be a chapter with a discussion of the results and a reference list.

1.1 The electricity derivative market

Nasdaq Clearing AB is a clearing house and provides clearing for a wide range of electricity derivatives, among them futures, forwards, options and electricity price area differentials (EPADS). Nasdaq Clearing acts as a Central Counter Party (CCP) for these contracts. This means that the clearing house (Nasdaq Clearing) acts as the counter party for both sides. That means that regardless if an investor has a long (has bought the contract) or short position (has sold the contracts) in a contract, Nasdaq Clearing is the contractual counter party. The reason for this is to minimize the consequences of a default, failure to meet the contractual obligations, of a participant. Should a participant default, Nasdaq Clearing ensures that the contract is honoured. Thus removing a sizeable portion of the credit risk for the participants.

We shall in this thesis, focus on forwards and futures on electricity, since these are the one highest in both open interest and trading volume.

A derivative is a standardized contract (written on an underlying) between two parties with a pay off function. Since this thesis will focus on futures and forwards, let us briefly discuss these contracts. When two parties enter into a future or forward contract, they agree on a price today for delivery of an underlying (eg. stocks, oil, seafood, electricity) on a later date, the delivery date. Such a contract may be traded on an exchange, and it is always possible to exit the contract (by for example, entering a contract in the opposite direction). As the contract can be traded on an exchange, there is a possibility of a gain/loss for the participants. For a futures contract, the losses and gains are settled on a daily basis and for a forward contract losses/gains are settled on delivery. This implies that forwards have a market value while futures do not.

In order to participate in the market, a participant, regardless whatever the participant is long or short, has to post collateral, henceforth called margin. The purpose of the margin is to ensure that the participants can handle sudden price changes in their positions. Should the amount of collateral deposited be insufficient due to market movements, a so called margin call is issued, and the participant has a set amount of time to either post more collateral or close positions so the amount of margin is sufficient. The margin is often in the form of cash, but can also be in the form of bonds, stocks etcetera.

For derivatives on electricity, the quoted price on the exchange is for one

1MVh[1]. Different contracts have different volumes. There are contracts available for the delivery of electricity for all days of a certain week, month, quarter and year. The contracts can also specify when during the day the electricity should be delivered. This can for example be during peak load (office hours) or during base load (all hours of the day).
At first, it might seem strange that one would want to trade with derivatives on electricity, since electricity can not (in a reasonable way) be stored. However, electricity is in fact a great candidate as underlying for a derivative due to the fact that quality is not an issue, it is power in the electricity grid, regardless of who produced it or how. To make this matter clearer we give the following example, from the perspective of an electricity producer taken from [2].
Consider a producer of electricity with a power plant. The producer has the choice to sell the electricity on the so called *spot market* or on the *forwards market*. For simplicity, assume that the power plant can produce 1 MW at all times. On September 1st there is a forward contract available for delivery of electricity for all hours of the month of April the following year. The price of the contract is 40 €/MWh. The producer chooses to enter this contract, as the price is what the producer would like to sell for. At the last day of March, the price of the contract has risen to 45 €/MWh, so the producer have lost 5 €/MWh. Say, again for simplicity, that the spot price for electricity is constant at 30 €/MWh for all hours in April. Then, since the producer lost 5 €/MWh on the forward contract, and the spot price remains constant, the total price in the end is 45-5+30-30=40 €/MWh. Which was what the producer aimed for in the beginning. If the producer instead exits the contract right before settlement, then the producer still have to pay 5 €/MWh as loss on the forward and will have to sell the electricity on the spot market. The gain is now 30-5=25 €/MWh. It is of course, a bit unrealistic that the producer would sell the entire production capability on the derivatives market and that spot price for electricity would be constant during an entire month. The idea with the above example is to show the usefulness of electricity derivatives as hedging instruments for producers and consumers of electricity. This is by no means a complete covering of different derivatives. For the reader interested to know more, a good introduction to the subject is given in [11].

1.2 Margin and Time Buckets

Currently Nasdaq Clearing employs the Nordic SPAN model for the calculation of margin. A key feature of the SPAN model is the so called time buckets, or time grid. The volatility of commodity futures and forwards has been found empirically to be dependent on time to delivery. As a contract

[1] Mega Watt hour

moves closer to delivery, the volatility of the contract increases. Volatility is simply a contract's ability to change in price. Volatility is, by its nature, an unobserved quantity. There are several ways to measure volatility. One way the volatility could be measured is by using the relative price increments

$$\frac{P_t - P_{t-1}}{P_t}$$

To make use of the fact that the volatility increases as the contract moves closer to delivery when calculating the margin, time to delivery is divided into a grid, where every cell is called a time bucket. The assumption is then that within each time bucket, volatility can be seen as more or less constant. The time buckets are constructed so that the beginning of the time bucket is included while the end is not. In a more mathematically formulated way, this means that if the time bucket begins when the time left to delivery is t_1 and ends when the time left to delivery is t_2, then the time bucket can be written as the interval

$$[t_1, t_2)$$

where $t_1 < t_2$. Figure 1.1 shows a time series of prices for a contract that has been divided into time buckets.

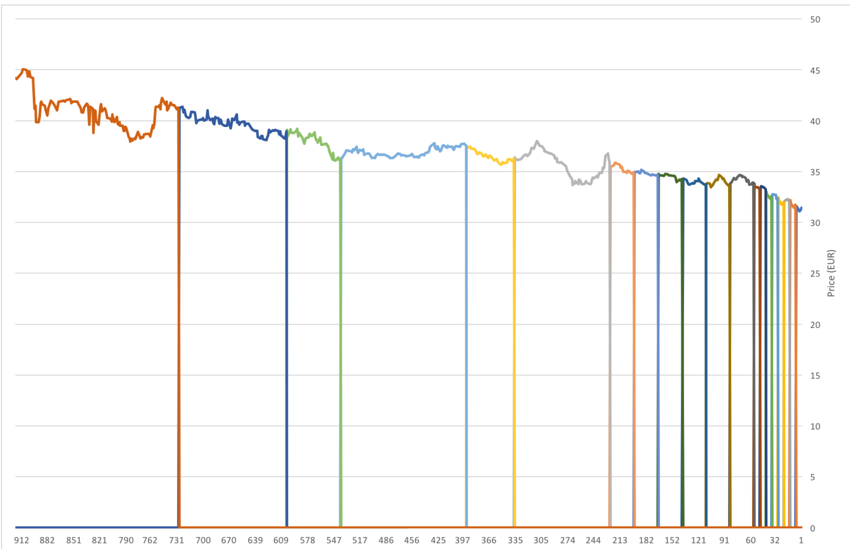


Figure 1.1: An example of how a contract can be divided into time buckets beginning when the contract starts being traded until the contract goes into delivery. The $y - axis$ is the price and the $x - axis$ is remaining days to delivery. Each colour corresponds to a time bucket.

This construction can be used for risk management in the following way. Let us say that on a market for a certain commodity there are several futures/forwards being traded. Let us say that there are several time buckets that cover the entire interval of delivery times for all contracts being traded. Then each day, we can observe absolute relative price changes

$$\left| \frac{P_t - P_{t-1}}{P_t} \right|$$

for each time bucket. If a time bucket contains several contracts, then for each day one takes the relative price change that is the highest in absolute value. Now one can for example use the empirical quantile $\hat{F}^{-1}(p)$ (or some more elegant method, such as Extreme Value Theory) to calculate margin M at level p for a contract belonging to time bucket k by using for example the formula

$$M = \hat{F}_k^{-1}(p) P_t V T$$

Here V is the volume of the contract, which, for a base load contract, is one MWh for 24 hours[2], and T is the duration of the contract which can be either one day, one week, one month, one quarter or one year, and finally P_t is the current price of the contract. An obvious problem with this approach is that the method is blind to the specific kind of contract. For example, one might calculate a margin for a yearly contract by using data belonging to a monthly contract.

Since the clearing house is the financial counter party for both the seller and the buyer of the contract, only the magnitude of the price change is of interest, not the direction. Then when observations have been collected for several days, a volatility curve is created by taking the 99 % quantile of the absolute relative price changes. The number of days for which observations have been collected is referred to as the look back period. This is typically a year.

The Figure 1.2 is an example of a volatility curve

This is by no means a complete explanation of time buckets and how they are used for risk management in the SPAN model. The reader interested in a full description of the SPAN model is directed to [1]

1.3 Some small illuminating examples

To get a more clear understanding of what it is we want to achieve, its potential role in risk management, and to explain why it is achieved by not by to just taking the well known linear correlation coefficient of two samples to model dependence and to provide an analytical expression for the ratio

$$\frac{F_{X+Y}^{-1}(p)}{F_X^{-1}(p) + F_Y^{-1}(p)}$$

[2]Contracts can also be, for example, specified for "peak load", i.e. office hours

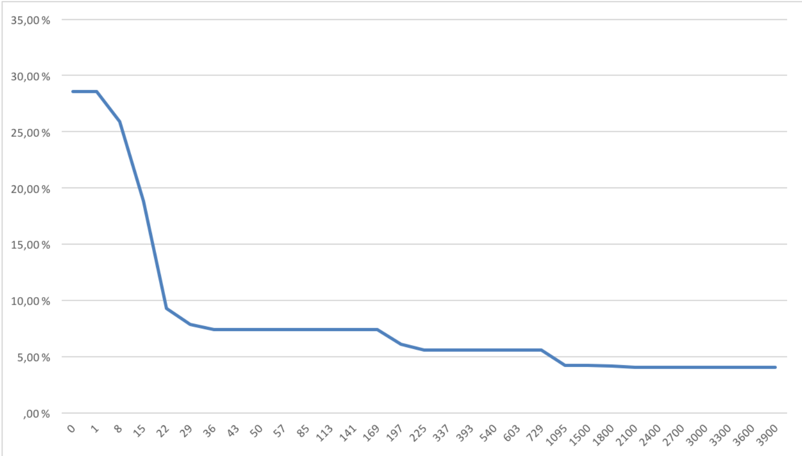


Figure 1.2: An example of a volatility curve, with days to delivery on the *x-axis*,s each number represents a time bucket

consider the following examples. On a market there are two different contracts available, A and B. In order to participate in the market, the clearing house requires that a participant can cover its losses from day to day in 99% of the cases. In order to guarantee this, the clearing house demands that collateral, margin, is deposited. Let us say, to make it a bit easier for us, that the profit or loss from each contract is given by normally distributed random variables X and Y. The clearing house has complete knowledge of the distribution of the gains and losses and thus requires that the amount of margin to be posted should equal the 99% quantile of the profits and losses multiplied by the size of the position. Let us first consider what would happen in two extreme cases, comonotonicity and countermonotonicity. We quickly remind our selves what we mean by this. We say that two random variables X, Y are comonotone if $X = F_X^{-1}(U)$ and $Y = F_Y^{-1}(U)$. Where $U \sim U(0,1)$ and F_X^{-1} and F_Y^{-1} are the quantiles of X and Y.
If X and Y are counter monotone then $X = F_X^{-1}(U)$ and $Y = F_Y^{-1}(1-U)$, with the same definitions as above. For $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\mu, \sigma^2)$, $X = Y$ in the case of comonotonicity and $X = -Y$ in the case of counter monotonicity. Comonotonicity is "the highest" correlation possible. Note that this does not mean that the linear correlation coefficient, ρ, is necessarily 1 or -1 in these two cases. A simple example is when $X \sim LN(0,1)$ and $Y \sim LN(\mu, \sigma^2)$. That is, $X = e^{Z_1}$ and $Y = e^{Z_2}$ for $Z_1 \sim N(0,1)$ and $Z_2 \sim N(\mu, \sigma^2)$. If one calculates the linear coefficient ρ

$$\rho = \frac{E[(X - E[X])(Y - E[Y])]}{\sqrt{E[(X - E[X])^2]}\sqrt{E[(Y - E[Y])^2]}} \tag{1.2}$$

for the random variables X and Y then one finds that for the case of monotonicity and counter monotonicity respectively,

$$\rho_{max} = \frac{e^{\sigma} - 1}{\sqrt{e-1}\sqrt{e^{\sigma^2} - 1}} \quad \text{and} \quad \rho_{min} = \frac{e^{-\sigma} - 1}{\sqrt{e-1}\sqrt{e^{\sigma^2} - 1}} \tag{1.3}$$

As can be readily seen, as $\sigma \to \infty$, $\lim_{\sigma\to\infty} \rho_{min} = \lim_{\sigma\to\infty} \rho_{max} = 0$. Even though, the pair X, Y are *perfectly* correlated. For a further discussion about this, see [17]. In these two (extreme) cases, either none, or a complete discount in margin could be offered, since for in the case of comonticity the two contracts move in the same direction or in the case of counter co monotonicity the contracts will move in opposite directions, so a loss in one contracts will always be countered by a gain of the same magnitude in the other.

Say that margin M is calculated as the quantile of the price change $P_t - P_{t-1}$ between one day $t-1$ and another t multiplied by the size of the position, that is $M = \pi F^{-1}_{P_t - P_{t-1}}(p)$, where π is the size of the position and $F^{-1}_{P_t - P_{t-1}}(p)$ is the quantile for the distribution of returns. Assume, a bit unrealistic, that the price change X_1, X_2 of two contracts are given by two normal distributions $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, with linear correlation coefficient ρ: $-1 \leq \rho \leq 1$. We may write for a bivariate normal distribution

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

Where Z_1, Z_2 I.I.D. $N(0,1)$. For the sum $Y = \alpha X_1 + \beta X_2$ it holds that $Y = \begin{pmatrix} \alpha & \beta \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ So that Y is $N(\alpha\mu_1 + \beta\mu_2, \alpha^2\sigma_1^2 + \beta^2\sigma_2 + 2\alpha\beta\rho\sigma_1\sigma_2)$ distributed see [9]. Then the quantile of Y, $F_Y^{-1}(p)$ is given by

$$F_Y^{-1}(p) = \alpha\mu_1 + \beta\mu_2 + \sqrt{\alpha^2\sigma_1^2 + \beta^2\sigma_2 + 2\alpha\beta\rho\sigma_1\sigma_2}\Phi^{-1}(p) \tag{1.4}$$

Where $\Phi^{-1}(p)$ is the standard quantile function for the $N(0,1)$ distribution.

So, in this scenario, the ratio in Equation 1.1 would be given by

$$\frac{\alpha\mu_1 + \beta\mu_2 + \sqrt{\alpha^2\sigma_1^2 + \beta^2\sigma_2^2 + 2\alpha\beta\rho\sigma_1\sigma_2}\Phi^{-1}(p)}{|\alpha|(\mu_1 + \sigma_1\Phi^{-1}(p)) + |\beta|(\mu_2 + \sigma_2\Phi^{-1}(p))} \tag{1.5}$$

Let us now return to margin. Say that the clearing house would like to offer a discount on margin in cases where it can be clearly established that there is a dependence in price movements between two contracts. If we continue to assume that the distribution of the price changes X_1, X_2 for the two contracts is normal, $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, we get that total margin for two contracts can be calculated (for a position of size α and β in the two contracts) as either

$$M = |\alpha| \left(\mu_1 + \sigma_2 \Phi^{-1}(p)\right) + |\beta| \left(\mu_1 + \sigma_2 \Phi^{-1}(p)\right) \tag{1.6}$$

or by using Equation 1.4. Now, a *discount* in margin γ that acknowledges eventual dependence between contracts can be defined as

$$\gamma = 1 - \frac{F^{-1}_{\alpha X_1 + \beta X_2}(p)}{|\alpha| F^{-1}_{X_1}(p) + |\beta| F^{-1}_{X_2}(p)} \tag{1.7}$$

If we use Equation 1.4 and Equation 1.6 the discount becomes

$$\gamma = \left(1 - \frac{\alpha\mu_1 + \beta\mu_2 + \sqrt{\alpha^2\sigma_1^2 + \beta^2\sigma_2^2 + 2\alpha\beta\rho\sigma_1\sigma_2}\Phi^{-1}(p)}{|\alpha|(\mu_1 + \sigma_1\Phi^{-1}(p)) + |\beta|(\mu_2 + \sigma_2\Phi^{-1}(p))}\right) \tag{1.8}$$

If we assume zero mean (for illustrative purposes) of the return, then this expression reduces to

$$\gamma = \left(1 - \frac{\sqrt{\alpha^2\sigma_1^2 + \beta^2\sigma_2^2 + 2\alpha\beta\rho\sigma_1\sigma_2}}{|\alpha|\sigma_1 + |\beta|\sigma_2}\right) \tag{1.9}$$

Notice that the discount is non-zero even if X_1 and X_2 are uncorrelated or have a linear correlation coefficient ρ that is $\rho = 1$ or $\rho = -1$[3]. If we further assume that $\sigma_1 = \sigma_2$ and $\alpha = \beta$ then γ in (1.9) further reduces to

$$\gamma = 1 - \frac{\sqrt{2\alpha^2\sigma^2 + 2\rho\alpha^2\sigma^2}}{2|\alpha|\sigma} \tag{1.10}$$

So that in the case $\rho = 0$ (independence), $\gamma = \frac{\sqrt{2}-1}{\sqrt{2}}$, and further the case $\rho = -1$ (countermonotonicity) $\gamma = 1$ and finally for the case $\rho = 1$ (comonotonicity) gives that $\gamma = 0$. Showing that a none or a complete reduction in margin is possible.

1.4 The data

In order to solve the problem, the following data is available.
The data set contains 59121 observations, of both futures and forwards. A typical data entry have the following fields given in Table 1.1.

Most entries are self explanatory. The entries that are of most interest is the relative increments and TIERS. The entry that needs explanation the

[3]Which, for the bivariate normal distribution implies comonotonicity and counter monotonicity, respectively. Moreover, $\rho = 0$ implies independence for the bivariate normal distribution.

Tier	Date 0	Date 1
DEXBQ003	2011-05-23	2011-05-24
Series	P_0	P_1
EDEBLQ4-11	65.15	65.2
Increment	Relative increment	Date of delivery start
0.05	0.0007674597	2011-10-01
Date of delivery stop	Days to delivery start	Days to delivery stop
2011-12-31	130	221

Table 1.1: Table of the different fields in a data point with illustrating values

most is the TIER. The tiers are keeping record on the sequence of delivery for a contract of a specific type. In the sample data given in Table 1.1, the tier is DEXBQ003. This is to be read as German electricity (DEX), base load (B), quarterly (Q) contract that has one quarterly contract being traded before it comes into delivery (003) (the contract that is in delivery has index 001). It follows this pattern for all the contract, with the obvious interpretation; DEXBY004, is a yearly contract with two yearly contracts that will go into delivery before it. The entry series follows a similar pattern. EDEBLQ4-11 is shorthand for Electricity in Germany (*Deustchland* EDE)) Base Load (all hours of the day, BL) for a quarter (Q), the forth of the year 2011.

Chapter 2

Mathematical background

The approach to capture the dependence will be based on the use of a combination of GARCH (Generalised Autoregressive Conditional Heteroskedasticity) processes and copulas.

2.1 GARCH processes

The ARCH (Autoregressive Conditional Heteroskedasticity) was first proposed by Engle in 1981 in [5]. The ARCH process is defined for the univariate case by

Definition 2.1 *A process $\{X_t\}$ is said to be an $ARCH(p)$ process if it for every $t \in \mathbb{Z}$ satisfies*

$$\begin{aligned} X_t =& \sqrt{h_t} Z_t \\ h_t =& \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 \\ & \omega > 0, \quad \alpha_j \geq 0 \end{aligned}$$

where $Z_t \sim WN(0,1)$ and h_t is a strictly positive process.

Here, $WN(0,1)$ is a white noise process (see Definition 2.5) with zero mean and variance/standard deviation one.
A generalisation of the above and one of the main focuses of this thesis, is the so called GARCH-process (Generalised Autoregressive conditional Heteroscedasticity), first proposed by Bollarslev in 1986 in [3]. It is for the univariate case defined as

Definition 2.2 *A process $\{X_t\}$ is said to be a $GARCH(p,q)$ process if it* *for every $t \in \mathbb{Z}$ satisfies*

$$X_t = \sqrt{h_t} Z_t \tag{2.1}$$

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} \tag{2.2}$$

$$\omega > 0, \quad \beta_j, \alpha_j \geq 0 \tag{2.3}$$

$$\tag{2.4}$$

where $Z_t \sim WN(0,1)$ and h_t is a strictly positive process.

The ARCH/GARCH model allows us to capture a few stylised facts about financial returns. Most prominently, volatility clustering. That is, the well known fact that changes in volatility tend to come in clusters. T As can be easily seen, the $ARCH(p)$ process is a special case of the $GARCH(p,q)$ process with $q = 0$.

