EXAMENSARBETE INOM MASKINTEKNIK,
AVANCERAD NIVÅ, 30 HP
STOCKHOLM, SVERIGE 2021

DCC-GARCH Estimation

Forecasting conditional correlation and covariance

CHRISTOFER NORDSTRÖM

KTH
SKOLAN FÖR TEKNIKVETENSKAP

Abstract

When modelling more that one asset, it is desirable to apply multivariate modeling to capture the co-movements of the underlying assets. The GARCH models has been proven to be successful when it comes to volatility forecasting. Hence it is natural to extend from a univariate GARCH model to a multivariate GARCH model when examining portfolio volatility. This study aims to evaluate a specific multivariate GARCH model, the DCC-GARCH model, which was developed by Engle and Sheppard in 2001. In this paper different DCC-GARCH models have been implemented, assuming both Gaussian and multivariate Student's t distribution. These distributions are compared by a set of tests as well as Value at Risk backtesting.

Keywords:

Multivariate GARCH, DCC-GARCH, Conditional Correlation, Forecasting

Sammanfattning

I portföljanalys så är det åtråvärt att applicera flerdimensionella modeller för att kunna fånga hur de olika tillgångarna rör sig tillsammans. GARCH-modeller har visat sig vara framgångsrika när det kommer till prognoser av volatilitet. Det är därför naturligt att gå från endimensionella till flerdimensionella GARCH-modeller när volatiliteten av en portfölj skall utvärderas. Den här studien ämnar att utvärdera tillvägagångssättet för prognoser av en viss typ av flerdimensionell GARCH-modell, DCC-GARCH-modellen, vilken utvecklades av Engle och Sheppard 2001. I den här uppsatsen har olika DCC-GARCH modeller blivit implementerade, som antar innovationer enligt både flerdimensionell normalfördelning samt flerdimensionell student's t-fördelning. Dessa jämförs med hjälp av en handfull tester samt Value-at-Risk backtesting.

Nyckelord:

Flerdimensionella GARCH-modeller, DCC-GARCH, Betingad Korrelation, Prognoser

Table of Contents

Page

1 Introduction **6**

2 Purpose **7**

3 Theory **8**
3.1 Univariate GARCH . . . 8
3.1.1 ARCH model . . . 8
3.1.2 GARCH model . . . 8
3.2 Multivariate GARCH . . . 9
3.2.1 Introduction to multivariate GARCH . . . 10
3.2.2 CCC-GARCH . . . 11
3.2.3 DCC-GARCH . . . 12

4 Forecasting Procedure and Evaluation **14**
4.1 Gaussian distributed innovations . . . 14
4.2 Multivariate Student's t distributed innovations . . . 16
4.3 Forecasting DCC-GARCH . . . 17

5 Evaluation of estimations **19**
5.1 Goodness of fit for univariate distributions . . . 19
5.2 Goodness of fit for multivariate distributions . . . 20

6 Empirical Application **22**
6.1 Data evaluation . . . 23
6.2 Estimations . . . 25
6.3 DCC-forecasts . . . 26
6.3.1 D_t-forecasts . . . 26
6.3.2 R_t-forecasts . . . 27
6.3.3 H_t-forecasts . . . 28

7 Goodness of fit **29**
7.1 Marginal distributions . . . 29
7.1.1 Visual error evaluation . . . 29
7.1.2 The auto correlation function . . . 30
7.1.3 Ljung Box test . . . 31
7.2 Multivariate goodness of fit . . . 32
7.2.1 Violations . . . 32
7.2.2 Kupiec test . . . 33
7.2.3 Christoffersen test . . . 33