2.1.1 Properties of GARCH processes

Before we start to work with the GARCH process, we need to establish some facts regarding the process. We first need to establish under which conditions the process is stationary.

2.1.1.1 Stationarity of GARCH processes

Generally for stochastic processes there are two kinds of stationarity, first order, and second order stationarity. We first give a quick definition of the two kinds of stationarity.

Definition 2.3 *A stochastic process $\{X_t\}$ is said to be strictly (first order) stationary if $\{X_{t_1}, \ldots . X_{t_n}\}$ and $\{X_{t_{1+h}}, \ldots, X_{t_{n+h}}\}$ have the same probability distribution for all $t_1, \ldots, t_n$ and $h > 0$.*

Definition 2.4 *A stochastic process $\{X_t\}$ is said to be weakly (second order) stationary stationary if $Cov(X_s, X_{s+t})$ is independent of s and thus a function of t and $E[X_s]$ and $E\left[X_s^2\right]$ are independent of s.*

With these definitions in mind, we can define a white noise.

Definition 2.5 *We say that a stochastic process $\{X_t\}$ is white noise $WN(0, \sigma^2)$ with centred mean 0 and variance σ^2 if it is weakly stationary (second order) and*

$$\rho(h) = Cov(X_t, X_{t+h}) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases} \tag{2.5}$$

Moreover, we say that $\rho(h)$ is the autocorrelation function for the process $\{X_t\}$.

What now follows might seem like a abundance of definitions and theorems. But we will need them to establish the conditions for two kinds of stationarity of the GARCH and to show that the GARCH process can be seen as a white noise.
We begin by writing a general squared $GARCH(p, q)$ process in vector form as

$$\bar{x}_t = b_t + A_t \bar{x}_{t-1} \tag{2.6}$$

Where A is the following $(p + q, p + q)$ matrix.

$$A_t = \begin{pmatrix} \alpha_1 Z_t^2 & & \cdots & & \alpha_q Z_t^2 & \beta_1 Z_t^2 & & \cdots & & \beta_p Z_t^2 \\ 1 & 0 & \cdots & & 0 & 0 & & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 & 0 & & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & 1 & 0 & 0 & & \cdots & 0 & 0 \\ \alpha_1 & & \cdots & & \alpha_q & \beta_1 & & \cdots & & \beta_p \\ 0 & & \cdots & & 0 & 1 & 0 & \cdots & & 0 \\ 0 & & \cdots & & 0 & 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & & \cdots & 0 & 0 & 0 & & \cdots & 1 & 0 \end{pmatrix} \tag{2.7}$$

and

$$\bar{b}_t = \begin{pmatrix} \omega Z_t^2 \\ 0 \\ \vdots \\ \omega \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{p+q} \qquad \bar{x}_t = \begin{pmatrix} X_t^2 \\ \vdots \\ X_{t-p+1}^2 \\ h_t \\ \vdots \\ h_{t-q+1} \end{pmatrix} \in \mathbb{R}^{p+q} \tag{2.8}$$

We call this the Markovian representation of the GARCH process. It is indeed a Markov process as the future value is dependent on the past value of x_{t-1} only. As an example, we can see that the $GARCH(1, 1)$ process can in this representation be written as

$$\begin{aligned} \bar{x}_t &= \bar{b}_t + A_t \bar{x}_t \\ &= \begin{pmatrix} \omega Z_t^2 \\ \omega \end{pmatrix} + \begin{pmatrix} \alpha_1 Z_t^2 & \beta_1 Z_t^2 \\ \alpha & \beta \end{pmatrix} \begin{pmatrix} X_{t-1}^2 \\ h_{t-1} \end{pmatrix} \\ &= \begin{pmatrix} (\omega + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1}) Z_t^2 \\ \omega + \alpha X_{t-1}^2 + \beta_1 h_{t-1} \end{pmatrix} \\ &= \begin{pmatrix} h_t Z_t^2 \\ h_t \end{pmatrix} \end{aligned}$$

We can by iterating (2.7), get that

$$\bar{x}_t = b_t + \sum_{k=1}^{\infty} A_t A_{t-1} A_{t-k+1} b_{t-k} \tag{2.9}$$

The idea now is to find conditions for the existence of this sum.
To continue, let us define a useful relation for the spectral radius ρ of matrices.

$$\lim_{t\to\infty} \frac{1}{t} \log ||A^t|| = \log(\rho(A)) \tag{2.10}$$

Here, the specific matrix norm is unimportant. It can be shown that the spectral radius is invariant with respect to the choice of norm. This result has the following extension to random matrices.

$$\gamma = \lim_{t\to\infty} a.s \frac{1}{t} \log ||A_t A_{t_1} \dots A_1|| \tag{2.11}$$

Here, *a.s.* stands for almost surely.

With this, we have the following theorem

Theorem 2.6 *A necessary and sufficient condition for the existence of a strictly stationary solution to the* $GARCH(p, q)$ *model is that*

$$\gamma < 0 \tag{2.12}$$

for the matrix A *in Equation (2.7) and* γ *as defined in Equation (2.11) When the strictly stationary solution exists, it is unique, nonanticipative and ergodic.*

A proof of the theorem can be found in [6].
We get the conditions for second order stationarity with the following theorem.

Theorem 2.7 *If there exists a GARCH(p,q) process, in the sense of Definition 2.2 which is second-order stationary and nonanticipative, and if* $\omega > 0$*, then*

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \tag{2.13}$$

Conversely, if condition (2.13) holds, the unique strictly stationary solution of the model in Definition 2.2 is a weak white noise (and thus is second-order stationary). In addition, there exists no other second-order stationary solution.

In the $GARCH(1, 1)$ case, the criterion for strict stationarity becomes simpler, by the following theorem.

Theorem 2.8 *If*

$$-\infty < \gamma := E\left[\log(\alpha Z_t^2 + \beta)\right] < 0 \tag{2.14}$$

then the infinite sum

$$h_t = \left\{1 + \sum_{i=1}^{\infty} \Gamma(Z_{t-1}) \dots \Gamma(Z_{t-i})\right\} \omega \tag{2.15}$$

where $\Gamma(y) = \alpha y^2 + \beta$*, converges almost surely (a.s.) and the process* X_t *defined by* $X_t = \sqrt{h_t} Z_t$ *is the unique strictly stationary solution of the model in Definition 2.2. This solution is nonanticipative and ergodic. If* $\gamma \geq 0$ *and* $\omega > 0$*, there exists no strictly stationary solution.*

We also have an easy condition for second order stationarity for the $GARCH(1,1)$ with the following theorem.

Theorem 2.9 *Let* $\omega > 0$ *. If* $\alpha + \beta \geq 1$*, a nonanticipative and second-order stationary solution to the GARCH(1, 1) model does not exist. If* $\alpha + \beta < 1$*, the process* X_t *defined by 2.2 is second-order stationary. More precisely,* X_t *is a weak white noise. Moreover, there exists no other second-order stationary and nonanticipative solution.*

2.1.2 ARMA-GARCH

As a further extension of the $GARCH(p,q)$ model, there is the $ARMA(r,s)-GARCH(p,q)$ model. The $ARMA(r,s) - GARCH(p,q)$ model is simply a regular ARMA model with GARCH errors.
An $ARMA(r,s)$ process is given by the definition below.

Definition 2.10 *The process* $\{X_t = 0, \pm1, \pm2, \dots\}$ *is said to be an* $ARMA(r,s)$ *process if* $\{X_t\}$ *is stationary and if for every* t*,*

$$X_t - \phi_1 X_{t-1} - \dots - \phi_r X_{t-r} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_s Z_{t-s}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$*.*
We say that $\{X_t\}$ *is an* $ARMA(r,s)$ *process with mean* μ *if* $\{X_t - \mu\}$ *is an* $ARMA(r,s)$ *process.*

Furthermore, in order to guarantee that an $ARMA(r,s)$ is invertible and casual, we require that the polynomials

$$1 - \phi_1 z - \dots - \phi_r z^r \tag{2.16}$$

$$1 + \theta_1 z + \dots + \theta_s z^s \tag{2.17}$$

have no common roots and that the roots lie outside the unit circle. For the reader inclined to know more about ARMA processes and to find proof of the above, a good source is [18].
We combine Definition 2.10 above with Definition (2.2) to get the $ARMA(r,s)-GARCH(p,q)$ process with mean μ

Definition 2.11 *We say that a process is an $ARMA(r,s) - GARCH(p,q)$ process if it satisfies*

$$X_t = \mu + \sqrt{h_t} Z_t \tag{2.18}$$

$$\mu_t = \mu + \sum_{i=1}^{r} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{s} \theta_j (X_{t-j} - \mu) \tag{2.19}$$

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i (X_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2.20}$$

$$Z_t \sim WN(0,1) \tag{2.21}$$

Where

$$\omega > 0 \quad \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1 \text{ and } \alpha_i, \beta_j \geq 0$$

Using the $ARMA(r,s) - GARCH(p,q)$ autoregressive and moving average effects in the data can be remedied.
By Theorem 2.8 for the $GARCH(1,1)$ case and by Theorem 2.7 for the general case we have the necessary conditions for a $GARCH(p,q)$-process to be a (weak) white noise.

2.1.3 Estimating the GARCH process

The GARCH process can be estimated using Maximum Likelihood Estimation. Maximum likelihood for the GARCH case means maximizing the expression

$$\mathcal{L}(\theta, x_1, \ldots, x_)) = \prod_{t=1}^{n} \frac{1}{\sqrt{h_t}} f\left(\frac{x_t}{\sqrt{h_t}}\right) \qquad h_t = \omega + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2 \tag{2.22}$$

where $f\left(\frac{x_t}{h_t}\right)$ is the density of the innovations (the white noise) and $x_1, \ldots, x_n$ are the observations. Here θ is a vector with the parameters we want to estimate $\theta = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \omega, \gamma_1, \ldots, \gamma_n)^T$. $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \omega$ are from definition 2.2 and $\gamma_1, \ldots, \gamma_n$ are parameters belonging to the distribution of innovations, e.g. the degree of freedom ν for a Student's t distribution.
When estimating, we look for parameters θ that maximize the likelihood. That is we search for

$$\arg\max_{\theta} \mathcal{L}(x_1, \ldots, x_n, \theta)$$

This is often done by instead using the log-likelihood, that is maximizing

$$\log(\mathcal{L}(x_1, \ldots, \omega)) = \sum_{t=1}^{n} \log\left(\frac{1}{\sqrt{h_t}} f\left(\frac{X_t}{\sqrt{h_t}}\right)\right) \tag{2.23}$$

The main difficulty in estimating the GARCH process, is how to make a good starting guess for h_t, since h_t is not directly observed. A common way to get a starting guess for $h_1, \ldots, h_q$ is to use the standard deviation for the entire sample, and then $h_{q+1}, \ldots, h_n$ are determined by the recursion.
The parameters are preferably found using numerical optimization techniques. To estimate the $ARMA(r,s) - GARCH(p,q)$ the maximum likelihood is again a viable approach. In this case, the expression to be maximized is

$$\mathcal{L}(x_1, \ldots, x_n, \theta,) = \prod_{t=1}^{n} \frac{1}{\sqrt{h_t}} f\left(\frac{X_t - \mu}{\sqrt{h_t}}\right) \tag{2.24}$$

Or the equivalent log likelihood version. X_t and h_t is in the above the same as in Definition 2.11 and f is the density of the innovations (the white noise). θ and $x_1, \ldots, x_n$ have here the same meaning as in Equation (2.22). Of course, in this case we also look for the parameters θ rather than the numerical value of objective function/likelihood.

2.2 Copulas

The dependence between the innovations Z_t of the GARCH for different contracts will be modelled using copulas. We will first given a formal definition of a copula and some useful properties.

2.2.1 Definition and properties

We begin by formally defining a two dimensional copula

Definition 2.12 *A two dimensional copula is a function C from $[0,1]^2$ to $[0,1]$ with the properties*

- *for every u, v in $[0,1]$:*
  - $C(0,v) = C(u,0) = 0$
  - $C(u,1) = u, \quad C(1,v) = v$
- *For every u_1, u_2, v_1, v_2 in $[0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$*

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

Copulas for dimensions higher than two can also be defined. However, this definition becomes very technical. Higher dimensional copulas are defined below. However, the definition of *grounded* and *n-increasing* are intentionally left out as to not make the text unnecessary technical. The interested reader is advised to consult [14]. Nevertheless, an n-dimensional copula can be defined as in 2.13

Definition 2.13 *An n-dimensional copula C is a function C with domain* $[0,1]^n$ *such that*

1. *C is grounded and n-increasing*
2. *C has margins* C_k $k = 1, \ldots, n$ *which satisfy* $C_k(u) = u$ *for all u in* $[0,1]$

The usefulness of copulas stems from Sklar's theorem.

Theorem 2.14 Sklar's theorem
Let H be an n-dimensional distribution function with marginal distributions $F_1, \ldots, F_n$. *Then there exists an n-copula C such that for all x in* $\boldsymbol{R}^n$

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \tag{2.25}$$

If $F_1, \ldots, F_n$ *are continuous, then C is unique. The converse is also true. A copula C with distributions* $F_1, \ldots, F_n$ *is a distribution function H.*

Sklar's theorem makes it possible for us to regard the marginals and the dependency structure separate from one another.
There is also another important theorem that tells us that transformed random variables have the same copula after transformation for a rather broad class of transformations.

Theorem 2.15 *Let* $(X_1, \ldots, X_n)^T$ *have copula C. If* $\alpha_1, \ldots, \alpha_n$ *are strictly increasing on* $RanX_1, \ldots, RanX_n$*, respectively,then* $(\alpha_1(X_1), \ldots, \alpha_n(X_n))^T$ *has copula C.*

A proof can be found in [14]

2.2.2 Some copulas

We will investigate if the dependence between the innovations can be modelled using five popular copulas; three *archimedian* copulas and two *elliptical* copulas.

Archimedian copulas

Archimedian copulas are defined by their *generator*, through the following theorem.

Theorem 2.16 *Let* φ *be a continuous strictly decreasing function from* $[0,1]$ *to* $[0,\infty)$*, such that* $\varphi(1) = 0$ *and let* $\varphi^{[-1]}$ *be the pseudo inverse of* φ *given by*

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

Then the function C from $[0,1]^2$ to $[0,1]$ given by

$$C(u,v) = \varphi^{[-1]}(\varphi(u)+\varphi(v))$$

is a copula if and only if φ is convex.

Using this theorem we can construct the following three popular archimedian copulas. The Clayton, Gumbel and Frank copulas.
The Clayton copula has generator

$$\varphi(t) = \frac{(t^{-\theta}-1)}{\theta} \tag{2.26}$$

which gives the Clayton copula

$$C_\theta^{Clayton}(u_1,u_2) = max[(u_1^{-\theta}+u_2^{-\theta}-1)^{-\frac{1}{\theta}},0] \tag{2.27}$$

The Gumbel copula has generator

$$\varphi(t) = (-\ln(t))^\theta$$

which gives the Gumbel copula.

$$C_\theta^{Gumbel}(u,v) = e^{-[(ln(u))^\theta+(-\ln(v))^\theta]^{\frac{1}{\theta}}}$$

Finally, the Frank copula has generator

$$\varphi(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right) \quad \theta\in\mathbb{R}\backslash\{0\}$$

Which gives the Frank copula

$$C_\theta^{Frank}(u,v) = \frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1}{e^{-\theta}-1}\right) \tag{2.28}$$

In order to construct archimedian copulas of dimension greater than two, we will use the following theorem.

Theorem 2.17 *Let φ be a continuous strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\varphi(0)=\infty$ and $\varphi(1)=0$ and let φ^{-1} denote the inverse of φ. Then, C^d, a function from $[0,1]^d$ to $[0,1]$, given by*

$$C^d(u_1,\dots,u_d) = \varphi^{[-1]}(\varphi(u_1)+\cdots+\varphi(u_d)) \tag{2.29}$$

is a copula for all $d\geq 2$ if and only if φ^{-1} is completely monotonic on $[0,\infty)$.

The condition of a completely monotonic inverse of the generator holds for the three archimedian copulas presented above. For these copulas it is thus possible to construct multivariate copulas using Theorem 2.17. For details see [14], The drawback of this approach however is that the dependency structure is controlled by just one parameter for all variables, which leads to inflexible models.

Elliptical copulas

It is possible to construct copulas using the well known elliptical distributions. This gives rise to so called *elliptical* copulas.
The bivariate Gaussian copula is given by

$$C_\rho^N(u,v) = \int_\infty^{\Phi^{-1}(u)} \int_\infty^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho}} e^{-\frac{s^2-2\rho st+t^2}{2(1-\rho^2)}} dsdt. \tag{2.30}$$

The bivariate t-copula is given by

$$C_{\nu,\rho}^t(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \left(1 + \frac{s^2-2\rho st+t^2}{\nu(1-\rho^2)}\right)^{-\frac{v+2}{2}} dsdt. \tag{2.31}$$

The parameters for the elliptical copulas (normal and Student's t) are for us familiar, they are simply the linear correlation coefficient and the degree of freedom. We will later see the importance of the degrees of freedom, the extra parameter for the Student's t copula.
The elliptical copulas can easily be extended to several variables. In fact, we have that the n-dimensional normal copula is given by

$$C_{\mathcal{P}}^N(u_1,\ldots,u_n) = \Phi_{\mathcal{P}}^n(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n)) \tag{2.32}$$

Where $\Phi_{\mathcal{P}}^n$ is the multivariate normal distribution function with correlation matrix $\mathcal{P}$, zero mean and unit variance, and Φ^{-1} is the quantile for the univariate standard normal distribution.
Likewise, the n-dimensional Student's t copula is given by

$$C_{\mathcal{P},\nu}^t(u_1,\ldots,u_n) = t_{\rho,\nu}^n(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_n)) \tag{2.33}$$

where $t_{\mathcal{P},\nu}^n$ is the multivariate t distribution function with correlation matrix $\mathcal{P}$ and ν degrees of freedom and t_ν^{-1} is the quantile for the univariate t distribution with ν degrees of freedom.

2.2.3 Estimation of copulas

Copulas, like the GARCH/GARCH-ARMA parameters, will be estimated using maximum likelihood. To be more specific, the copulas are estimated using pseudo maximum likelihood. The method gets its name from the fact that we use a pseudo uniform sample for the marginal distributions. We get pseudo uniform observations from a multivariate sample of size n by using a modified version of the empirical distribution function

$$F_{j,n}(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}(X_{i,j} < x) \quad j = 1,\ldots,d \tag{2.34}$$

for a multivariate sample of a random vector $(X_1, \ldots, X_j, \ldots, X_n)^T$ of size n. Here $\mathbf{1}$ is the indicator function and $\frac{1}{n+1}$, instead of the usual $\frac{1}{n}$, is to ensure that pseudo observations remain in the interior of $[0,1]^d$. This modification is done so that the the likelihood of the copula can be evaluated in each point.
Let the the copula have parameters given by the vector θ, then given U_i in $[0,1]$ the log likelihood of the copula is given by

$$\ln \mathcal{L}(\theta, \hat{U}_1, \ldots, \hat{U}_n) = \sum_{i=1}^{n} \ln c_\theta(\hat{U}_i) \tag{2.35}$$

where c_θ is the density of the copula, θ is a vector of the parameters of the copula and and $\hat{U}_i$ refers to the pseudo uniform sample obtained using equation (2.34).
Then one seeks the parameters θ that maximizes Equation 2.35, i.e. one seeks

$$\arg\max_{\theta} \ln L(\theta, \hat{U}_1, \ldots, \hat{U}_n)$$

this can be found by using numerical optimization techniques.

As in the GARCH case, the log likelihood is also a viable approach. To provide some examples of the objective function under consideration in the optimization the log likelihood for the normal copula and t copula is provided.
To estimate the parameters of the normal copula, we look for

$$\arg\max_{\mathcal{P}} \sum_{i=1}^{n} \ln f_{\mathcal{P}}(\Phi^{-1}(\hat{U}_{1,i}), \ldots, \Phi^{-1}(\hat{U}_{d,i})) - \sum_{i=1}^{n} \sum_{j=1}^{d} \ln(\phi(\Phi(\hat{U}_{i,j}))) \tag{2.36}$$

Where $f_{\mathcal{P}}$ is the density of the multivariate normal distribution with correlation matrix $\mathcal{P}$, unit variances and zero mean. Likewise, to estimate the parameters of the t copula, we look for

$$\arg\max_{\nu,\mathcal{P}} = \sum_{i=1}^{n} \ln f_{\nu,\mathcal{P}}(t_\nu^{-1}(\hat{U}_{i,1}), \ldots, t_\nu^{-1}(\hat{U}_{i,d})) - \sum_{i=1}^{n} \sum_{j=1}^{d} \ln f_\nu(t_\nu^{-1}(\hat{U}_{i,j})) \tag{2.37}$$

Where $f_{\nu,\mathcal{P}}$ is the density function for the multivariate student's t-distribution with correlation matrix $\mathcal{P}$ and ν degrees of freedom and zero mean. Lastly f_ν is the density of the univariate t-distribution with zero mean and unit variance and t_ν^{-1} is the corresponding quantile.