8 Discussion and conclusion **34**

9 References **35**

10 Appendix A **37**

List of Figures

1 For plots for the mean corrected returns, $\mathbf{y}_t$ 23
2 For plots for the ACF, $\mathbf{y}_t^2$ 24
3 For plots for the ACF, $\mathbf{y}_t$ 24
4 Plots for $\mathbf{D}_{t+k}$, Univariate Normal distribution 26
5 Plots for $\mathbf{R}_{t+k}$, Gaussian distribution 27
6 Plots for $\mathbf{R}_{t+k}$, MV Student's t distribution 27
7 Plots for $\mathbf{H}_{t+k}$, Gaussian distribution 28
8 Plots for $\mathbf{H}_{t+k}$, MV Student's t distribution 28
9 Plots for ϵ_t, Gaussian errors 29
10 Plots for ϵ_t, MV Student's t errors 29
11 ACF for ϵ_t, Gaussian errors 30
12 ACF for ϵ_t, MV Student's t errors 30
13 Ljung Box test, Gaussian errors 31
14 Ljung Box test, MV Student's t errors 31

1 Introduction

Understanding of volatility is crucial when it comes to financial modelling. It is of importance since it can be used as a measurement of risk for an asset. It is well known that variance tend to vary over time and that it tends to come in clusters. This is a phenomenon which is called heteroscedasticity, that small changes in an index tends to be followed by small changes and great changes in an index follows by great changes. Before the Autoregressive Conditional Heteroscedasticity (ARCH) process was introduced by Robert F. Engle in 1982 [10], traditional econometric models assumed constant one period variance, which of course is not realistic. With the introduction of the ARCH-model the conditional variance could be modelled as a function of past errors, leaving the unconditional variance constant. In 1986, Bollerslev [2] further developed the ARCH model, into the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. The GARCH-models has been proven successful when it comes to forecast volatility for an asset.

However, the univariate GARCH model can only be used on one asset at a time. In financial mathematics it is often relevant to analyze situations from a portfolio perspective rather than looking at individual assets, which is why a multivariate extension of the GARCH process is of interest. As the multivariate extension allows for analysis of the co-movements in the portfolio. The multivariate GARCH models follow the same structure of the univariate GARCH model, however it makes it possible to make joint analysis and forecasts. There is a wide range of multivariate GARCH models. This thesis evaluates a specific type of multivariate model, the Dynamic Conditional Correlation (DCC) GARCH model which was introduced by Engle and Sheppard in 2001 as an extension of the already existing Constant Conditional Correlation (CCC) GARCH model [12].

The DCC-model has been used in this thesis to forecast an equally distributed currency basket for two different error distributions. The practical implementation of has been done in the softwares of R and MATLAB. The MFE-toolbox[1] has been used in MATLAB and the rmgarch-package[2] has been used in R.

The structure of this thesis is to introduce the reader to univariate and multivariate GARCH-theory in chapter 3, the forecasting procedure in chapter 4, evaluation methods in chapter 5, showing results in chapter 6, discussing goodness of fit in chapter 7 and finally with a discussion in chapter 8.

[1]Which is made by the co-author of the DCC model, Kevin Sheppard, 2013.

[2]The rmgarch package in R is created by Alexios Ghalanos, 2019

2 Purpose

The purpose of this study is to estimate and evaluate time varying covariances and correlations by applying a DCC-GARCH model. The purpose is stated in the following research questions:

1. How are the parameters chosen for a DCC-GARCH model?

2. How can a DCC-GARCH model be used for forecasts?

3 Theory

3.1 Univariate GARCH

3.1.1 ARCH model

Before the Autoregressive Conditional Heteroscedasticity (ARCH) process was introduced by Robert F. Engle in 1982 [10], traditional econometric models assumed constant one period variance. With the introduction of the ARCH-model the conditional variance could be modelled as a function of past errors, leaving the unconditional variance constant. The ARCH-processes are zero mean, serially unconditional processes with nonconstant variance conditional on the past. Hence, information about the recent past can with an ARCH-model describe future variance.

$$y_t = \epsilon_t h_t^{1/2} \tag{1}$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + ... + \alpha_q y_{t-q}^2 \tag{2}$$

Where y_t is a zero mean process at time t, ϵ_t is error term at time t, h_t conditional variance at time t, α_t model parameter for lag t and q order of ARCH-model.