2.2.4 Goodness of fit for copulas

We want to test the hypothesis $H_0 : \mathcal{C} \in \mathcal{C}_0$, that our estimated copula belongs to a specific family of copulas. A review of goodness of fit tests

for copulas is given in [4]. The test that is to be presented relies on the copula, so we will use the pseudo observations. This means that the narrower hypothesis if the marginal distributions are also correctly specified is not covered under the test.
In this thesis, focus will be on a test that is based around the so called von-Mises statistic

$$S_n = \int_{[0,1]^d} n\{C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u})\}^2 dC_n(\mathbf{u}) \tag{2.38}$$

where C_n is the so called empirical copula

$$C_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(U_1 \leq u_1, \ldots, U_d \leq u_d) \tag{2.39}$$

Here $\mathbf{u}$ is a random number in $[0,1]^d$ and C_{θ_n} is the pseudo maximum likelihood estimate of the copula, estimated from a sample of size n. The standard method to obtain a p-value for this statistic when testing copulas is to use bootstrap. That is, draw with replacement from the sample and estimate the copula for the *new* (bootstrapped) sample and do this a large number of times. Bootstrapping is however a very computer intensive method, and the computational time for even a moderate sample can be hours, as the copula i re-estimated in every step. However an alternative to the bootstrap is outlined in [12], and is based on multipliers.
According to [12], the estimator is found to give satisfactory results for low sample sizes (around a 100 or so), and reduces the needed computational time to seconds.
The idea with the multiplier method introduced in [12] is to use an asymptotic representation of the pseudo maximum likelihood estimator as follows. Let $\hat{\mathbf{U}}_i$ be the pseudo $[0,1]^d$ distributed numbers given by Equation (2.34), the von Mises statistic in Equation (2.38) can be approximated by

$$S_n^{(k)} = \int_{[0,1]^d} \{\mathbb{C}_n^{(k)}(\mathbf{u}) - \dot{C}_{\theta_n}^T(\mathbf{u})\hat{\Theta}_n^{(k)}\}^2 dC_n(u) = \tag{2.40}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\{\mathbb{C}_n^{(k)}(\hat{\mathbf{U}}_i) - \dot{C}_{\theta_n}^T(\hat{\mathbf{U}}_i)\hat{\Theta}_n^{(k)}\}^2 \tag{2.41}$$

Here we have

$$\mathbb{C}_n^{(k)}(\mathbf{u}) = \alpha_n^{(k)}(\mathbf{u}) - \sum_{j=1}^{d} \frac{\partial C_\theta}{\partial u_k} \alpha_n^{(k)}(1, \ldots, 1, u_j, 1, \ldots, 1) \tag{2.42}$$

$$\alpha_n^{(k)}(\mathbf{u}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i^{(k)}\{\mathbf{1}(\hat{U}_i \leq \mathbf{u}) - C_n(\mathbf{u})\} \tag{2.43}$$

$$\hat{\Theta}_n^{(k)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i^{(k)} \hat{J}_{\theta_n}(\hat{U}_i) \tag{2.44}$$

where $Z_i^{(k)}$ are random numbers with zero mean and unit variance. Moreover

$$\hat{J}_{\theta_n}(\hat{U}_i) = \mathbf{\Sigma}_n^{-1} \left\{ \frac{\dot{c}_{\theta_n}(\hat{U}_i)}{c_{\theta_n}(\hat{U}_i)} - \frac{1}{n} \sum_{j=1}^{d} \sum_{k=1}^{n} \mathbf{1}(\hat{U}_{i,j} < \hat{U}_{k,j}) \frac{c_{\theta_n}^{(j)}(\hat{U}_k)}{c_{\theta_n}(\hat{U}_k)} \frac{\dot{c}_{\theta_n}(\hat{U}_k)}{c_{\theta_n}(\hat{U}_k)} \right\} \tag{2.45}$$

Here $\mathbf{\Sigma_n}$ is the covariance matrix for

$$\frac{\dot{c}_{\theta_n}(\hat{\mathbf{U}}_1)}{c_{\theta_n}(\hat{\mathbf{U}}_1)}, \ldots, \frac{\dot{c}_{\theta_n}(\hat{\mathbf{U}}_n)}{c_{\theta_n}(\hat{\mathbf{U}}_n)}$$

and

$$\dot{C}_\theta = \left(\frac{\partial C_\theta(u)}{\partial \theta_1}, \ldots, \frac{\partial C_\theta(u)}{\partial \theta_q} \right)^T \quad u \in [0,1]^d \tag{2.46}$$

and finally

$$\dot{c}_\theta = \left(\frac{\partial c_\theta(u)}{\partial \theta_1}, \ldots, \frac{\partial c_\theta(u)}{\partial \theta_q} \right)^T \quad u \in [0,1]^d \tag{2.47}$$

where $c_\theta(u)$ is the density of the copula, assuming it exists.
An approximate p-value for the test can then be obtained by following the steps below

1. Estimate the Cramer von Mises statistic

$$S_n = \int_{[0,1]^d} n\{C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u})\}^2 dC_n(\mathbf{u}) = \tag{2.48}$$

$$= \sum_{i=1}^{n} \left(C_n(\hat{\mathbf{U}}_i) - C_{\theta_n}(\hat{\mathbf{U}}_i) \right)^2 \tag{2.49}$$

2. For some large integer N do for $k \in \{1, 2 \ldots, N\}$

   (a) Generate n random variables $Z_j^{(k)}, j = 1, \ldots, n$ with $E\left[Z_j^{(k)}\right] = 0$ and
   $E\left[\left(Z_j^{(k)}\right)^2\right] = 1$

   (b) Form an approximate realization of S_n, $S_n^{(k)}$ using (2.41).

3. An approximate p-value is now given by

$$\frac{1}{N} \sum_{k=1}^{N} \mathbf{1}(S_n^{(k)} \geq S_n) \tag{2.50}$$

The test does not take into account the specified margins, but tests the dependence structure, provided by the copula, only.

2.2.5 Tail dependence

The copulas differ in what is known as tail dependence. For two $U(0,1)$ distributed random variables U, V with copula C, upper and lower tail dependence λ_U and λ_L is given by

$$\lambda_U = \lim_{z \to 1} 2P(U > z | V = z) \tag{2.51}$$

$$\lambda_L = \lim_{z \to 0} 2P(U \leq z | V = z) \tag{2.52}$$

In words, tail dependence measures in a sense, the probability that an extreme event in one variable, will be followed by an extreme event in the other variable. We may also refer to this as asymptotic dependence or asymptotic independence in the upper or lower tail. Strictly speaking, Equation (2.51) and (2.52) is only valid for exchangeable copulas. Copulas for which

$$C(u, v) = C(v, u).$$

All copulas we have presented have this property.
For the Gaussian copula given by Equation (2.30) we have that

$$\lambda_U = \lim_{z \to 1} 2P(U > z | V = z) = 2 \lim_{x \to \infty} P(\Phi^{-1}(U) | \Phi^{-1}(V) = x) = P(X > x | Y = x)$$

We know that for (X, Y) from a bivariate normal distribution with linear correlation coefficient ρ, $X | Y = x \sim N(\rho x, 1 - \rho^2)$. So that we get

$$2 \lim_{x \to \infty} P(X > x | Y = x) = 2 \lim_{x \to \infty} \bar{\Phi} \left(\frac{x - \rho x}{\sqrt{1 - \rho^2}} \right) = 2 \lim_{x \to \infty} \bar{\Phi} \left(x \sqrt{\frac{1 - \rho}{1 + \rho}} \right) = 0 \tag{2.53}$$

where $\bar{\Phi}(x) = 1 - \Phi(x)$ and $\Phi(x)$ is the normal distribution function. Since the normal copula is symmetric, the lower tail dependence λ_L is zero as well. For the t-copula similar calculations give that the tail dependence is given by

$$\lambda_U = 2\bar{t}_{\nu+1} \left(\sqrt{\frac{(\nu + 1)(1 - \rho)}{1 + \rho}} \right) \tag{2.54}$$

where $\bar{t}_{\nu+1}(x) = 1 - t_{\nu+1}(x)$ and $t_{\nu+1}$ is the distribution function for the Student's t distribution with ν degrees of freedom. As can be seen, the coefficient of tail dependence is decreasing in ν. and increasing in ρ. The asymptotic limit of the t copula as $\nu \to \infty$ is the normal copula, so the fact that the tail dependence is zero as $\nu \to \infty$ is expected. Since the t-copula is symmetric the upper tail dependence is the same as the lower tail dependence. For the tree Archimidean copulas under study, we find that, for an archimedian copula, the tail dependence is given by the following theorem from [16]. We start with a definition.

Definition 2.18 *We say that the generator φ of an archimedean copula is* strict *if $\varphi(0) = \infty$.*

With this technicality out of the way, we move on to the theorem.

Theorem 2.19 *Let φ be a strict generator such that φ^{-1} belongs to the class of Laplace transforms of strictly positive random variables. If $\varphi^{-1'}(0)$ is finite, then*

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$$

does not have upper tail dependence. If the copula has upper tail dependence, then $\varphi^{-1'}(0) = -\infty$ and the coefficient of upper tail dependence is given by

$$\lambda_U = 2 - 2 \lim_{s \to 0} [\frac{\varphi^{-1'}(2s)}{\varphi^{-1'}(s)}].$$

The coefficient of lower tail dependence is given by

$$\lambda_L = 2 \lim_{s \to \infty} [\frac{\varphi^{-1'}(2s)}{\varphi^{-1'}(s)}],$$

without the criteria of $\varphi^{-1'}(0)$.

Using this theorem, we can see that for the Clayton copula

$$\varphi(s)^{-1'} = \frac{1}{\theta}(1 + \theta s)^{\frac{1}{\theta} - 1}$$

so that the Clayton copula does not have upper tail dependence. The lower tail dependence is however given by

$$\lim_{s \to \infty} \frac{1 + \theta 2s^{-\frac{1}{\theta} - 1}}{1 + \theta s^{-\frac{1}{\theta} - 1}} = 2^{-\frac{1}{\theta}}.$$

For the Frank copula, tedious calculations give that

$$\varphi^{-1'} = -\frac{e^{\theta} - 1}{\theta}$$

which is finite, so the Frank copula does not have upper tail dependence. Moreover, since the Frank copula is symmetric, it does not have lower tail dependence either.
For the Gumbel copula we have that

$$\varphi(s)^{-1'} = -s\frac{1}{\theta}s^{-\frac{1}{\theta}}e^{-s\frac{1}{\theta}},$$

so that the Gumbel copula has upper tail dependence, which is given by

$$2 - 2 \lim_{s \to 0} \frac{\varphi(2s)^{-1'}}{\varphi(s)^{-1'}} = 2 - 2^{\frac{1}{\theta}}$$

and the lower lower tail dependence is zero.
Tail dependency allows us to model the fact that heavy losses in one contract are likely to come with a large loss in another contract.

2.2.6 Simulating from a Copula

Using a method for simulating from an elliptical distribution, simulating from the elliptical copulas can be achieved in the same way with a small modification. The steps for simulating from the elliptical copulas under consideration, with correlation matrix $\mathcal{P}$ and degree of freedom ν (in the case of the t copula) are given below.

- Simulate a random vector $\mathbf{X} = (x_1, \ldots, x_d)^T$ with d independent standard normal or student's t with ν degree of freedom random numbers.

- Multiply $\mathbf{X}$ with the Cholesky decomposition of the correlation matrix $\mathcal{P}$, denoted by $\mathbf{A}$ to obtain $(z_1, \ldots, z_d)^T = \mathbf{Z} = \mathbf{AX}$.

- Set $\mathbf{U} = (u_1, \ldots u_d)^T = \left(F^{-1}(z_1), \ldots, F^{-1}(z_d)\right)^T$ where F^{-1} is the quantile of the univariate normal/Student's t distribution with ν degrees of freedom.

- $\mathbf{U}$ can now be seen as a sample from the normal or student's t copula.

For the archimedian copulas there is a theorem that allows for simulation. However, this theorem is very technical and involves many definitions. For brevity reasons, we content ourselves with knowing that it exists and the knowledge that simulating from archimedian copulas is possible. For the theorem, please see [16].

2.3 Other considerations

2.3.1 Fat tails

Another *stylized fact* about the distribution of financial returns is that it usually demonstrates what is known as *fat tails*. Here, a small introduction to the concept will be provided
There is no absolute definition of a fat tails. A common [9] definition is to consider a left or right tail fat if

$$\lim_{x \to -\infty} \frac{F(x)}{e^{-\lambda(-x)}} = \infty \quad \lim_{x \to \infty} \frac{1 - F(x)}{e^{-\lambda x}} = \infty \quad \text{for all } \lambda > 0$$

for the left and right tail, respectively. Another way to view this is that the decay is slower than any exponential.
We investigate this by investigating how the left tail of the normal distribution behaves as $x \to -\infty$. We begin by investigating the limit of

$$\lim_{x \to -\infty} \frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{\phi\left(\frac{x-\mu}{\sigma}\right) \sigma/(-x)}.$$

An application of l'Hospital's rule gives that

$$\frac{d\Phi\left(\frac{x-\mu}{\sigma}\right)}{dx} = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$$

$$\frac{d}{dx}\frac{\sigma\phi\left(\frac{x-\mu}{\sigma}\right)}{(-x)} = \frac{\sigma}{x^2}\phi\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{(-x)}\phi'\left(\frac{x-\mu}{\sigma}\right) =$$

$$= \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\left(1 + \frac{\mu}{(-x)} + \frac{\sigma^2}{x^2}\right)$$

So that

$$\lim_{x\to-\infty} \frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{\phi\left(\frac{x-\mu}{\sigma}\right)\sigma/(-x)} = 1$$

By that we can conclude that the rate of decay of the normal distribution left tail is slower than any exponential. That is, we have that

$$\lim_{x\to-\infty} \frac{F_{\mu,\sigma}(x)}{e^{-\lambda(-x)}} = 0 \text{ for all } \lambda > 0$$

and that the left tail of the normal distribution is not heavy tailed. With similar calculations we can also show that the Student's t distribution is heavy tailed in this sense, i.e.,

$$\lim_{x\to-\infty} \frac{F_{\nu,\mu,\sigma}(x)}{e^{-\lambda(-x)}} = \infty \text{ for all } \lambda > 0.$$

The calculations are excluded for brevity.
Both the normal and Student's t distribution is symmetrical. This gives us that the right tail of the normal distribution is not heavy tailed and that the right tail of the Student's t distribution is heavy tailed. We can detect fat tails by using a QQ-plot. A QQ-plot (Quantile Quantile plot) is the empirical quantile (the ordered sample) plotted against the quantile of a reference distribution. The points of the reference distribution is usually chosen to be those of the empirical distribution function. If the sample have heavier tails than the reference distribution, the plot will resemble the letter *s*, as shown in Figure 2.1 where 500 simulated Student's t distributed random numbers with 3 degrees of freedom are plotted against the standard normal distribution as the reference distribution.

2.3.2 Kendall's tau

The parameters of the different copulas can be related to each other through the use of Kendall's tau (τ). Kendall's τ is given by the following definition

Definition 2.20 *Kendall's tau $\tau(X,Y)$ for a random vector (X,Y) is given by*

$$\tau(X,Y) = P\Big((X-\hat{X})-(Y-\hat{Y})>0\Big) - P\Big((X-\hat{X})-(Y-\hat{Y})<0\Big)$$

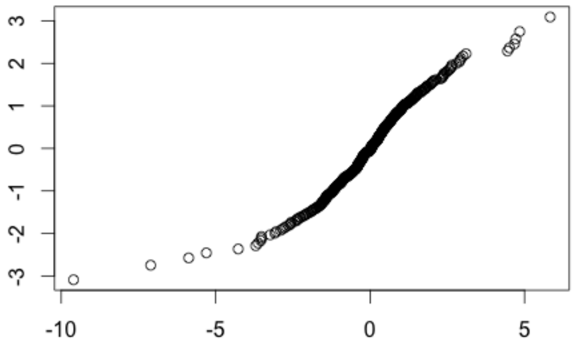


Figure 2.1: Example of QQ-plot with 500 simulated random numbers with a Student's t distribution with 3 degrees of freedom ($x - axis$) against a standard normal distribution ($y - axis$)

where $(\hat{X}, \hat{Y})$ *is an independent copy of* (X, Y).

In words, Kendall's τ is the probability of concordance minus the probability of disconcordance. For copulas, we have a theorem that allows us to calculate Kendall's τ for two random variables with copula C.

Theorem 2.21 *Let X and Y be random variables with an Archimedean copula C generated by* φ. *Then Kendall's* τ *for* X *and* Y *is given by*

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} du$$

A proof of this theorem can be found in [16]. A table of Kendall's τ as a function of of the copula parameter θ is given in Table 2.1 [1]

Copula	Kendall's tau
Gumbel	$1 - \frac{1}{\theta}$
Clayton	$\frac{\theta}{\theta+2}$
Frank	$1 - 4\frac{(1-D_1(\theta))}{\theta}$

Table 2.1: Kendall's τ as function of the copula parameter θ for the three arcimedian copulas Gumbel, Clayton and Frank copulas.

[1] $D_k(\theta)$ is the Debye function $D_k(\theta) = \frac{k}{\theta^k} \int_0^\theta \frac{x^k}{e^x - 1} dx$

We also have that for two random variables with an elliptical copula Kendall's τ is given by

$$\tau(X, Y) = \frac{2}{\pi} \arcsin(\rho_{X,Y})$$

Where $\rho_{X,Y}$ is the usual linear correlation coefficient between X and Y.

Chapter 3

Method

In order not to make the problem too large, we will focus only on yearly and quarterly forward contracts. Even with that restriction, all possible combinations will not be presented. The theory and considerations we have presented is combined into the following model. The corner stone of the method will be the so called *tiers*. The tiers will be in essence a bookkeeping tool for keeping track of how many contracts of the same kind that are being traded that goes in to delivery before it. For example, the tier DEXBQ003 refers to a quarterly contract, that has one quarterly contract that will go into delivery before this contract goes into delivery. Say for example that today is in the first quarter of the year. Then DEXBQ003 would correspond to the contracts that starts delivery in the third quarter of the year. The dividing into tiers enables us to have a consistent, or consistent enough, time series of prices or returns. The data divided into tiers do not have *gaps* in the price time series. By the absence of gaps, we mean that each business day will have recorded price data. This is important, as for the yearly contracts, there might be as long as a year between a yearly contract leaving a time bucket until a yearly contract returns to the bucket. The time series for the prices P_t is divided by P_{t-1} i.e. taking $Y_t = \frac{P_t}{P_{t-1}}$. We then take the logarithm of Y_t to create

$$X_t = \log(Y_t) \tag{3.1}$$

We refer to X_t in Equation (3.1) as the *log returns*. The two figures below (Figure 3.1 and Figure 3.2) shows the time series of returns for a tier with a time bucket colour marked.

What we try to achieve is that we want to find the nature of the dependence between (in this example) the blue and orange regions in the figures above, assuming it exists. We will try to do this with a combined GARCH and copula approach.