Adding the assumption of normality, the ARCH-model can be written as the following. Given Ψ_{t-1} as the information set (σ-field) up to and including time $t-1$ generated by the observed series y_t:

$$y_t \mid \Psi_{t-1} \sim N(0, h_t) \tag{3}$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + ... + \alpha_q y_{t-q}^2 \tag{4}$$

3.1.2 GARCH model

The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model was developed by Bollerslev in 1986 [2] as a generalization of the ARCH process. The extension from the ARCH-model to the GARCH-model allows for past conditional variances in the current conditional variance equation. The GARCH(p, q)-model is defined as, with normal innovations:

$$y_t \mid \Psi_{t-1} \sim N(0, h_t) \tag{5}$$

$$r_t = \mu_t + y_t \tag{6}$$

$$y_t = h_t^{1/2} \epsilon_t \tag{7}$$

$$h_t = \alpha_0 + \underbrace{\sum_{i=1}^{q} \alpha_q \epsilon_{t-q}^2}_{\text{ARCH-terms}} + \underbrace{\sum_{i=1}^{p} \beta_p h_{t-p}}_{\text{GARCH-terms}} \tag{8}$$

With:

$$p \geq 0, q \geq 0, \alpha_0 > 0, \quad \alpha_i \geq 0 \quad i = 1, ..., q \quad \beta_i \geq \quad i = 1, ..., 0 \quad (\alpha_i + \beta_i) \leq 1$$

Where ϵ_t is a real valued discrete-time stochastic process, Ψ_t the information set (σ-field) up to time t, h_t is the conditional variance, y_t dependent variable, μ is the mean vector, r_t is a vector of log returns, β_t model parameter for lag t and q, p lag orders.

In this thesis univariate GARCH(1,1)-models will be used, however for the sake of writing with general terminology in both the univariate and multivariate cases, the GARCH models are defined as GARCH(p, q).

3.2 Multivariate GARCH

The univariate GARCH-model can only be used on one asset at a time. The extension from a univariate model to a multivariate model allows for multiple assets to be modelled together. The multivariate GARCH (MGARCH) models follow the same structure of the univariate GARCH model, however it makes it possible to make joint analysis and forecasts.

If we consider a stochastic vector $\mathbf{r}_t$ with dimension $n \times 1$, with $E[\mathbf{r}_t] = 0$ and let Ψ_{t-1} denote the information set generated by the observed series of $\mathbf{r}_t$ up to time $t-1$ and assume that r_t is conditionally heteroskedastic we get [18]:

$$\mathbf{y}_t \mid \Psi_{t-1} \sim \mathbf{H}_t^{1/2} \epsilon_t \tag{9}$$

$$\mathbf{r}_t = \mathbf{y}_t + \mu_t \tag{10}$$

Where $\mathbf{r}_t$ log return at time t of n assets, $n \times 1$-vector. μ_t is the expected value of the conditional $\mathbf{y}_t$ with n assets at time t ($n \times 1$-vector). $\mathbf{y}_t$ mean corrected return at time t ($n \times 1$-vector). $\mathbf{H}_t$ is the conditional covariance ($n \times n$-matrix) of ϵ_t and $\mathbf{y}_t$. ϵ_t is an IID error vector process s.t. $E[\epsilon_t \epsilon_t'] = I, E[\epsilon_t] = 0$. $\mathbf{H}_t^{1/2}$ can be obtained from $\mathbf{H}_t$ by using Cholesky decomposition.

There are many ways of computing $\mathbf{H}_t$. Different methods of computing the conditional covariance matrix will be discussed in the next section. μ_t can be modelled as a function of time or as a constant. In this thesis the modelling of μ_t will not be in focus and will be modelled as a constant vector.

3.2.1 Introduction to multivariate GARCH

Silvennoinen and Teräasvirta [18] divided the multivariate GARCH models into four different types of models. In this section these different models will be discussed, which are different approaches of computing $\mathbf{H}_t$ in (9). A problem with MGARCH-models is to make them parsimonious enough while maintaining flexibility, as the number of parameters can grow rapidly with the increase of the number of assets n. Further it is important to ensure the positive definiteness of matrices used in the MGARCH-models. This is usually done through eigenvalue-eigenvalue decomposition, which can be difficult in large systems. It is therefore of interest to ensure that MGARCH-models avoid excessive inversion of matrices.