A key assumption made when estimating copulas is that the multivariate sample should be independent and identically distributed. Thus, if the sam-

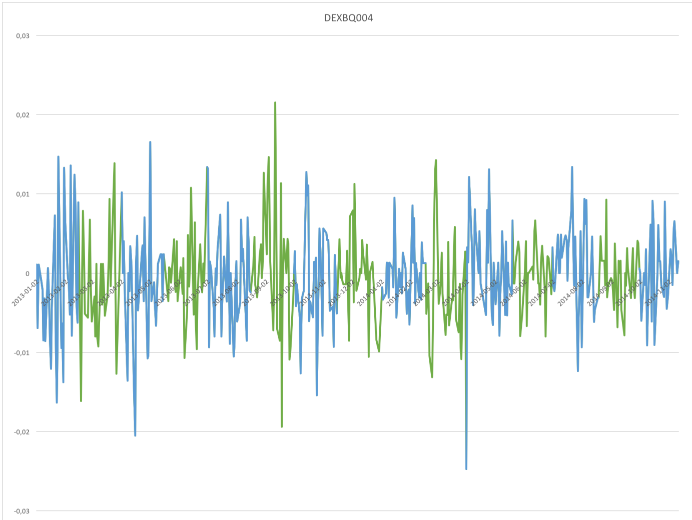


Figure 3.1: Times series of the log returns with the time bucket [225, 337) marked in blue

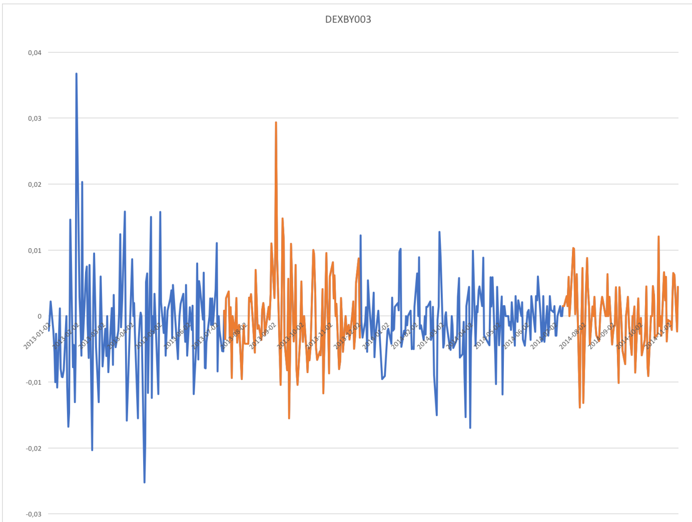


Figure 3.2: Times series of the log returns with the time bucket [393, 540) marked in orange

ple shows heteroskedecedicity then a direct approach using copulas on the data will fail, as the sample will have time dependent standard deviation. By filtering out the variance, using the estimated GARCH process we hope to achieve better estimation of the copulas.

The hypothesis is that the log returns can be modelled as some GARCH(p,q) or ARMA(r,s)-GARCH(p,q) process and then we can use the maximum likelihood estimator provided in the previous chapter to accurately get an estimation of the parameters of the innovations process Z_t, $Z_t = \frac{X_t}{h_t}$ from (2.2) or (2.11). From this estimation we are primarily interested in the innovations

$$Z_t = \frac{X_t}{h_t}.$$

From the estimated innovations Z_t from the GARCH process fitted to the time series of log returns on the tiers, we can estimate copulas with the methods presented in the previous chapter. We can then apply the test outlined in 2.2.4 to asses if we should reject some family of copulas at some confidence level α.
Using the copula(s) we could not reject, we simulate a large number of new random variables. These new random variables, in $[0, 1]^d$ are then transformed using the quantile transform to the distribution of the innovations Z_t of the proposed GARCH process.
Using the simulated innovations with hypothetically the same dependence structure as the original sample, the process is simulated one step into the future. We use these simulated random variables to form an estimate of

$$\frac{F^{-1}_{\sum_{i=1}^{d} X_i}(p)}{F^{-1}_{X_1}(p) + \cdots + F^{-1}_{X_d}(p)}$$

Since what we really are interested in is the price increments $P_{t+1} - P_t$, not the return $\frac{P_{t+1}}{P_t}$, the results will have to be transformed. The simulated log return data X_t can be transformed into price increments by the following transformation

$$P_{t-1}\left(e^{X_t} - 1\right) = P_{t-1}\left(\frac{P_t}{P_{t-1}} - 1\right) = P_{t-1}\left(\frac{P_t - P_{t-1}}{P_{t-1}}\right) = P_t - P_{t-1} \quad (3.2)$$

This is a strictly monotone transformation. By Theorem 2.15 the price increments will have the same copula and by that, the same dependence structure, as the log returns. A problem with the implementation of the copula goodness of fit test is that there is no algorithm, to the knowledge of the author, for the calculation of the distribution for the multivariate t-distribution with non integer degrees of freedom ν. So, as to overcome this numerical problem when testing the null hypothesis of a t-copula the degrees of freedom is rounded to the nearest integer that is highest in likelihood. This has the consequence that the degree of freedom ν is not a parameter under the test.

In short, we want to investigate whether the $GARCH$ or $ARMA - GARCH$ are a suitable models for the dynamics of the price differences. We also want to, using the copula, investigate the dependence between the time buckets to get an estimate of the ratio given in (1.1).

Chapter 4

Results

We will present results of the model for a few illustrative cases. First, we will look at a simple example consisting of two contracts (one time bucket each) and examine the dependence for an arbitrarily chosen time bucket. We will then examine how the ratio behaves when two contracts are far from each other in delivery and when they are close to each other in delivery. Lastly, we will investigate how the model behaves for two contracts belonging to time buckets far from each other.

4.1 A first example

As starting point, the tiers DEXBQ004 and DEXBY003 are chosen. With DEXBQ04 belonging to the time bucket [225, 337) and DEXBY003 belonging to the time bucket [393,540). The reason for this rather arbitrary choice is that the large amount of observations available for these two contracts in the two time buckets. We are looking at a quarterly contract, that is the third in line to go into delivery and a yearly contract with one yearly contract which will go into delivery before it. The time series of the log returns for the two tiers DEXBY003 and DEXBQ003 are given in Figure 4.1.

A scatter plot for the data that belong to the time bucket $[225, 339)$ for the tier DEXBQ004 and time bucket $[393, 540)$ for tier DEXY003 is given in Figure 4.2

In total, we have 2189 observations for both the tiers DEXBQ004 and DEXBY003. The subset of the samples for which the tier DEXBY003 belongs to the time bucket $[393, 540)$ and DEXBQ004 belongs to the time bucket $[225, 337)$ simultaneously consists of 532 observations. It it is this subset that will later be used to fit the copula. Plots of the estimated autocorrelation function for the observations and the squared observations is given in Figure 4.3 and 4.4

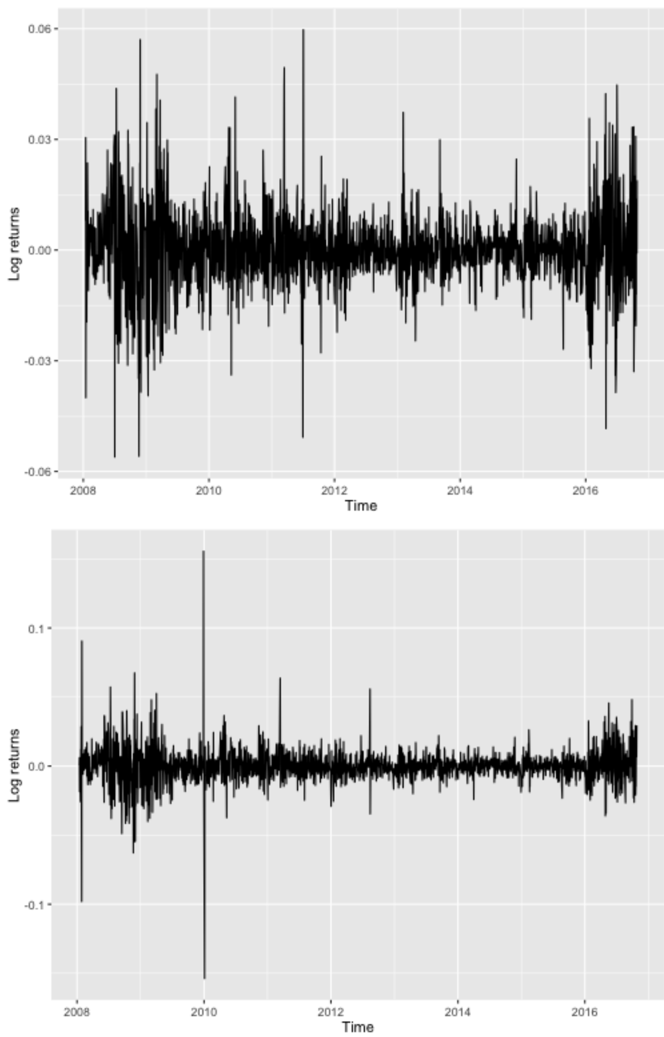


Figure 4.1: Time series of the log returns for the tier DEXBY003 (top) and DEXBQ004 (bottom).

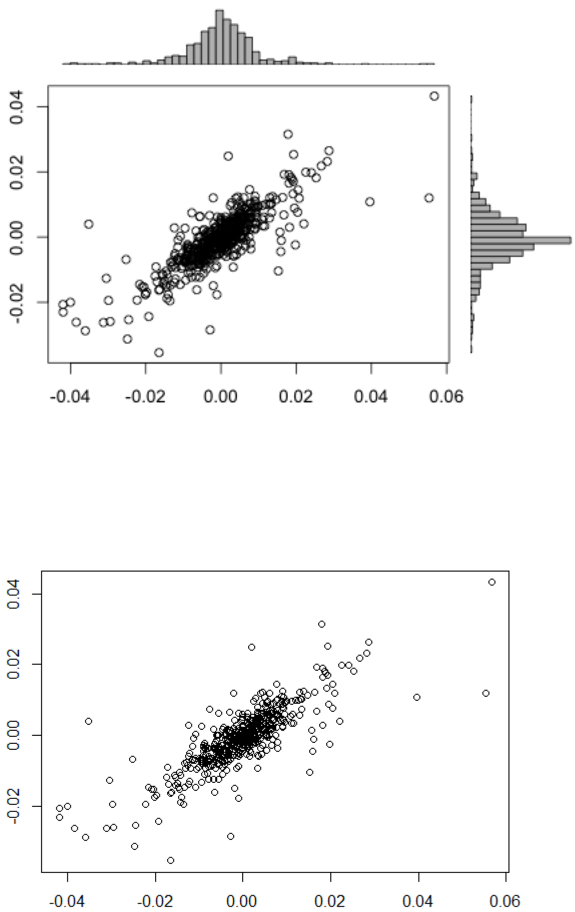


Figure 4.2: Top: Scatter plot for the data. The data belonging to the tier DEXBQ004 on the $y - axis$ and data belonging to the tier DEXBY003 on the $x - axis$. On the axes there is also a histogram. Bottom: Same as above with only the data that belong to time bucket $[225, 337)$ and time bucket $[393, 540)$ (minus the histograms)

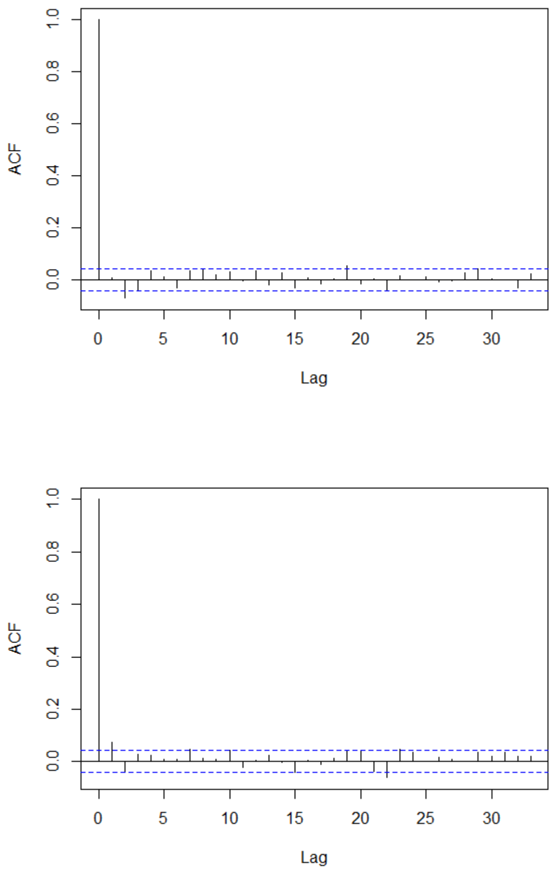


Figure 4.3: Plot of the estimated auto correlation function for the observations X_t for the tier DEXBQ004 (top) and DEXBY003 (bottom). The banded blue line is a 95% confidence interval.

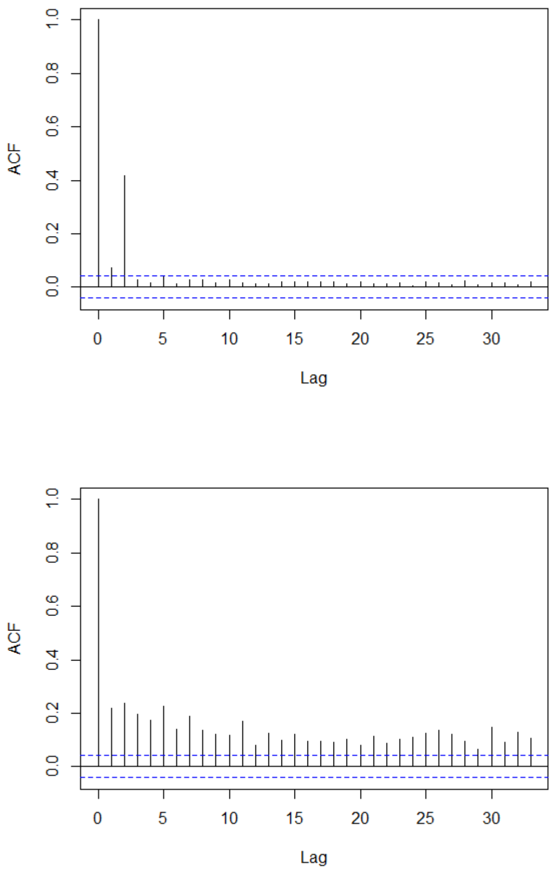


Figure 4.4: Plot of the estimated auto correlation function for the squared observations X_t^2 for the tier DEXBQ004 (top) and DEXBY003 (bottom). The banded blue line is a 95% confidence interval.

4.1.1 GARCH(1,1) with normally distributed innovations

The first approach is to use the normal distribution for the innovations of the GARCH process. Fitting a GARCH(1,1) with normal distributed innovations process using maximum likelihood gives the parameters given in Table 4.1

Tier	DEXBQ005	DEXBY004
μ	-5.396$\cdot 10^{-4}$	-5.249$\cdot 10^{-4}$
ω	2.137$\cdot 10^{-6}$	2.0259$\cdot 10^{-6}$
α_1	1.250$\cdot 10^{-1}$	1.688$\cdot 10^{-1}$
β_1	8.7340$\cdot 10^{-1}$	8.211$\cdot 10^{-1}$

Table 4.1: Parameters for the GARCH(1,1) series with normal distributed innovations fitted to the tiers DEXBQ004 and DEXBY003

Using this fitted model, we can extract the estimated time conditional standard deviation given in Figure 4.5. The autocorrelation function for the estimated innovations Z_t are given in Figure 4.1.1-

The squared innovations have the estimated autocorrelation functions given in Figure 4.7.

QQ-plots of the proposed normal distribution against the sample is given in Figure 4.8

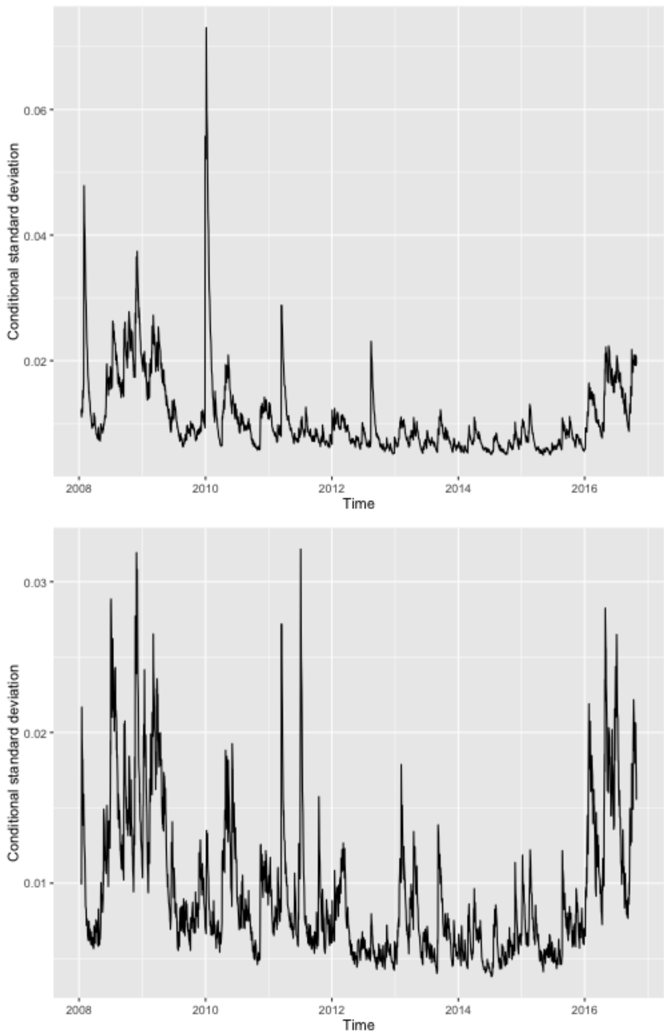
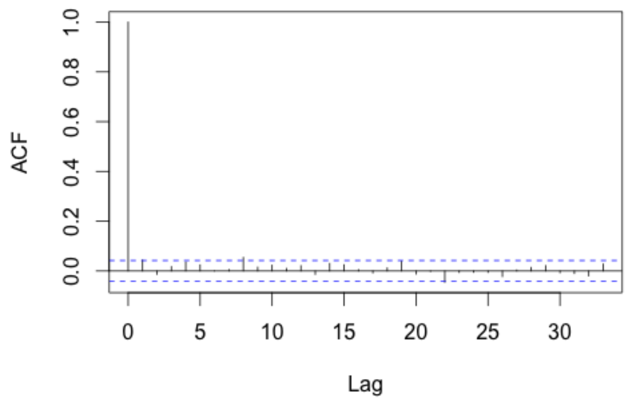


Figure 4.5: Estimated conditional standard deviation for the tier DEXBQ004 (top) DEXBY003 (bottom) with normally distributed innovations



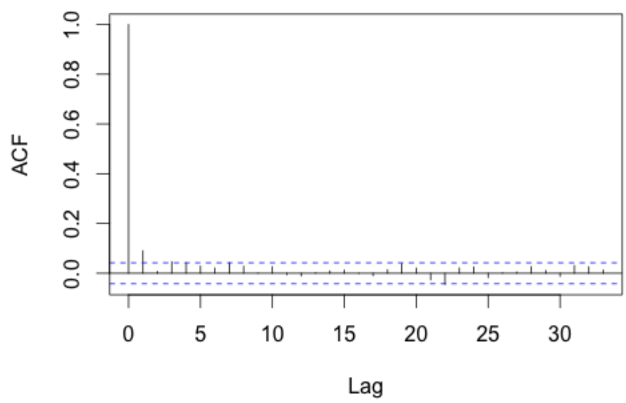
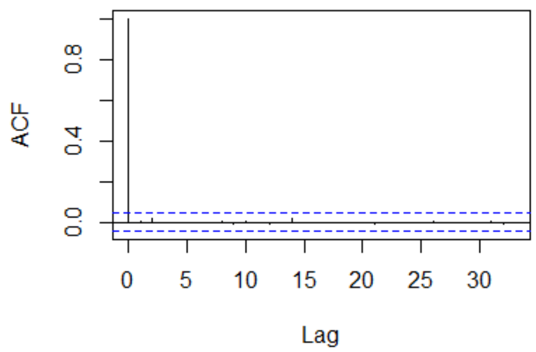


Figure 4.6: Plot of the estimated auto correlation function for the innovations Z_t from the estimated $GARCH(1,1)$ process with normal innovation for the tier DEXBQ004 (top) and DEXBY003 (bottom). The banded blue line is a 95% confidence interval.



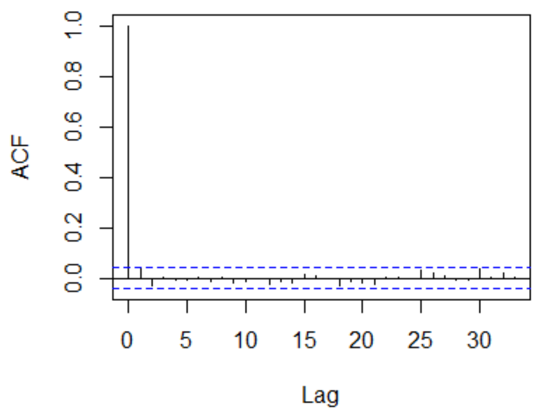


Figure 4.7: Plot of the estimated auto correlation function for the squared innovations Z_t^2 from the estimated $GARCH(1,1)$ process with normal innovation for the tier DEXBQ004 (top) and DEXBY003 (bottom). The banded blue line is a 95% confidence interval.