- **Models of the conditional covariance matrix:** The VEC-model was introduced by Engle, Wooldridge and Bollerslev [4] in 1988. It is a straightforward generalization of the univariate GARCH model. The conditional variance and covariance is a function of lagged conditional variance and covariances.

$$\mathrm{VECH}(\mathbf{H}_t) = \mathbf{c} + \sum_{j=1}^{q} \mathbf{A}_j \mathrm{VECH}(\mathbf{r}_{t-j}\mathbf{r}'_{t-j}) + \sum_{j=1}^{p} \mathbf{B}_j \mathrm{VECH}(\mathbf{H}_{t-1}) \quad (11)$$

  Where VECH$(\cdot)$ is a matrix operator, $\mathbf{c}$ is a $n(n+1)/2 \times 1$ vector, and $\mathbf{A}_j$ and $\mathbf{B}_j$ are parameter matrices. However, the computation of the VEC-model is unpleasant since conditions has to be made in order to make $\mathbf{H}_t$ positive definite. A VECH-model is also very computationally demanding. As model that was developed as an improvement of the original VEC-model is the BEKK-model, which has the property that $\mathbf{H}_t$ is positive definite by definition. The BEKK-model was created by Engle and Kroner in 1995 [8]:

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \sum_{j=1}^{q}\sum_{k=1}^{K} \mathbf{A}'_{kj}\mathbf{r}_{t-j}\mathbf{r}'_{t-j}\mathbf{A}'_{kj} + \sum_{j=1}^{p}\sum_{k=1}^{K} \mathbf{B}'_{kj}\mathbf{H}_{t-j}\mathbf{B}'_{kj} \quad (12)$$

  Where $\mathbf{B}_{kj}$, $\mathbf{B}_{kj}$, $\mathbf{C}$ are parameter matrices.

- **Factor models:** Factor models are motivated by pricing theory. The first factor GARCH-model was created by Engle, Ng, Rothschild in 1990 [11]. In this model it is assumed that the underlying factors are conditionally heteroscedastic and has the attributes the be modelled as GARCH-processes:

$$\mathbf{H}_t = \mathbf{\Omega} + \sum_{k=1}^{K} \mathbf{w}_k \mathbf{w}'_k \mathbf{f}'_{k,t} \quad (13)$$

Where $\mathbf{\Omega}$ is a n×n positive, semi definite matrix, $\mathbf{w}_k$ are the portfolio weights and $\mathbf{f}_{k,t}$ factors. One of the most widely used factor GARCH-models is the GO-GARCH model, which was introduced by Van der Weide in 2002 [20]:

$$\mathbf{H}_t = \mathbf{W}\mathbf{H}_t^z\mathbf{W} = \sum_{k=1}^{N} \mathbf{w}_k \mathbf{w}_k' h_{k,t}^z \tag{14}$$

Where $\mathbf{w}_k$ are the columns in $\mathbf{W}$ and $h_{k,t}^z$ are the diagonal elements of $\mathbf{H}_t^z$. The main difference between (13) and (14) is that the factors in (14) are uncorrelated.

- **Non-parametric and semi-parametric approaches:** This class of GARCH-models that form an alternative to parametric estimation of the conditional covariance structure. These models has the advantage of not imposing a particular structure on the data. Long and Ullah [15] developed an approach where after a parametric model is estimated, the standardized residuals $\hat{\eta}_t$ are extracted. If the model is not correctly specified, use of nonparametric estimation is applied to extract the information that is not correctly specified:

$$\mathbf{H}_t = \hat{H}_t^{1/2} \frac{\sum_{r=1}^{T} \hat{\eta}_t \hat{\eta}_t' K_h(s_t - s_{t-1})}{\sum_{t=1}^{T} K_h(s_t - s_{t-1})} \hat{H}_t^{1/2} \tag{15}$$

  Where, $\mathbf{H}_t$ is the conditional covariance matrix, estimated from an MGARCH-model and $s_t \in \Psi_{t-1}$ is the observed variable on the information set Ψ_{t-1} and $K_h(\cdot) = K(\cdot/h)/h$ where $K(\cdot)$ is a kernel function and h is the bandwidth parameter.