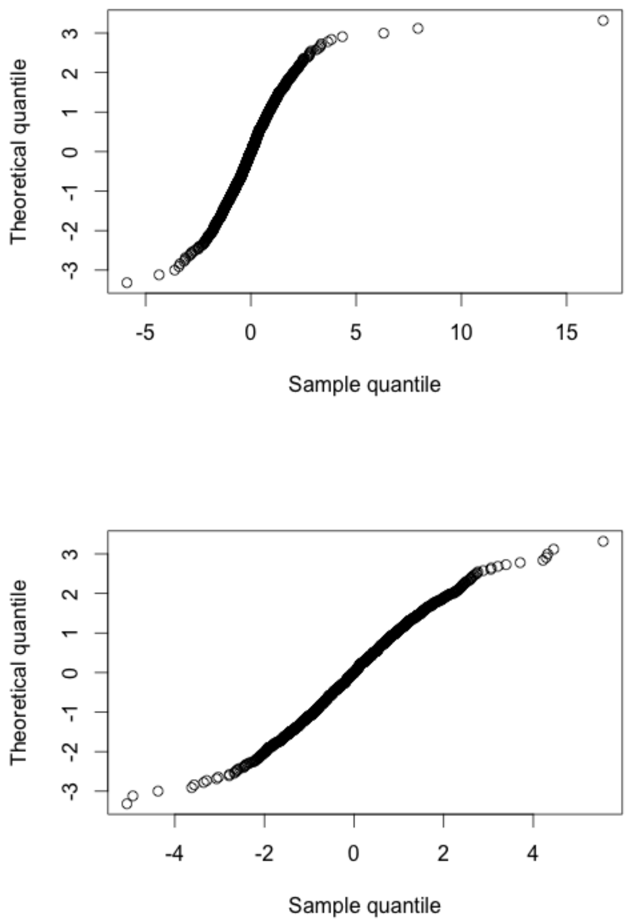


Figure 4.8: QQ plots of the sample quantile ($x - axis$) against the theorized normal distribution ($y - axis$) for DEXBQ004 (top) and DEXBY003 (bottom)

4.1.2 ARMA(1,1)-GARCH(1,1) with Student's t distributed innovation

As the GARCH(1,1) model with normally distributed innovations doesn't seem to capture the heavy tails of the sample, and since there still seems to be some autocorrelation in the estimated innovations of the sample we instead try an ARMA(1,1)-GARCH(1,1) model with Student's t distributed innovations. The parameters are for the fitted ARMA(1,1)-GARCH(1,1) with Student's t distributed innovations using maximum likelihood are given in Table 4.2.

Tier	DEXBY003	DEXBQ004
μ	-5.863·10^{-4}	-5.221·10^{-4}
ϕ_1	-3.900·10^{-4}	-2.989·10^{-4}
θ_1	4.437·10^{-1}	3.259·10^{-1}
ω	1.470·10^{-6}	2.324·10^{-6}
α_1	1.403·10^{-1}	1.439·10^{-1}
β_1	8.524·10^{-1}	8.477·10^{-1}
ν	6.921	4.822

Table 4.2: Coefficients for the $ARMA(1,1) - GARCH(1,1)$ with Student's t distributed innovations fitted to tiers at the top

From the fitted ARMA(1,1)-GARCH(1,1) we get the estimated conditional standard deviations in Figure 4.9.

The estimated autocorrelation function for the estimated innovations Z_t are given in Figure 4.10.

Plots of the estimated autocorrelation function of the squared innovations Z_t^2 is given in Figure 4.11.

QQ-plots for the proposed Student's t distributions against the sample is given in Figure 4.12.

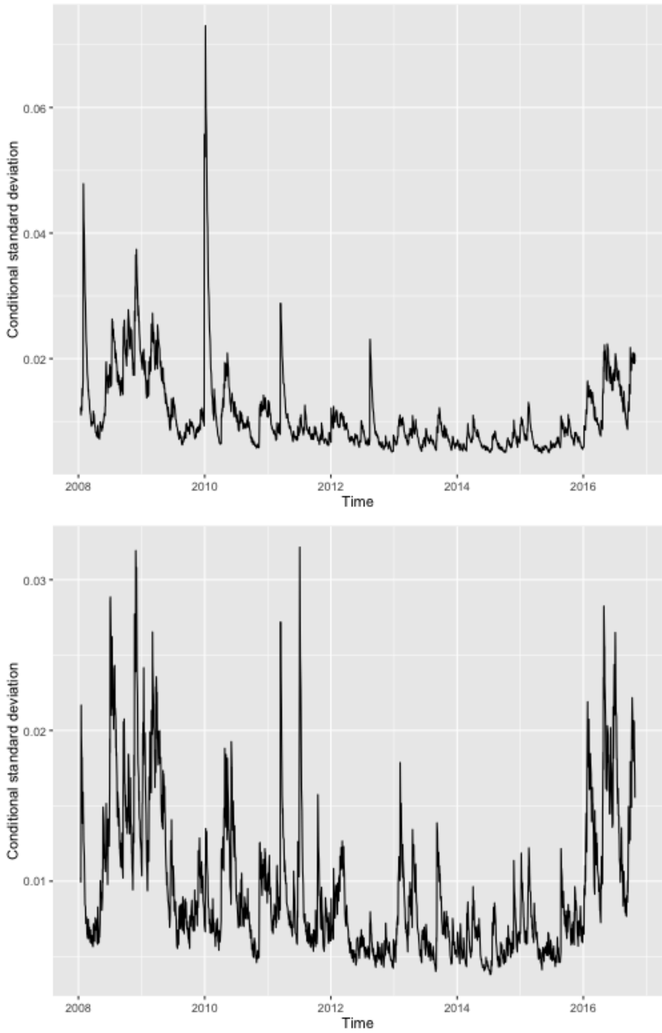


Figure 4.9: Estimated conditional standard deviation for the tier DEXBQ004 (top) and DEXBY003 (top).

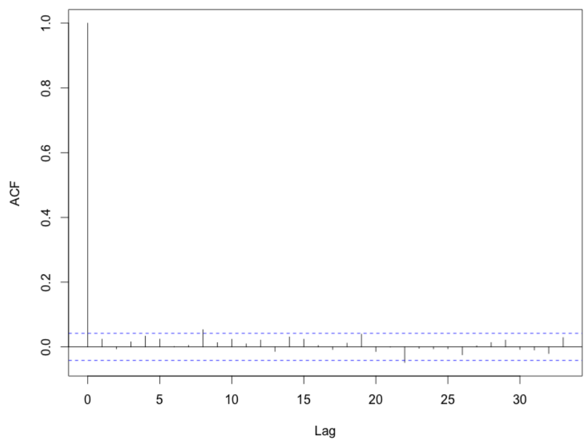
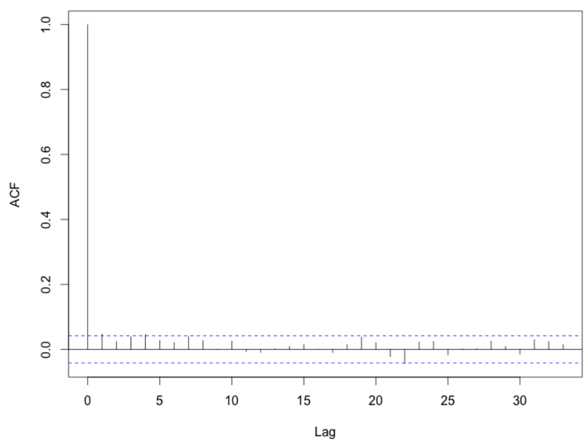
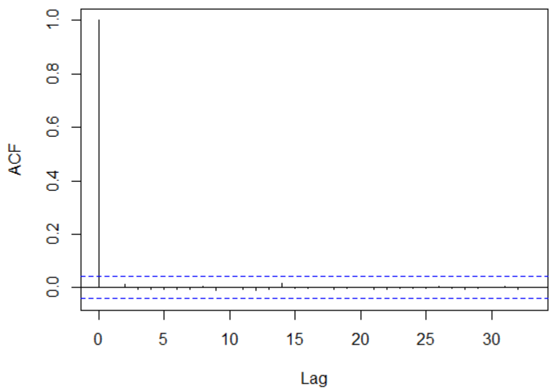


Figure 4.10: Plot of the estimated auto correlation function for the innovations Z_t from the estimated $ARMA(1,1) - GARCH(1,1)$ process with Student's t distributed innovations for the tier DEXBQ004 (top) and DEXBY003 (bottom). The banded blue line is a 95% confidence interval.





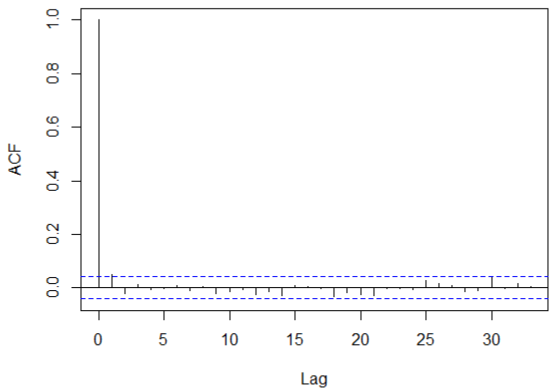


Figure 4.11: Plot of the estimated auto correlation function for the squared innovations Z_t^2 from the estimated $ARMA(1,1) - GARCH(1,1)$ process with Student's t distributed innovations for the tier DEXBQ004 (top) and DEXBY003 (bottom). The banded blue line is a 95% confidence interval.

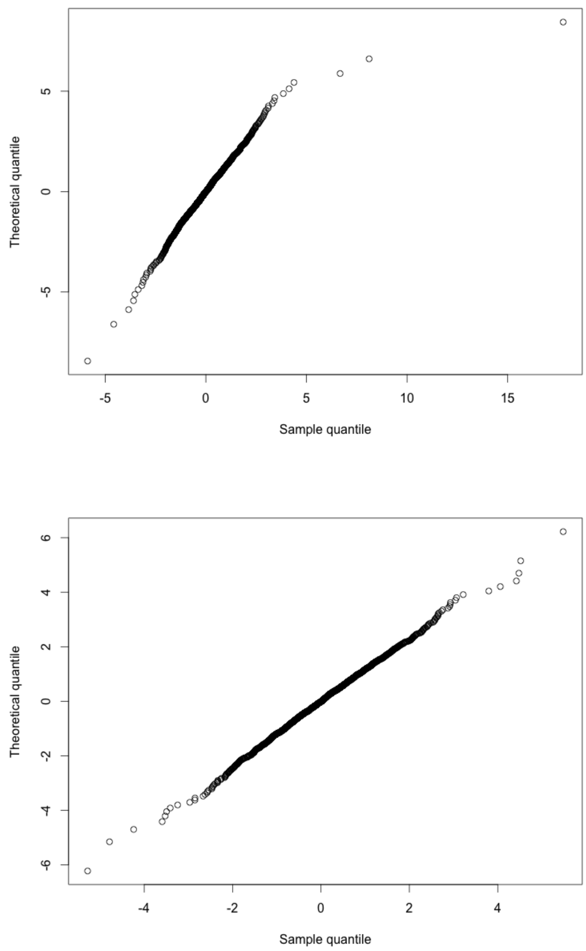


Figure 4.12: QQ plots of the sample quantile ($x - axis$) against the theorized Student's t distribution ($y - axis$) with degrees of freedom as given in Table 4.2 for DEXBQ004 (top) and DEXBY003 (bottom)

4.1.3 Fitting copulas to the innovations

The results obtained using a $ARMA(1,1) - GARCH(1,1)$ and Student's t distributed are considered sufficient for continuing to fitting copulas. We fit the copula to data given in the scatter plot in Figure 4.13. The data corresponds to the pair of innovations Z_t belonging to the time bucket $[225, 339)$ and $[393, 540)$.

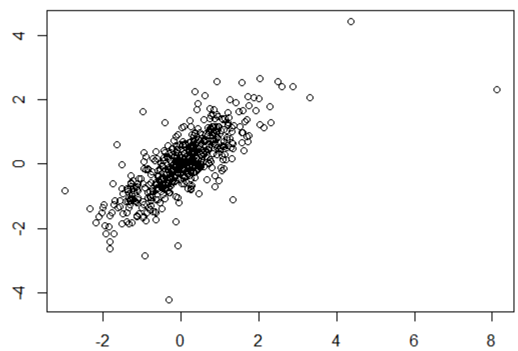


Figure 4.13: A scatter plot of the innovations Z_t from the tiers DEXBQ003 and DEXBY004 belonging to time bucket $[225, 339)$ and $[393, 540)$.

Five different copulas are fitted using maximum pseudo likelihood, the five copula that was introduced in chapter 3. The resulting estimated parameters together with result from the goodness of fit test from Chapter 2.2.4 and AIC values are given in Table 4.3.

Copula	Parameter(s)	p-value	AIC
Normal	$\rho = 0.7974$	0.0524	-527.0017
t	$\rho = 0.8073$ $\nu = 6.2551$	0.1723	-549.5125
Clayton	$\theta = 1.848$	0.0004995	-405.4068
Frank	$\theta = 8.137$	0.003497	-526.4015
Gumbel	$\theta = 1.848$	0.0004995	-518.2817

Table 4.3: Parameters for the five different copulas fitted to innovations from the $ARMA(1,1) - GARCH(1,1)$ from section 4.1.2. The p-value refers to the p-value obtained from the goodness of fit test presented in chapter 2 and AIC is Aikake's Information Criterion

By the rule we established earlier, that we should use a 5 % confidence, we can reject the hypothesis that the dependence is captured by the Clayton, Frank or Gumbel copula. We fail to reject the hypothesis that the copula is

Normal or Student's t copula. A plot of the copula density, with the pseudo observations superimposed is given in Figure 4.14.

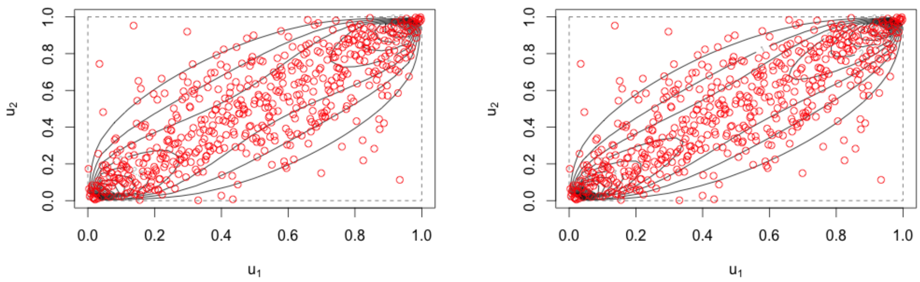


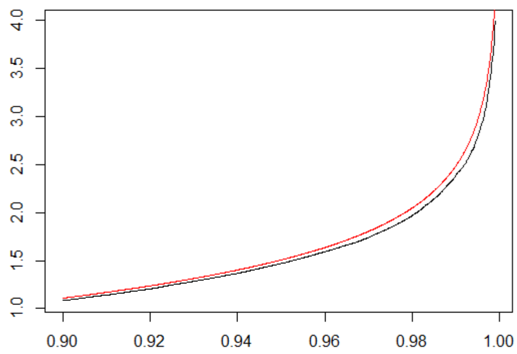
Figure 4.14: Copula densities for the estimated copulas that could not be rejected at the 5% level with parameters from Table 4.3 with the pseudo observations superimposed (in red). Left: Normal copula, right t copula

4.1.4 Estimating the ratio

With the established dynamics of the log returns and dependence from above, we can move on to the main question, the ratio

$$\frac{F_{X+Y}(p)}{F_X(p) + F_Y(p)}$$

Figure 4.15 show the quantile for $F^{-1}_{X+Y}(p)$ and the sum of the two quantiles $F^{-1}_X(p) + F^{-1}_Y(p)$ with the dependence according to the normal copula and t copula.



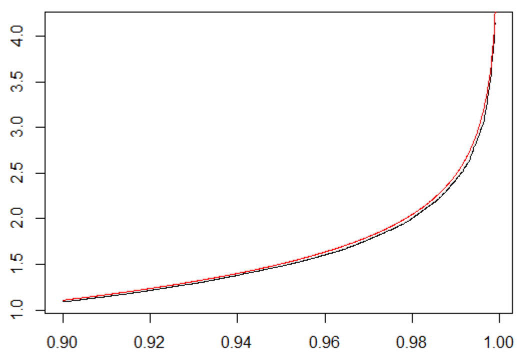
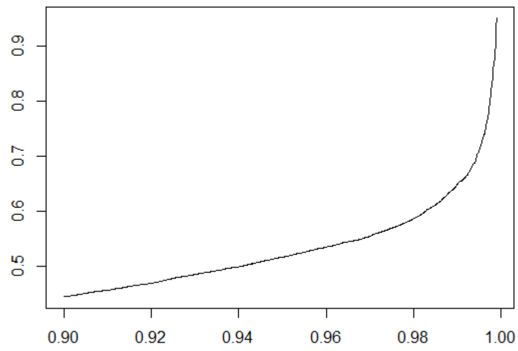


Figure 4.15: Plot of the quantile $F^{-1}_{X+Y}(p)$ (black) and the sum of the two quantiles $F^{-1}_X(p) + F^{-1}_Y(p)$ (red) with $p \in [0.9, 1)$ on the $x - axis$ for dependence given by the normal copula (top) and the t-copula (bottom).

In Figure 4.16 the ratio $\frac{F_{X+Y}(p)}{F_X(p)+F_Y(p)}$ is given as a function of p for the normal and t copula.



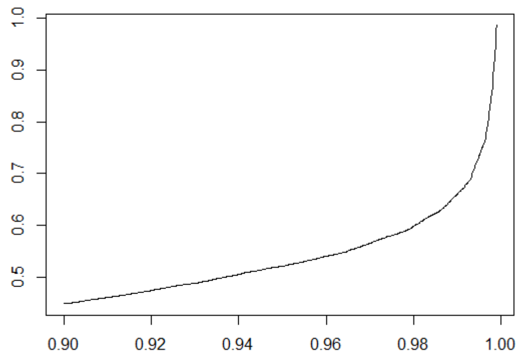


Figure 4.16: The ratio $\frac{F_{X+Y}^{-1}(p)}{F_X^{-1}+F_Y^{-1}}$ as function of $p \in [0.9, 1)$ (on the $x - axis$) with the dependence structure given by the normal copula (top) and t copula (bottom)

Table 4.4 gives a summary of the result at the 99% level.

Copula	$\frac{F_{X+Y}^{-1}(0.99)}{F_X^{-1}(0.99)+F_Y^{-1}(0.99)}$	$F_{X+Y}^{-1}(0.99)$	$F_X^{-1}(0.99) + F_Y^{-1}(0.99)$
Normal	0.961	2.390	2.488
t	0.968	2.407	2.488

Table 4.4: A summary of the different quantiles at the 99% level for the different copulas that could not be rejected.

4.2 Time buckets far from each other

Let us again look at quarterly and yearly contracts. This time, let the yearly contract belong to the time bucket $[729, 1095)$ and the quarterly contracts belong to the time bucket $[57, 85)$. The time series of the log returns is given in Figure 4.17.

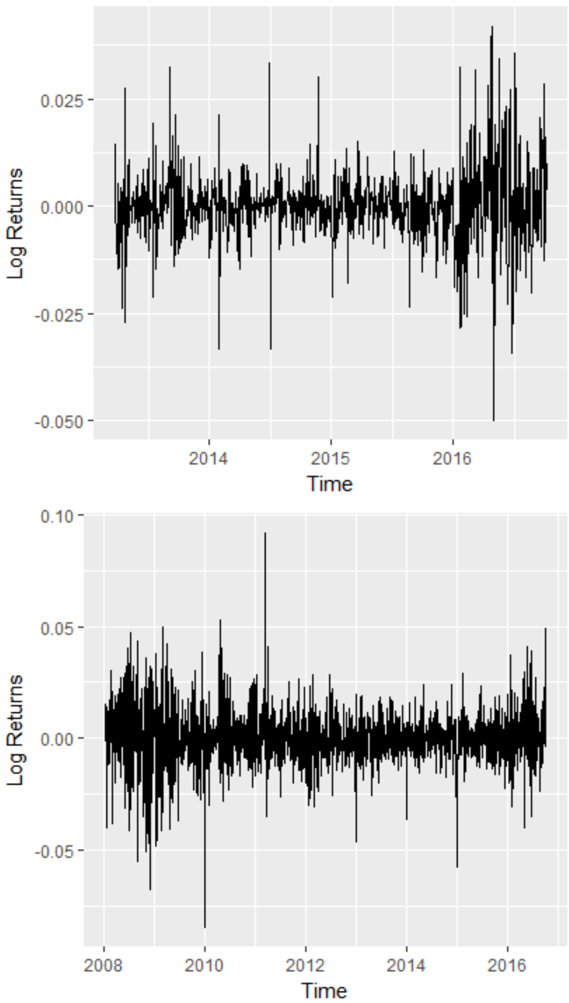


Figure 4.17: Time series of the log returns for the tier DEXBY004 (top) and DEXBQ002 (bottom)

The tier DEXBQ002 and DEXBQ004 contains 866 combined observations of which 273 belongs to the time bucket t $[57, 85)$ for the tier DEXBQ002 and time bucket $[729, 1095)$ for tier DEBXY004 simultaneously. A scatter plot of the data that belong to the tier DEXBQ002 and tier DEXBY004 as well as a scatter plot for the data that belong to time bucket $[57, 85)$ for the tier DEXBQ002 and time bucket $[729, 1095)$ for tier DEBXY004 is given in Figure 4.18.

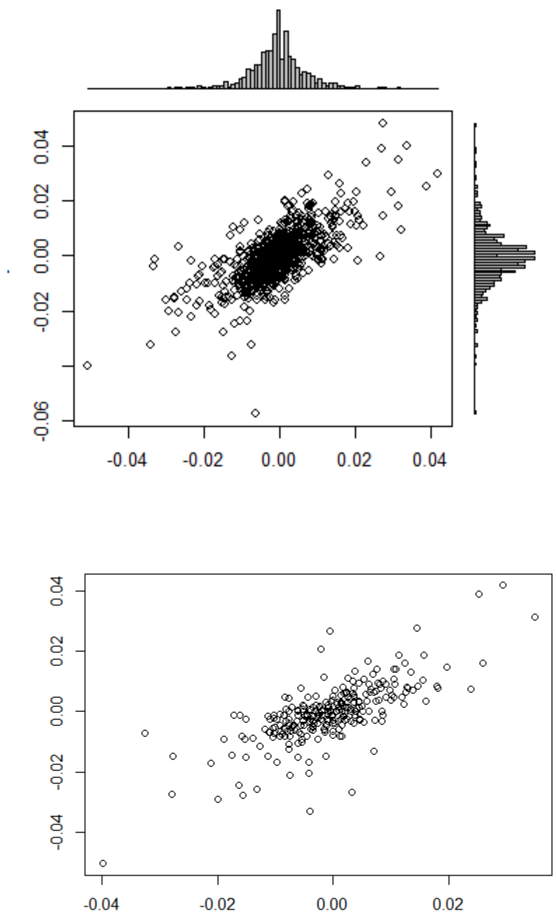


Figure 4.18: Top: Scatter plot for the data. The data belonging to the tier DEXBQ002 on the $y-axis$ and data belonging to the tier DEXBY003 on the $x-axis$. On the axes there is also a histogram. Bottom: Same as above with only the data that belong to time bucket $[57, 85)$ and time bucket $[729, 1095)$ (minus the histograms)

Assuming the same ARMA(1,1)-GARCH(1,1) with Student's t distributed innovations as in Section 4.1 we get the parameters given in Table 4.5 as estimated by the maximum likelihood method.

From the estimated process, we can produce the plots of the time conditional standard deviation given in Figure 4.19.

That the ARMA(1,1)-GARCH(1,1) with Student's t distibuted innovations is a suitable choice of model is supported by the below plots of the

TIER	DEXBY004	DEXBQ002
μ	$-5.786 \cdot 10^{-4}$	$-8.679 \cdot 10^{-4}$
ϕ	$-3.174 \cdot 10^{-1}$	$3.399 \cdot 10^{-2}$
θ	$3.445 \cdot 10^{-1}$	$5.586 \cdot 10^{-2}$
ω	$3.330 \cdot 10^{-6}$	$3.380 \cdot 10^{-6}$
α_1	$2.107 \cdot 10^{-1}$	$1.093 \cdot 10^{-1}$
β_1	$7.776 \cdot 10^{-1}$	$8.731 \cdot 10^{-1}$
ν	4.045	4.807

Table 4.5: Parameters for the ARMA(1,1)-GARCH(1,1) processes fitted to the tiers DEXBY004 and DEXBQ002 with Student's t distributed innovations.

autocorrelation function for the observations (Figure 4.20), squared observations (Figure 4.21), (estimated) innovations (Figure 4.22) and squared (estimated) innovations (Figure 4.23) and QQ-plots (Figure 4.24).

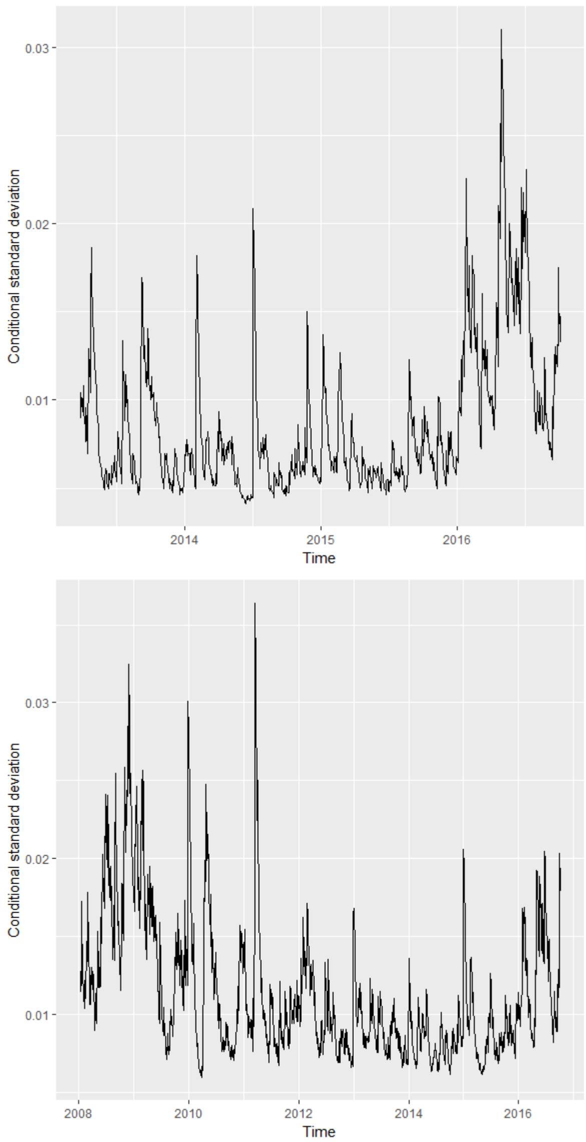
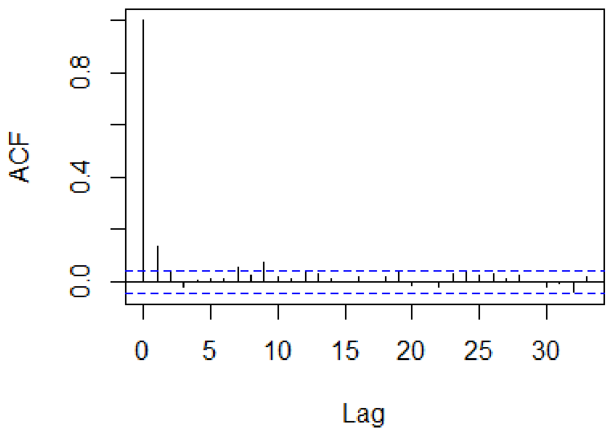


Figure 4.19: Conditional standard deviation for the tier DEXBY004 (top) and DEXBQ002 (bottom)



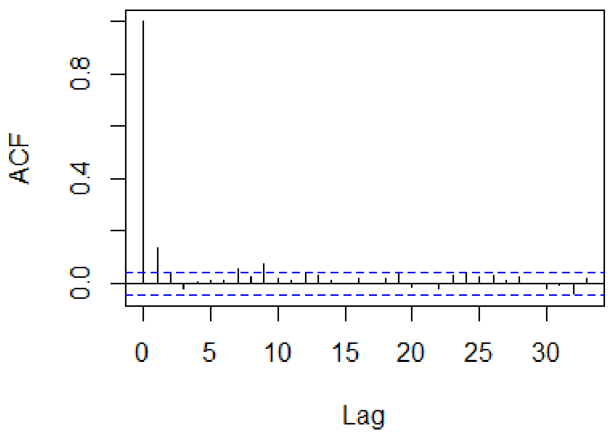


Figure 4.20: Autocorrelation function of the observations for the tier DEXBY004 (top) DEXBQ002 (bottom).

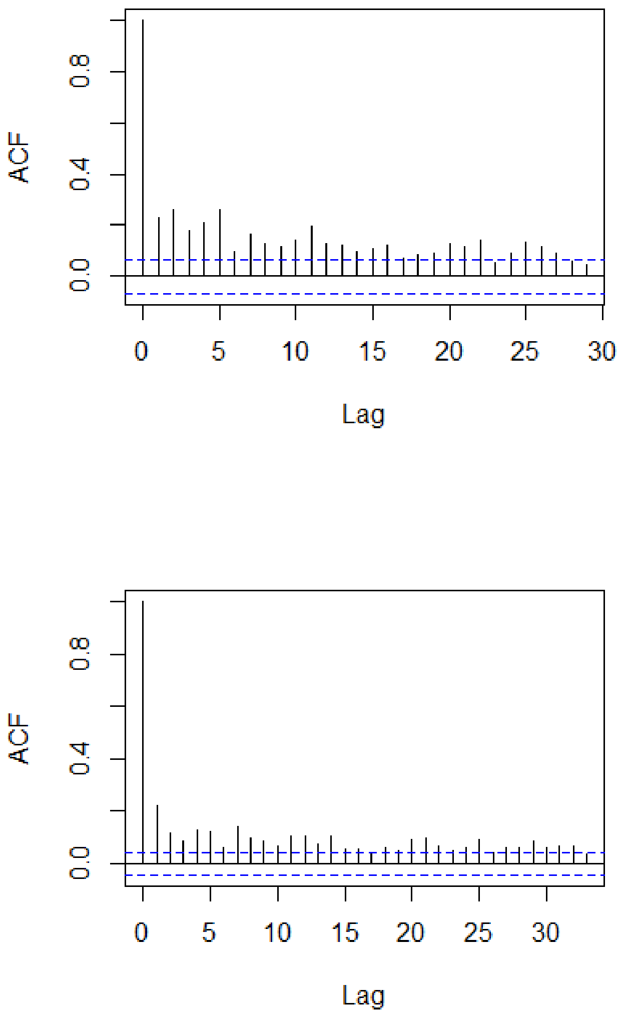
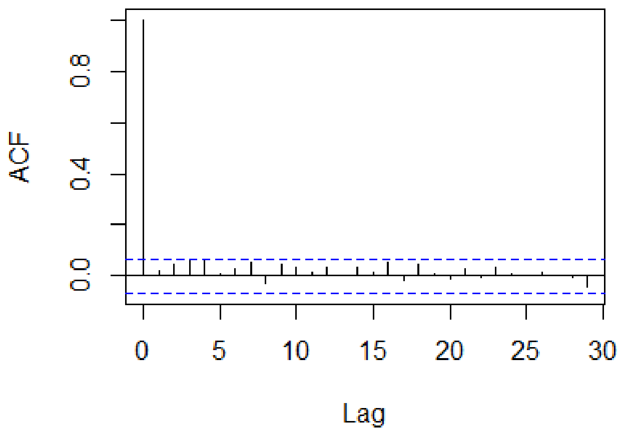


Figure 4.21: Autocorrelation function of the squared observations for the tier DEXBY004 (top) DEXBQ002 (bottom).



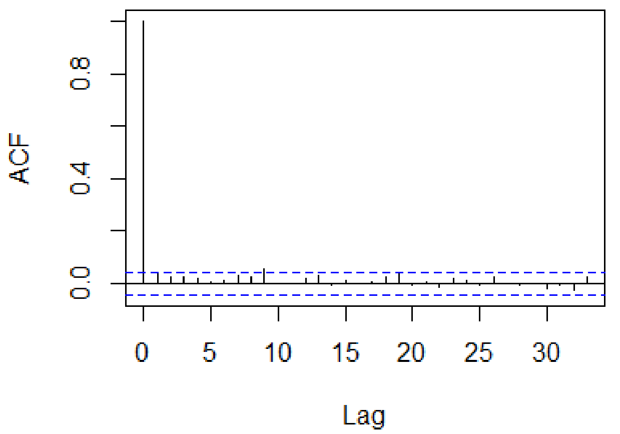
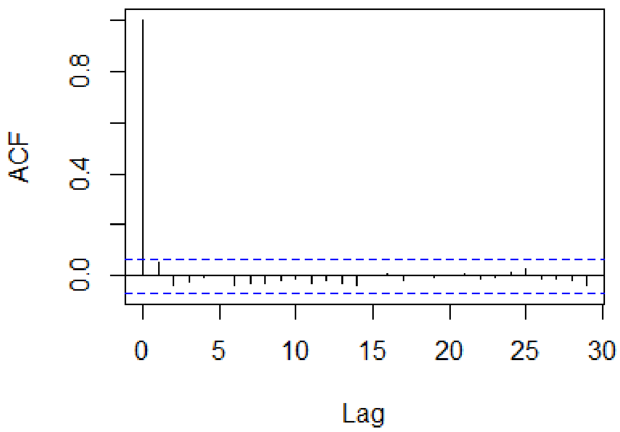


Figure 4.22: Autocorrelation function of the (estimated) Z_t innovations for the tier DEXBY004 (top) and DEXBQ002 (bottom).



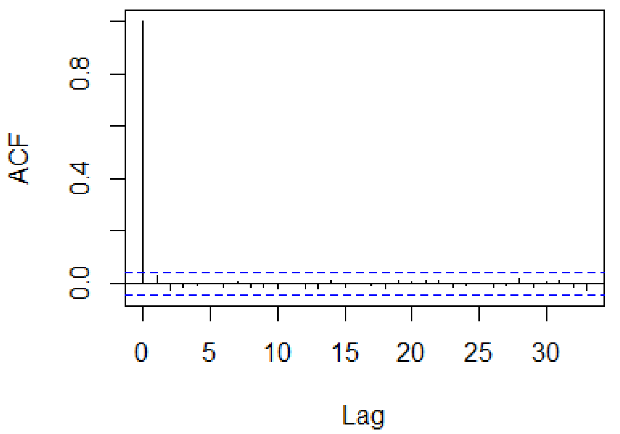
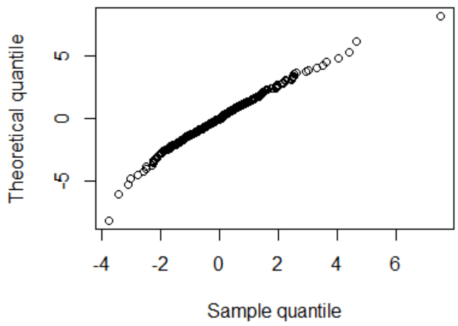


Figure 4.23: Autocorrelation function of the (estimated) squared innovations Z_t^2 for the tier DEXBY004 (top) and DEXBQ002 (bottom).



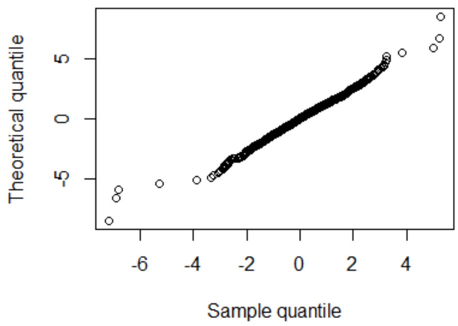


Figure 4.24: QQ plots of the sample quantile ($x-axis$) against the theorized Student's t distribution ($y-axis$) with degrees of freedom as given in Table 4.5 for DEXBY004 (top) and DEXBQ002 (bottom)

4.2.1 Fitting copulas to the innovations

As the ARMA(1,1)-GARCH(1,1) with Student's t distributed innovations again performs satisfactory we fit copulas to the innovations Z_t that belongs to the time bucket $[57, 85)$ for the tier DEXBQ002 and time bucket $[729, 1095)$ for the DEXBY004 and perform the goodness of fit test from 2.2.4. The results are given in Table 4.6. A scatter plot of the innovations Z_t belonging to the time buckets is given in Figure 4.25.

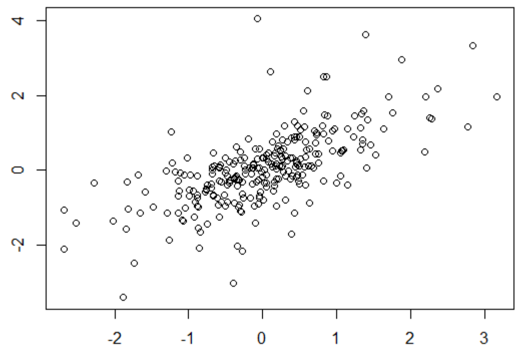


Figure 4.25: A scatter plot of the innovations Z_t from the tiers DEXBQ002 and DEXBY004 belonging to time bucket $[57, 85)$ and $[729, 1095)$.

Copula	Parameter(s)	p-value	AIC
Normal	$\rho = 0.700$	0.103	-175.3028
t	$\rho = 0.706$ $\nu = 9.081$	0.157	-177.3911
Clayton	$\theta = 1.184$	0.0004995004995005	-120.612218234378
Frank	$\theta = 5.800$	0.0436	-171.834
Gumbel	$\theta = 1.928$	0.265	-176.744686740601

Table 4.6: Parameters, p-value and AIC value for five different copulas for the innovations of yearly contracts in the time bucket $[729, 1095)$ and quarterly contracts in the time bucket $[57, 85)$.

A plot of the copulas we cannot reject at the 0.05 level given the results in 4.6 is given in Figure 4.26

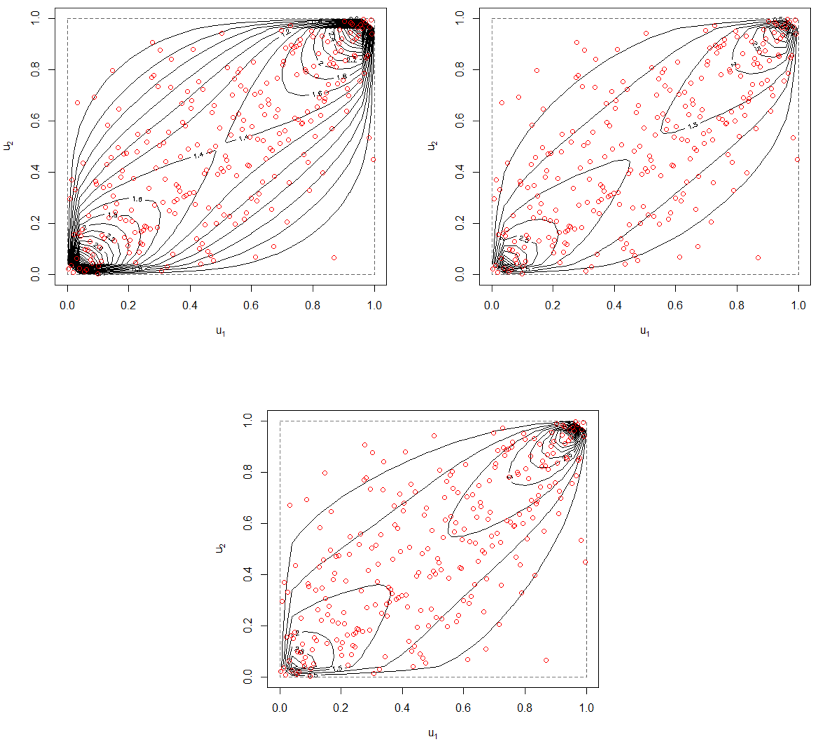


Figure 4.26: Plots of the copula densities with parameters from Table 4.6 with the pseudo observations superimposed (in red). Upper left if the normal copula, upper right is the Student's t copula and bottom is the Gumbel copula

4.2.2 Estimating the ratio

With the estimated copulas we can now estimate the ratio

$$\frac{F_{X+Y}^{-1}(p)}{F_X^{-1}(p) + F_Y^{-1}(p)}$$

for this particular combination of time buckets. Figure 4.27 show the quantiles on the interval $[0.9, 1)$ when the dependence structure is given by the copulas that could not be rejected at the 5% level. In Figure 4.28 the ratio as a function of p is plotted. In Table 4.7 there is a summary of the results at the 99% level.

Copula	$\frac{F_{X+Y}^{-1}(0.99)}{F_X^{-1}(0.99)+F_Y^{-1}(0.99)}$	$F_{X+Y}^{-1}(0.99)$	$F_X^{-1}(0.99) + F_Y^{-1}(0.99)$
Normal	0.931	2.575	2.766
t	0.916	2.534	2.766
Gumbel	0.944	2.612	2.766

Table 4.7: A summary of the different quantiles at the 99% level for the different copulas that could not be rejected.