- **Models of the conditional variances and correlations:** This type of MGARCH-models are based on the decomposition of the conditional covariance matrix into conditional standard deviations and correlations. It is within this class that the CCC-model and DCC-model belongs. Since the DCC-model is what this thesis aims to evaluate, it will be discussed further in the upcoming section together with the CCC-model.

3.2.2 CCC-GARCH

The CCC-model was developed by Bollerslev in 1990 as a multivariate time series model with time varying conditional variances and covariances, but with constant correlation [3]. As the conditional correlation matrix is not time dependent, the computation of the conditional covariance matrix $\mathbf{H}_t$ is computed as follows:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t \tag{16}$$

Where $\mathbf{R}$ is the constant correlation matrix with elements $\mathbf{R} = \rho_{ij}$, $\rho_{ii} = 1$ for $i = 1, .., n$ with n number of assets. $\mathbf{D}_t$ is the diagonal matrix containing $h_{nt}^{1/2}$ of the fitted univariate GARCH-models for each asset, s.t $\mathbf{D}_t = \text{diag}(h_{1t}^{1/2}, ..., h_{nt}^{1/2})$. The elements that are off the diagonals of $\mathbf{H}_t$ then becomes:

$$[\mathbf{H}_t]_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij}, \quad i \neq j \tag{17}$$

Where $i \geq 1, j \geq 1$. Since the processes r_{it} are modelled as univariate GARCH(p, q)-models, the conditional variances can be written on vector form as following:

$$\mathbf{h}_t = \omega_t + \sum_{j=1}^{q} \mathbf{A}_j \mathbf{r}_{t-j}^{*} + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{h}_{t-j} \tag{18}$$

Where $r_{t-j}^{*} = r_t \odot r_t$ and $\odot$ is the Hadamard product. ω_t is a $n \times 1$ vector, $\mathbf{A}_j$ and $\mathbf{B}_j$ are diagonal $n \times n$ matrices. For $\mathbf{H}_t$ to be positive definite, $\mathbf{R}$ has to be positive definite and the diagonal elements of $\mathbf{A}_j$ and $\mathbf{B}_j$ has to be positive. The requirement of $\mathbf{A}_j$ and $\mathbf{B}_j$ to be positive is not necessary unless $p = q = 1$. This is further discussed in Nelson and Cao [16].

Even though the CCC-GARCH model has an attractive parametrization method, empirical studies has implied that assumption of constant correlation is too trivial. Applying constant correlation is unrealistic in practise. To give an example, J.Chevallier mentions this in [6] (2011). This leads us to the DCC-GARCH model.

3.2.3 DCC-GARCH

The Dynamic Conditional Correlation-model (DCC-GARCH) was introduced by Engle and Sheppard in 2001 as an extension of the CCC-GARCH model [7]. The DCC-model use (16), but instead of modelling $\mathbf{R}$ as a constant matrix, it is modelled in a dynamic fashion with $\mathbf{R}_t$ depending on the time t s.t:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \tag{19}$$

The equation in (17) then becomes:

$$[\mathbf{H}_t]_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij,t} \tag{20}$$

$\mathbf{D}_t$ in (19) is on the form:

$$\mathbf{D}_t = \begin{pmatrix} h_{1,t}^{1/2} & 0 & \cdots & 0 \\ 0 & h_{2,t}^{1/2} & \ddots & \vdots \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{n,t}^{1/2} \end{pmatrix} \tag{21}$$

With $h_{it} = \omega_{it} + \sum_{j=1}^{q_i} \alpha_{ij} r_{it-j}^2 + \sum_{j=1}^{p_i} \beta_{ij} h_{i,t-j}$. In models that are defined as (19), positive definiteness of $\mathbf{H}_t$ is ensured if, not only the conditional variances $h_{it}, i = 1, ..., n$ are well defined, but that the matrix $\mathbf{R}_t$ is positive definite for every t. This fact makes the DCC-model more computationally demanding than the CCC-model, as the conditional correlation matrix has to be inverted with each t during every iteration. The DCC-model which was introduced by Engle has a dynamic conditional correlation structure on the form:

$$\mathbf{Q}_t = (1 - \alpha - \beta)\mathbf{S} + \alpha \epsilon_{t-1} \epsilon_{t-1}' + \beta \mathbf{Q}_{t-1} \tag{22}$$

$$\mathbf{R}_t = (\mathbf{I} \odot \mathbf{Q}_t^{1/2}) \mathbf{Q}_t (\mathbf{I} \odot \mathbf{Q}_t^{1/2}) \tag{23}$$

Where in (22) $\mathbf{Q}_t$ is a covariance matrix, $\mathbf{S}_t$ is the unconditional correlation matrix of the epsilons ϵ_t, s.t. $\mathbf{S}_t = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t \epsilon_t'$, since $\bar{Q} = \mathrm{Cov}[\epsilon_t \epsilon_t'] = \mathrm{E}[\epsilon_t \epsilon_t']$. α and β are the DCC-garch parameters s.t. $\alpha > 0, \quad \beta > 0, \quad \alpha + \beta < 1$. This process ensures positive definiteness, but not necessarily producing valid correlation matrices. The correlation matrices are obtained through the re scaling in (23), with $[R_t]_{i,j} = \rho_{i,j,t}, \quad \rho_{i,j,t} \leq 1$ and $\rho_{i,i,t} = 1$. To clarify, $(\mathbf{I} \odot \mathbf{Q}_t^{1/2})$ is computed as:

$$(\mathbf{I} \odot \mathbf{Q}_t^{1/2}) = \begin{pmatrix} \sqrt{q_{1,t}} & 0 & \cdots & 0 \\ 0 & \sqrt{q_{2,t}} & \ddots & \vdots \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{q_{n,t}} \end{pmatrix} \tag{24}$$

The rescaling of $\mathbf{Q}_t$ ensures that $\rho_{i,j,t} \leq 1$. Further, $\mathbf{Q}_t$ has to be positive definite for $\mathbf{R}_t$ to be positive definite. Finally, we get $\mathbf{r}_t$ on the following form:

$$\mathbf{r}_t \mid \Psi_{t-1} \sim N(0, \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t) = N(0, \mathbf{H}_t) \tag{25}$$

The assumption of normality gives rise to a Likelihood estimation. The normality aspect is used even when computing r_t with student's t distribution errors, which is part of the 2-stage Quasi Maximum Likelihood function. Which will be described in the next section.

4 Forecasting Procedure and Evaluation

Engle and Sheppard suggested a two stage estimation method of the parameters in the DCC model [12]. The procedure is to divide the quasi likelihood estimator into two parts, one volatility part and one correlation part. It is a two stage method, since the volatility part is used to calculate the correlation part. The first stage of the calculation, the volatility part, is the same regardless of the underlying innovation distribution. Several authors has shown that changing the underlying univariate distributions for the volatility part does not change the the outcome of the parameters in the second stage estimation, see e.g. [21]. The two stage estimation process of the quasi likelihood function for Gaussian distributed innovations as well as multivariate student's t distributed innovations will be discussed in this chapter.

4.1 Gaussian distributed innovations

First we define the multivariate Gaussian distribution according to [5]:

$$f(\epsilon_t) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{\frac{2}{n}}} e^{-\frac{1}{2}\epsilon_t'\epsilon_t}$$