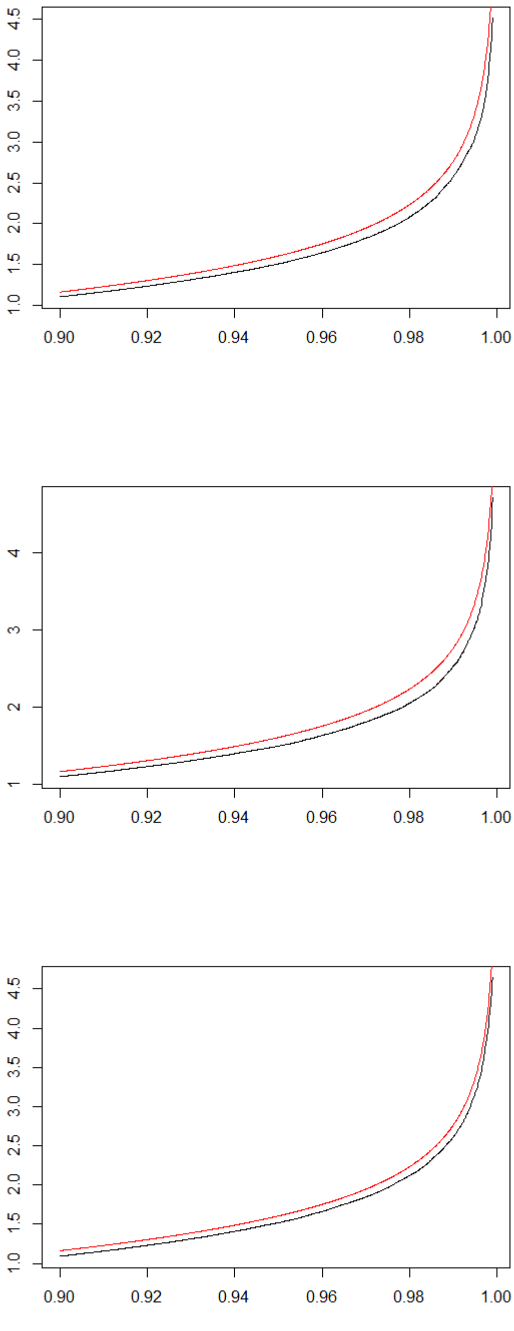


Figure 4.27: Plot of the quantile $F_{X+Y}^{-1}(p)$ (black) and the sum of the two quantiles $F_X^{-1}(p) + F_Y^{-1}(p)$ (red) with $p \in [0.9, 1)$ on the $x - axis$ for dependence given by the normal copula (top), t copula (middle) and Gumbel copula (bottom)

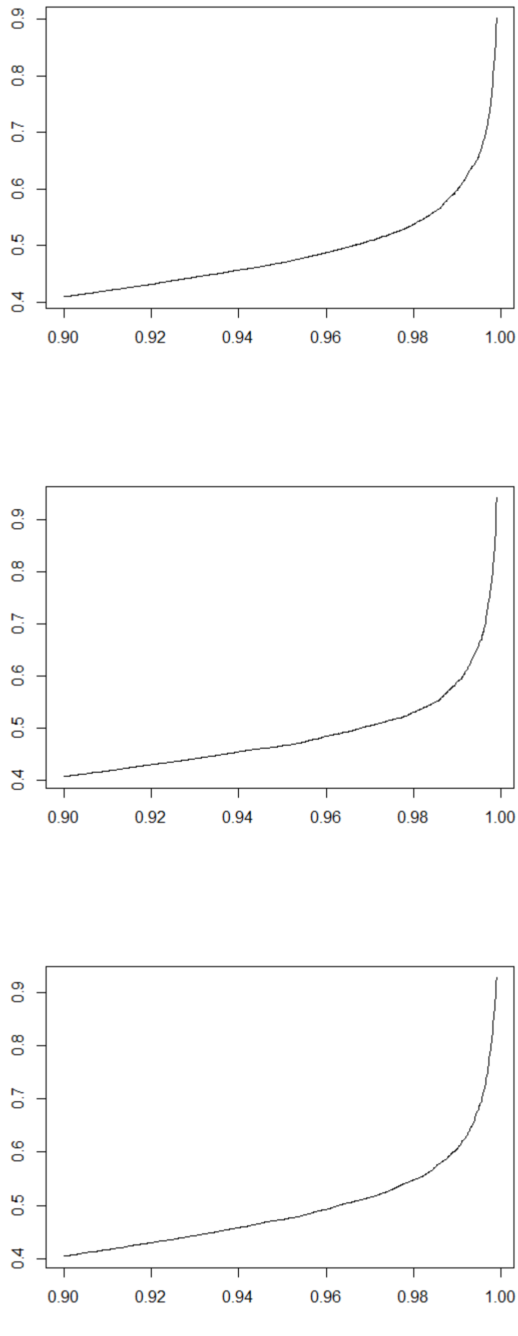


Figure 4.28: The ratio $\frac{F_{X+Y}^{-1}(p)}{F_X^{-1}(p)+F_Y^{-1}(p)}$ as a function of p for $p \in [0.9, 1)$ when dependence structure is given by the Normal copula (top), t copula (middle) and Gumbel copula (bottom).

4.3 Time buckets close to each other

As a last example, let us again look at the tier DEXBQ002 but this time in combination with the tier DEXBY002. Let us in this example look at the time bucket $[57, 85)$ for the tier DEXBQ002 and the time bucket $[225, 337)$ for the tier DEXBY002. In total, we have 2138 observations for both the tiers DEXBQ002 and DEXBY002. The subset of the samples for which the tier DEXBQ002 belongs to the time bucket $[57, 85)$ and DEXBY002 belongs to the time bucket $[225, 337)$ simultaneously consists of 193 observations. It it is this subset that will later be used to fit the copula. A scatter plot of the data that belong to the tier DEXBQ002 and tier DEXBY002 as well as a scatter plot for the data that belong to time bucket $[57, 85)$ for the tier DEXBQ004 and time bucket $[225, 337)$ for tier DEBXY004 is given in Figure 4.29.

Again, we will use a ARMA(1,1)-GARCH(1,1) process with Student's t distributed innovations for the log returns. We have already presented the tier DEXBQ002 and made a good fit with ARMA(1,1)-GARCH(1,1) with Student's t distributed innovations for that tier. For that reason we omit to present that tier again. The time series of the log returns and the estimated auto correlation function for the observations X_t and the squared observations X_t^2 is given in Figure 4.30, 4.31 and 4.32 respectively for the tier DEXBY002.

Given the previous success for the ARMA(1,1)-GARCH(1,1) specification with Student's t distributed innovations we again try this approach. Since we have already estimated the parameters of the ARMA(1,1)-GARCH(1,1) for the tier DEXBQ002 (Table 4.5), we will below only give the parameters for DEXBY002 (Table 4.8) fitted using maximum likelihood.

Tier	DEXBY002
μ	$-4.867 \cdot 10^{-4}$
ϕ_1	$-4.869 \cdot 10^{-1}$
θ_1	$5.395 \cdot 10^{-1}$
ω	$7.002 \cdot 10^{-7}$
α_1	$9.768 \cdot 10^{-2}$
β_1	$9.002 \cdot 10^{-1}$
ν	8.600

Table 4.8: Parameters for the ARMA(1,1)-GARCH(1,1) process with Student's t distributed innovations for log returns for the tier DEXBY002

For the same reason as stated above, below in Figure 4.33 and Figure 4.34 are plots of the estimated autocorrelation function for the estimated innovations Z_t and squared innovations Z_t for the tier DEXBY002.

In Figure 4.35 is a QQ-plot of the theorized Student's innovations distribution against the sample.

We can from the fitted model extract the conditional standard deviation for the tier DEXBY002. It is given as a time series in Figure 4.36

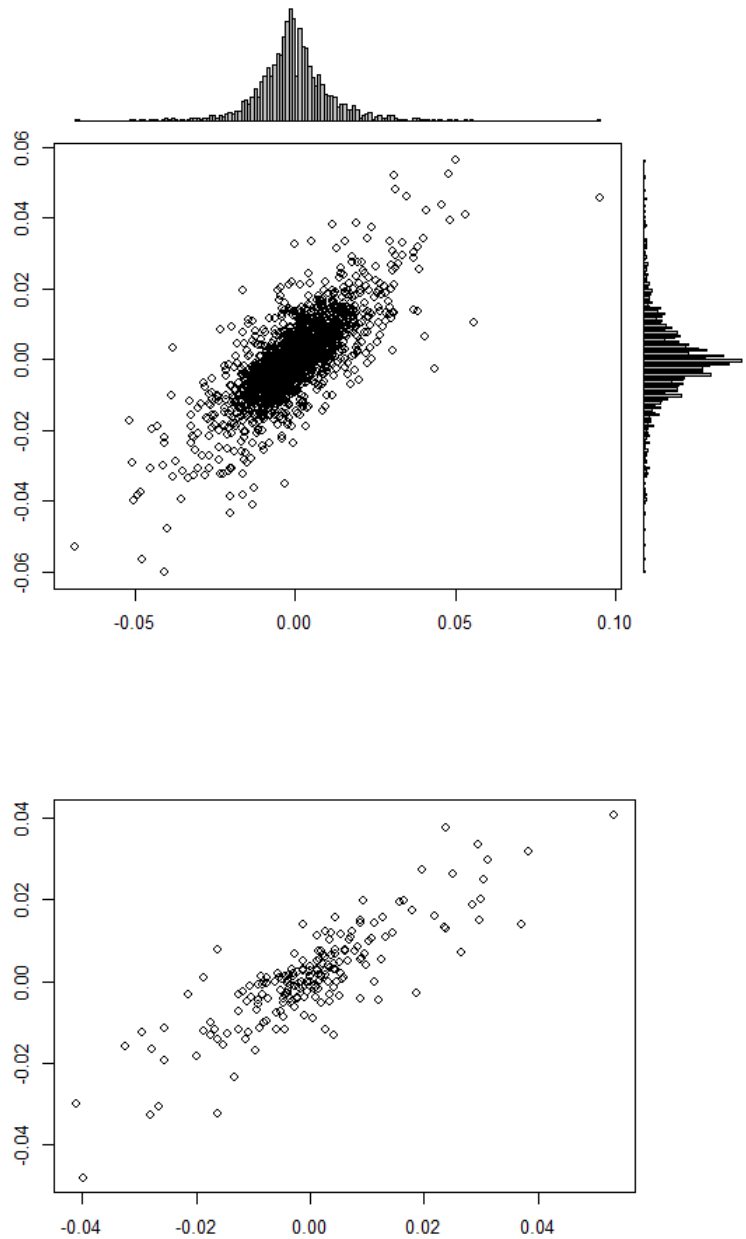


Figure 4.29: Top: Scatter plot for the data. The data belonging to the tier DEXBQ002 on the $y-axis$ and data belonging to the tier DEXBY002 on the $x-axis$. On the axes there is also a histogram. Bottom: Same as above with only the data that belong to time bucket $[57, 85)$ and time bucket $[225, 337)$ (minus the histograms)

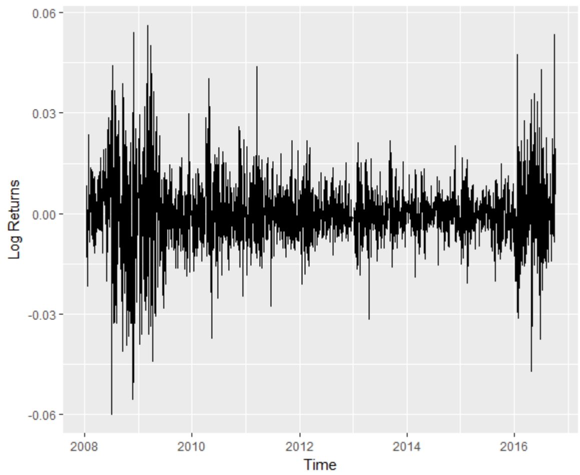


Figure 4.30: Time series of the log returns for the tier DEXBY002

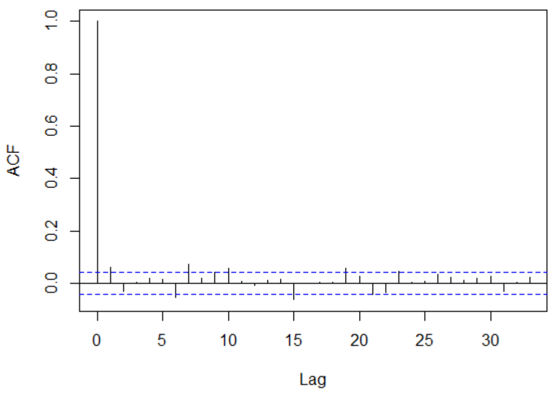


Figure 4.31: Estimated autocorrelation function for the observations X_t for the tier DEXBY002. The banded blue line is a 95% confidence interval.

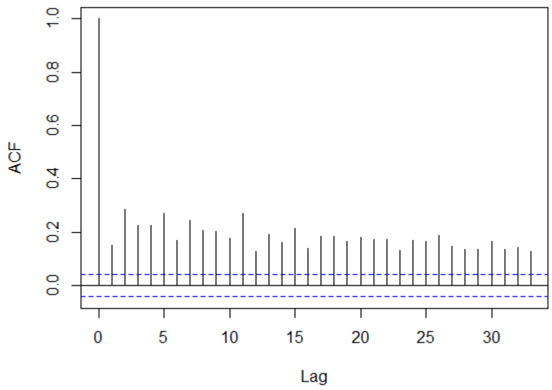


Figure 4.32: Estimated autocorrelation function for the squared observations X_t^2 for the tier DEXBY002. The banded blue line is a 95% confidence interval.

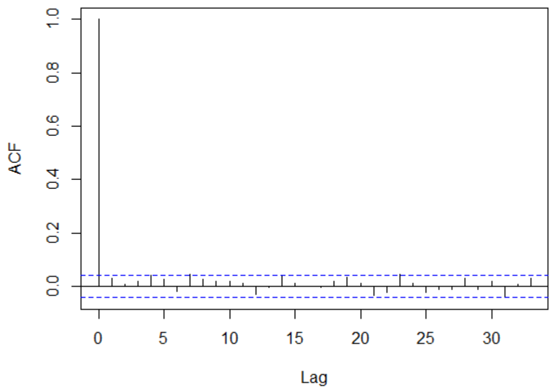


Figure 4.33: Autocorrelation function of the (estimated) innovations Z_t for the tier DEXBY002. The banded blue line is a 95% confidence interval.

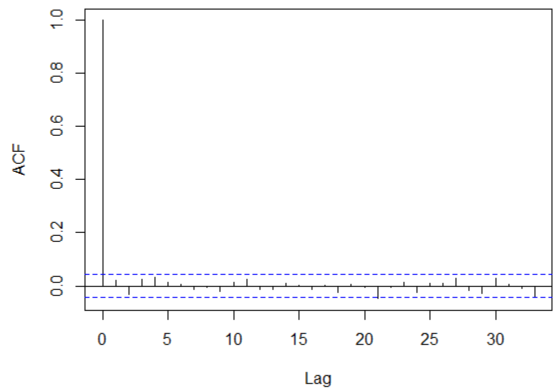


Figure 4.34: Autocorrelation function of the (estimated) squared innovations Z_t^2 for the tier DEXBY002. The banded blue line is a 95% confidence interval.

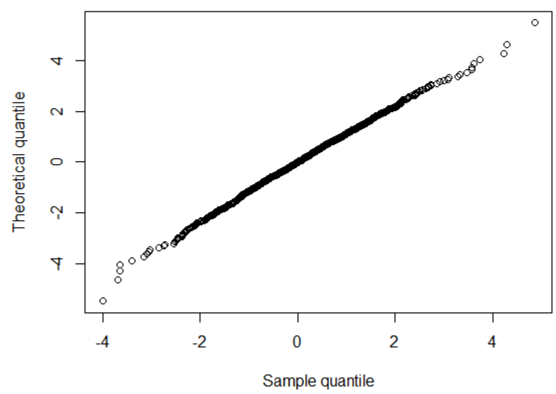


Figure 4.35: QQ plot of the sample quantile $(x - axis)$ against the theorized Student's t distribution $(y - axis)$ with degrees of freedom as given in Table 4.8 for DEXBY002

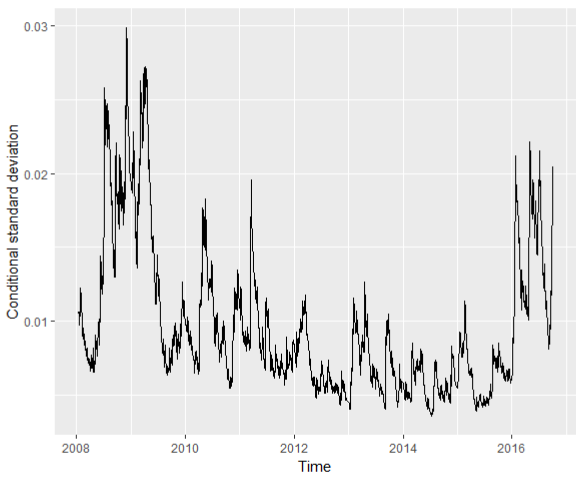


Figure 4.36: Estimated conditional standard deviation for the tier DEXBY002

4.3.1 Fitting copulas to the innovations

Copulas are fitted to the pseudo observations to the innovations of the data given in the scatter plot in Figure 4.37.

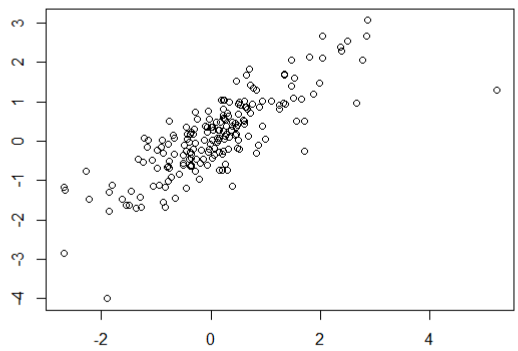


Figure 4.37: Scatterplot of the estimated innovations for quarterly contracts belonging to time bucket $[57, 85)$ $(y - axis)$ and yearly contracts belonging to time bucket $[225, 337)$ $(x - axis)$.

Which gives the coefficients for the different copulas together with the p-values from the goodness of fitness test and AIC-values in Table 4.9.

Copula	Parameter(s)	p-value	AIC
Normal	$\rho = 0.830$	0.506	216.563
t	$\rho = 0.830$ $\nu = 14.125$	0.580	215.462
Clayton	$\theta = 0.961$	0.001	158.525
Frank	$\theta = 8.211$	0.011	194.22828043
Gumbel	$\theta = 2.550$	0.509	214.149584

Table 4.9: Parameters, p-values and and AIC-values for five different copulas fitted to the innovations to quarterly and yearly contracts belonging to the time buckets $[57, 85)$ and $[225, 337)$

The copulas that can not be rejected at the 5% level have the densities given in Figure 4.38

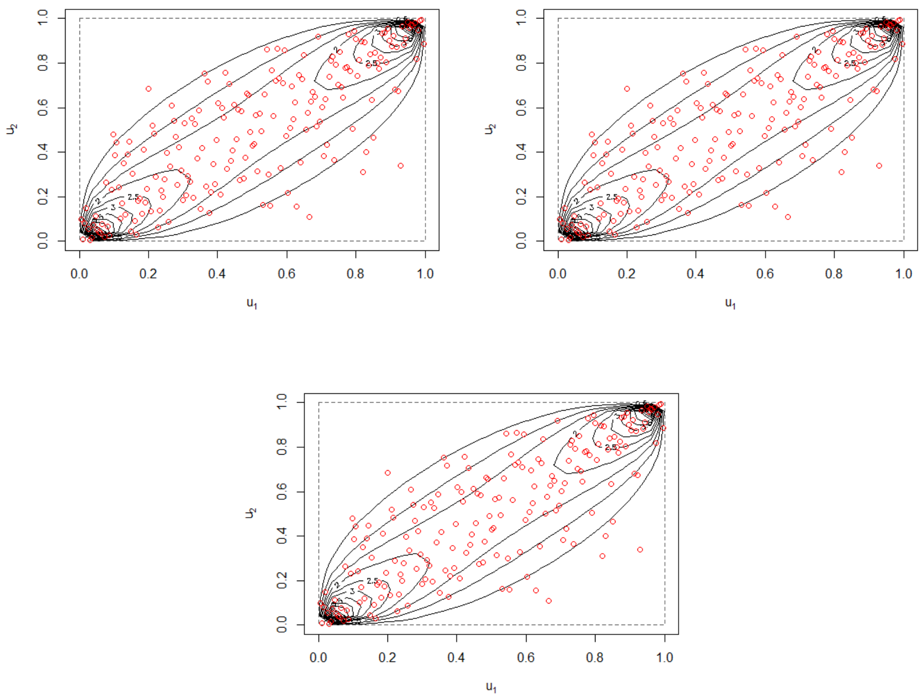


Figure 4.38: Densities for the copulas fitted to the quarterly contracts in time bucket [57, 85) and the yearly contracts in time bucket [225, 337). The order is clockwise, Normal copula, t copula and Gumbel copula. Superimposed in red is the pseudo observations.