A return vector of size n at time t, $\mathbf{r}_t$, with mean μ and conditional covariance matrix $\mathbf{H}_t$ as defined in (9) to (10), follows a multivariate normal distribution if $\mathbf{r}_t \sim MN(\mu, \mathbf{H}_t)$. The linear transformation and scaling possibilities of this distribution, makes the errors $\mathbf{r}_t - \mu = \epsilon_t \sim MN(0, \mathbf{H}_t)$ also Guassian distributed. The scaled errors $\mathbf{H}_t^{-1/2}\epsilon_t = z_t \sim MN(0, \mathbf{I}_N)$ also follows this distribution. The likelihood function at time t of the errors is given by:

$$f(\epsilon_t \mid \theta) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{\frac{2}{n}} |\mathbf{H}_t|^{\frac{1}{2}}} e^{-\frac{1}{2}\epsilon_t'\mathbf{H}_t^{-1}\epsilon_t} \tag{26}$$

Where $|\mathbf{H}_t|$ is the determinant. Given (26) the log likelihood estimator can then be expressed as:

$$
\begin{aligned}
L &= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + \log|\mathbf{H}_t| + \mathbf{r}_t'\mathbf{H}_t^{-1}\mathbf{r}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + \log|\mathbf{R}_t\mathbf{D}_t\mathbf{R}_t| + \mathbf{r}_t'\mathbf{D}_t^{-1}\mathbf{R}_t^{-1}\mathbf{D}_t^{-1}\mathbf{r}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + 2\log|\mathbf{D}_t| + \log|\mathbf{R}_t| + \epsilon_t'\mathbf{R}_t\epsilon_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + 2\log|\mathbf{D}_t| + \mathbf{r}_t'\mathbf{D}_t^{-1}\mathbf{D}_t^{-1}\mathbf{r}_t' \\
&\quad - \epsilon_t'\epsilon_t\log|\mathbf{R}_t| + \epsilon_t'\mathbf{R}_t\epsilon_t)
\end{aligned} \tag{27}
$$

If we let the parameters in $\mathbf{D}_t$ be denoted as θ and the parameters in $\mathbf{R}_t$ to be denoted as ϕ, the log likelihood function can be written as the sum of a volatility part and a correlation part:

$$
L(\theta,\phi) = L_v(\theta) + L_c(\theta,\phi) \tag{28}
$$

Where:

$$
\begin{aligned}
L_v(\theta) &= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + \log|\mathbf{D}_t|^2 + \mathbf{r}_t'\mathbf{D}_t^{-2}\mathbf{r}_t') \\
L_c(\theta,\phi) &= -\frac{1}{2}\sum_{t=1}^{T}(\log|\mathbf{R}_t| + \epsilon_t'\mathbf{R}_t^{-1}\epsilon_t - \epsilon_t'\epsilon_t)
\end{aligned} \tag{29}
$$

Further, the volatility part $L_v(\theta)$ is the sum of the univariate GARCH likelihoods:

$$
L_v(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}(n\log(2\pi) + \log(h_{i,t} + \frac{r_{i,t}^2}{h_{i,t}})) \tag{30}
$$

To summarize, the two step procedure to maximize the likelihood is to find:

$$
\begin{aligned}
&\text{Step 1: } \hat{\theta} = \max\{L_v(\theta)\} \\
&\text{Step 2: } \hat{\phi} = \max\{L_c(\hat{\theta},\phi)\}
\end{aligned} \tag{31}
$$

Which is the procedure to estimate Gaussian distribution innovations.

4.2 Multivariate Student's t distributed innovations

As mentioned earlier, $L_v(\theta)$ is the same for multivariate student's t distribution errors as for the Gaussian distribution errors. However, the second step of the likelihood estimation procedure is different in the multivariate student's t distribution case. Hence we need to specify $L_c(\theta, \phi)$ for this application. First, we define the multivariate student's t distribution:

$$f(\epsilon_t \mid \nu) = \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})[\pi(\nu-2)]^{n/2}}[1 + \frac{\epsilon_t' \epsilon_t}{\nu - 2}]^{-\frac{n+\nu}{2}}$$