4.3.2 Estimating the ratio

We can just as before by simulation estimate the ratio for this choice of time buckets.

$$\frac{F_{X+Y}^{-1}(p)}{F_X^{-1}(p) + F_Y^{-1}(p)}$$

Figure 4.39 show the quantiles on the interval $[0.9, 1)$ when the dependence structure is given by the copulas that could not be rejected at the 5% level.

In Figure 4.40 we have plots of the ratio as a function of p

In Table 4.10 there is a summary of the results at the 99% level.

Copula	$\frac{F_{X+Y}^{-1}(0.99)}{F_X^{-1}(0.99)+F_Y^{-1}(0.99)}$	$F_{X+Y}^{-1}(0.99)$	$F_X^{-1}(0.99) + F_Y^{-1}(0.99)$
Normal	0.968	3.146	3.250
t	0.972	3.159	3.250
Gumbel	0.985	3.199	3.250

Table 4.10: A summary of the different quantiles at the 99% level for the different copulas that could not be rejected.

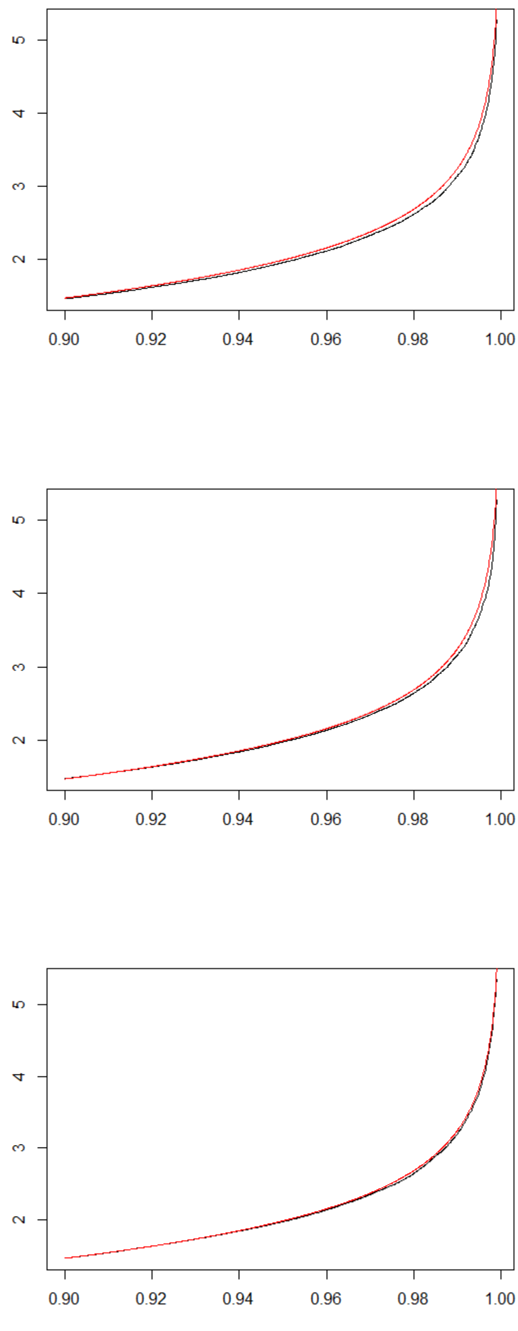
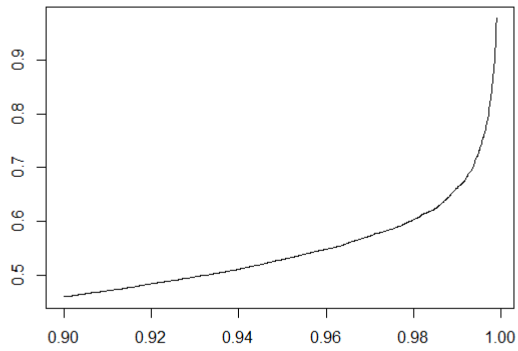
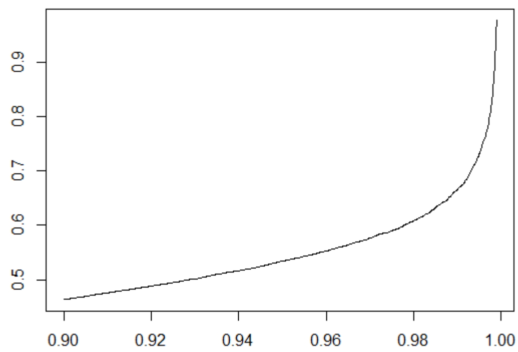


Figure 4.39: Quantiles $F^{-1}_{X+Y}(p)$ (black line)$F^{-1}_X(p) + F^{-1}_Y(p)$ (red line) for $p \in [0.9, 0.99)$ when the dependence structure is given by the normal copula (top), t copula (middle) and Gumbel copula (bottom).





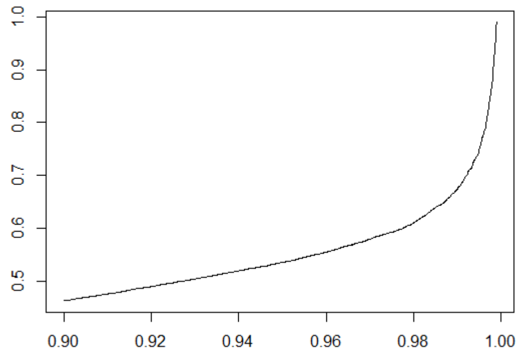


Figure 4.40: The ratio $\frac{F_{X+Y}^{-1}(p)}{F_X^{-1}(p)+F_Y^{-1}(p)}$ as a function of p for $p \in [0.9, 1)$ when dependence structure is given by the Normal copula (top), t copula (middle) and Gumbel copula (bottom).

Chapter 5

Analysis and conclusion

5.1 Analysis

The ARMA-GARCH process shows promise to model the dependency between returns for electricity derivatives. The main advantage of this model is the possibility to take care of heteroscedeicity in the return data. However, this comes with some notable limitations.

5.1.1 The role of time buckets

A question that still is not answered is what role the time buckets play relative to the tiers. If we return to the two tiers DEXBQ004 and DEXBY003 and this time instead fit a copula to all the pairs of innovations from the fitted ARMA(1,1)-GARCH(1,1) processes (from Table 4.2) that belongs to that particular choice of tiers, we get a somewhat different result. Table 5.1 shows the result of fitting copulas to the entire sample. In leftmost column of the table, the parameters haven been transformed into Kendall's τ using the formulas from Section 2.3.2. For ease of comparison, the result from Section 4.1.3 are included in table 5.2. In that table the parameters have been transformed into Kendall's τ as well using the formulas from Section 2.3.2.

Copula	Parameter(s)	Kendall's τ
Normal	$\rho = 0.821$	0.6129036
t	$\rho = 0.835 \nu = 5.149$	0.6295034
Clayton	$\theta = 2.036$	0.5044616
Frank	$\theta = 9.007$	0.6369347
Gumbel	$\theta = 2.573$	0.6113263

Table 5.1: Parameters for the five different copulas fitted to innovations from the $ARMA(1,1) - GARCH(1,1)$ from Section 4.1.2 for all observed pairs belonging to the tiers DEXBQ004 and DEXBY003

Copula	Parameter(s)	Kendall's τ
Normal	$\rho = 0.7974$	0.5875605
t	$\rho = 0.8073$ $\nu = 6.2551$	0.5981841
Clayton	$\theta = 1.848$	0.4801768
Frank	$\theta = 8.137$	0.6076125
Gumbel	$\theta = 1.848$	0.5745010

Table 5.2: Parameters for the five different copulas fitted to innovations from the $ARMA(1,1) - GARCH(1,1)$ from Section 4.1.2 for observations belonging to the tiers DEXBQ004 and DEXBY003 and the time buckets $[225, 337)$ and $[393, 540)$

We can from Table 5.2 and Table 5.1 see that there is difference between the time bucket sample and the *full* sample.

The book *Nonparametric statistical methods*[10] provides a method for constructing a two sided confidence interval for Kendall's tau. We find that when using the entire data set a two sided confidence interval for Kendall's tau is given by $(0.620, 0.655)$ at the 95% level. In comparison, for the sub sample in time buckets $[225, 337)$ and $[393, 540)$ we find that a two sided confidence interval at the 95% level is given by $(0.570, 0.645)$. There is an overlap between these two confidence intervals with the confidence interval for the full sample and the confidence interval for the sub setted sample. The confidence interval for the full sample is smaller and numerically larger than the confidence interval for the subsetted sample. This may imply that the true dependence is best captured using the whole sample, rather than the subsetted sample with time buckets. We also find that for the whole sample, Kendall's *tau* as given by the copula parameters is higher, implying a greater level of concordance (in this case, prices moving in the same direction). We also find, since the tail dependence is an increasing function[1] of the parameters, that this implies a higher level of tail dependence as shown in Table 5.3

We conclude from this small analysis that the time buckets, in the sense of looking for dependence between contracts, could be ignored for this example and that it is better in a sense to use the tiers only. We do not however reject the idea of there being a dependence conditional on time between the contacts, only that it is hard to statistically observe using the time buckets.

5.1.2 The use of GARCH processes

Another question is if there is any need to use the ARMA-GARCH process. To answer that, let us try a naive approach. Let us once again return to the first case we studied, the case with the tiers DEXQ004 and DEXBY003 If we

[1]except for the t copula, where it is increasing in ρ and decreasing in ν

Subsetted sample	Normal	t	Clayton	Frank	Gumbel
Lower	0	0.4068396	0.7114549	0	0.0000000
Upper	0	0.4068396	0.0000000	0	0.6908107
Whole sample	Normal	t	Clayton	Frank	Gumbel
Lower	0	0.4850863	0.6871595	0	0.0000000
Upper	0	0.4850863	0.0000000	0	0.6569631

Table 5.3: The implied lower and upper tail dependence for the different copulas fitted to the pair of observations from DEXBQ004 and DEXBY003 for the subsetted sample belonging to the time buckets $[225, 337)$ and $[393, 540)$ (first three rows) and the whole sample (last three rows).

try to fit a copula to the "raw" data, the log returns, we get different results from when using the ARMA-GARCH approach. Since we use a semiparametric method, there is no need for us to first specify a distribution for the returns, and we can fit copulas directly on the pseudo observations. Doing so to unfiltered observations results in the copulas given in table 5.4

Copula	Parameter(s)
Normal	$\rho = 0.81878054$
t	$\rho = 0.8321460\nu = 2.3988$
Clayton	$\theta = 2.414$
Frank	$\theta = 8.796$
Gumbel	$\theta = 2.595$

Table 5.4: Parameters for copulas fitted to quarterly and yearly contracts belonging to the time buckets $[225, 337)$ and $[393, 540)$ without filtering the observations by means of a ARMA-(1,1)-GARCH(1,1)

The difference is small, but noticeable especially in the case of the degrees of freedom for the t-copula. Without GARCH filtering, the degrees of freedom ν is estimated to be 2.3998 compared to 6.2551 (Table 5.2) for the case with GARCH filtering. This has, as mentioned above an effect on the tail dependence, which is 0.6115 for the upper and lower tail dependence in the unfiltered case compared to 0.407 (Table 5.3) for upper and lower tail dependence for the filtered case. This is effect suspected to be caused by the use of the semi parametric pseudo log likelihood method. The assumed reason is that for the unfiltered case, large values in absolute value for the log returns may be from periods of high volatility without necessarily corresponding to high or low value of the quantile of the innovations. With this small analysis we draw the conclusion that filtering using a GARCH process has the effect of providing better estimates for the copulas. This is though hardly surprising. One of the core assumptions when fitting copulas is that the sample is to be independently identically distributed. Which, in the presence of heteroskedasticity and/or autoregressive/moving average effects, it is not. The effect of the GARCH filtering is that we get more "I.I.D like" observations.

5.1.3 The nature of the dependence

We can from the results see that the dependence between yearly and quarterly contracts change as the contacts move closer to delivery. Not only does the specifics of the dependence change (change in parameters) but we can also see that as the contracts move closer to delivery, the nature of dependence changes as well. We have looked at a total of three different scenarios. If we from each scenario take the copula that has the lowest AIC from those that could not be rejected using the goodness of fit test, we arrive at Table 5.5

Time bucket	Copula lowest in AIC	Paramters(s)
225, 337, Q 393, 540 Y	t	$\rho = 0.807$ $\nu = 6.255$
57,85, Q 729, 1095 Y	t	$\rho = 0.705$ $\nu = 9.081$
57, 85 Q 225, 337 Y	Normal	$\rho = 0.830$

Table 5.5: The copula and the corresponding parameter(s) that is lowest in AIC that could not be rejected at the 95% level for the different scenarios investigated. The letter to the right in the Time bucket column refers to Quarterly contracts (Q) and Yearly contracts (Y).

As can be seen in Table 5.5, the dependence structure between the contracts is dependent on the time to delivery. It can be seen that not only does the dependence change in correlation, it also changes in nature. It should be noted here again, that as $\nu \to \infty$ the t copula becomes the normal copula. For the last case, when the normal copula was lowest in AIC, the degrees of freedom ν for the estimated t copula was quite high. It was in that case estimated to $\nu \approx 14$ (Table 4.9).
We have seen that in general, with the exception of the Gumbel copula, the archimedian copulas could be rejected at the 95% confidence level. The dependence is, according to the copulas studied, best described by the elliptical copulas, with the t-copula being the probable choice. Most likely for its symmetrical tail dependence property.

5.2 Conclusions

A clear result is that the use of copulas, despite the higher computational effort to estimate them, provide a better understanding of the dependence, than say elliptical distributions. In this work the time buckets have been taken as a given fact. However, as we saw in Section 5.1.1, the time buckets may not contain the whole truth from a dependence point of view. We can however see that the dependence is conditional on time, that is, the dependence between the contracts is dependent on time to delivery. A shift of focus from the time buckets to tiers only would dramatically lower the amount of effort needed to get a more complete picture of the dependence and how it evolves through time. For example. If the quarterly contracts passes through 21 time bucket before they begin delivery and the yearly contracts passes through 24, that means that there are 504 different combinations of time buckets, only for this particular choice of contract types. If a quarterly contract can be traded three years before delivery and if the yearly contract can be traded five years before delivery that means there are *only* 72 different combinations, with a two dimensional approach. As these numbers show, getting the fuller picture is a no small task, and that is why so few examples are provided.
One big unaddressed problem of the approach used is that we do not know how the sought ratio (Equation (1.1)) depends on volatility (conditional standard deviation). As this approach used an ARMA-GARCH model to filter away the heteroscedacity of the data and could find the sought ratio it only did so at a particular date (the end of the time series) and thus at a particular volatility level. A question that arises then is how the ratio behaves when two contracts in different time buckets behave when they are at different volatility levels. This question remains unanswered.
The model has also not been tested in higher dimensions. While there is no theoretical obstacle for increasing the number of time buckets under investigation, this will probably result in a poor fit. As we have seen, the most probable candidate to best capture the dependence is the t-copula. But, the tail dependence for the t copula is determined by two variables. One is the linear correlation coefficient that can be determined for two random variables independent of the choice of dimension. The other variable, the degree of freedom ν is fixed for every pair of random variables. This leads to a somewhat inflexible model, that may fail to describe the tails properly in higher dimensions. While we have shown in Theorem 2.17 that multivariate archimedian copulas are also possible, they are determined by one parameter only, regardless of dimension. This leads to very inflexible models and would probably fail to properly describe the higher dimensional dependence.
The model also does not take into account a stressed market. Conventional wisdom say that during a stressed market, dependence takes a different form. Another issue not addressed in this thesis is that results obtained were not

back tested. That is, if there was a breach in margin in one or two contracts, was the combined price movement less than the summed price movements multiplied by the ratio and how often was it not. However performing such a test in a realistic way would be time and computational power consuming and would require many decisions and assumptions regarding how often the ratio is calculated, how often should the GARCH processes estimated, can we assume we have all the data, should margin be determined by the current margin model and so on.

5.3 Further Work

In order to get a full understanding of the problem, the questions raised in 5.2 and other questions need to be addressed.

A first approach for the problem of finding the ratio (1.1) at different volatility levels that perhaps would work from a risk management perspective is to look at the ratio for the date when the squared sum of the volatility is the largest, ie. when

$$\sqrt{\sigma_1^2 + \sigma_2^2}$$

is largest or corresponds to some chosen quantile. This would have the advantage of creating a bound for the ratio.

Given sufficient computing power, one could perhaps look at the ratio as not only as a function of p but also of the different volatility levels, i.e. look at the ratio

$$\frac{F_{X+Y}^{-1}(p, \sigma_x, \sigma_Y)}{F_X^{-1}(p, \sigma_X) + F_Y^{-1}(p, \sigma_Y)} \tag{5.1}$$

for some fixed value of p, and create a surface plot of the ratio. This approach would have to find some way to take into account the dependence between σ_X and σ_Y given by the copula and the ARMA-GARCH processes.

Another issue is to found out when do the dependence structure change? That is, at which point in the time before delivery can we find statistically significant changes in the dependence, and is this change in the form of different parameters, or in a completely different copula? This could perhaps be achieved using a Hidden Markov Model (HMM). Then the hidden states of the model would be certain dependence structure specified by some copula. Perhaps it is possible to achieve this in some sort of Bayesian framework. The models performance in higher dimensions would also need to be investigated. Is there a possibility to use Pair Copula Construction, (*"vines"*) to describe the dependence in higher dimensions? As was mentioned in Section 5.2, stressed markets was not taken into account. If we can identify stressed markets in the data, does the dependence stay the same as in *normal* market conditions, or would we need to use so called extreme value copulas in that case?

Bibliography

[1] Nasdaq Clearing AB. *Nasdaq SPAN - Margin Methodology guide for Commodity derivatives.* 2017.

[2] Rene Aid. *Electricity Derivatives.* Springer International Publishing, 2015.

[3] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327, 1986.

[4] David Beuadoinc Christian Genesta, Bruno Remillard. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 2009.

[5] R.F Engle. Autoregressive conditional heteroscedasticity with estimate of the variance of uk inflation. *Econometrica*, 50:987–1007, 1432.

[6] Christian Francq and Jean-Michel Zakoïan. *Estimating GARCH Models by Quasi-Maximum Likelihood*, pages 141–184. John Wiley Sons, Ltd, 2010.

[7] M. J. Frank, R. B. Nelsen, and B. Schweizer. Best-possible bounds for the distribution of a sum — a problem of kolmogorov. *Probability Theory and Related Fields*, 74(2):199–211, 1987.

[8] Alexander J. McNeil Rudiger Frey and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press, 2005.

[9] Carl Johan Rehn Henrik Hult. Filip Lindskog, Ola Hammarlid. *Risk and portfolia analysis - Principles and Methods.* Springer Verlag, 2012.

[10] Myles Hollander, Douglas A. Wolfe, and Eric Chicken. *Nonparametric Statistical Methods.* John Wiley Sons Inc., 2015.

[11] John C. Hull. *Options, future, and other derivatives.* Prentice Hall, 2012.

[12] Ivan Kojadinovic and Jun Yan. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21(1):17–30, 2011.

[13] G. D. Makarov. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory Probab. Appl.,*, 1981.

[14] Rober B. Nelsen. *An introduction to copulas.* Springer Science+Business Media, Inc., 2006.

[15] A.J. Patton. Modelling asymmetric exchange rate dependence. *International Economic Review*, 2006.

[16] Alexander McNeil Paul Embrecht, Filip Lindskog. Modelling dependence with copulas and applications to risk management. *None*, 2012.

[17] Daniel Straumann Paul Embrechts, Alexander Mcneill. Correlation and dependency in risk management: properties and pitfalls. -, 1999.

[18] Richard A Davis Peter J Brockwell. *Time Series Theory and methods.* Springer publishing, 1991.

TRITA -MAT-E 2017:49
ISRN -KTH/MAT/E--17/49--SE