Where $\Gamma(\bullet)$ is the Gamma-function. The multivariate student's t distribution consists of one extra parameter if it is compared to the Gaussian distribution, the shaping parameter ν. This parameter exists in the multivariate student's t distribution as it is a mixture of two other distributions, the multivariate gamma distribution and the multivariate normal distribution. As a consequence, the multivariate student's t distribution has a symmetrical tail dependence just like the multivariate normal distribution. Further, if $\mathbf{r}_t$, with mean μ, conditional covariance matrix $\mathbf{H}_t$ and shape parameter ν follows the multivariate student's t distribution, then $\mathbf{r}_t \sim MT(\mu, \mathbf{\Omega}_t, \nu)$. Where $\mathbf{\Omega}$ is a scaling matrix such that $\mathbf{H}_t = \frac{\nu}{\nu-2}\Omega$. Finally, we get that the likelihood at time t of the errors is given by:

$$f(\epsilon_t \mid \nu) = \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})[\pi(\nu-2)]^{n/2}|\mathbf{\Omega}|^{\frac{1}{2}}}[1 + \frac{\epsilon_t'|\mathbf{\Omega}^{-1}|\epsilon_t}{\nu - 2}]^{-\frac{n+\nu}{2}} \tag{32}$$

Given (32) we formulate the log likelihood estimator as:

$$\begin{aligned} L = \sum_{t=1}^{T} &(\log(\Gamma(\frac{\nu+n}{2}) - \log(\Gamma(\frac{\nu}{2})) - \frac{n}{2}\log(\pi(\nu-2)) \\ &- \frac{1}{2}\log(|\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t|) - \frac{\nu+2}{2}\log(1 + \frac{\epsilon_t'\mathbf{D}_t^{-1}\mathbf{R}_t^{-1}\mathbf{D}_t^{-1}\epsilon_t}{\nu-2}) \end{aligned} \tag{33}$$

Just like in the Gaussian case, we want to make a two step estimation procedure. Step 1 is identical to the Gaussian parameter estimation, and step 2 includes the extra parameter that needs to be estimated ν.

$$\begin{aligned} &\text{Step 1: } \hat{\theta} = \max\{L_v(\theta)\} \\ &\text{Step 2: } \hat{\phi}, \hat{\nu} = \max\{L_c(\hat{\theta}, \phi, \nu)\} \end{aligned} \tag{34}$$

Since the variance estimation of $\max\{L_v(\theta)\}$ already determined, we can treat $\mathbf{D}_t$ in (33) as a constant term and exclude it from the log likelihood estimator. As such, we get:

$$\begin{aligned} L_c(\hat{\theta}, \phi, \nu) = \sum_{t=1}^{T} &(\log(\Gamma(\frac{\nu+n}{2}) - \log(\Gamma(\frac{\nu}{2})) - \frac{n}{2}\log(\pi(\nu-2)) \\ &- \frac{1}{2}\log(|\mathbf{R}_t|) - \frac{\nu+2}{2}\log(1+\frac{\epsilon_t'\mathbf{R}_t^{-1}\epsilon_t}{\nu-2})) \end{aligned} \tag{35}$$

Which is the second stage estimation with assumed multivariate student's t distribution innovations.

4.3 Forecasting DCC-GARCH

Forecasting variances is a requirement of GARCH-models. Most GARCH models provide a pleasant way of generating a k-step ahead forecast. If we look at the univariate GARCH(1,1) process, the k-step ahead forecast is [1]:

$$h_{t+k} = \sum_{i=1}^{k-2} \omega(\alpha+\beta)^i + (\alpha+\beta)^{k-1} h_{t+1} \tag{36}$$

However, it is not as simple when it comes to the DCC-model, as it is non-linear:

$$\mathbf{Q}_{t+k} = (1-\alpha-\beta)\mathbf{S} + \alpha\epsilon_{t+k-1}\epsilon_{t+k-1}' + \beta\mathbf{Q}_{t+k-1} \tag{37}$$

Where $E_t[\epsilon_{t+k-1}\epsilon_{t+k-1}'] = E_t[\mathbf{R}_{t+k-1}]$ and $R_{t+k} = (\mathbf{I} \odot \mathbf{Q}_{t+k}^{1/2})\mathbf{Q}_{t+k}(\mathbf{I} \odot \mathbf{Q}_{t+k}^{1/2})$. Therefore, the k-step ahead forecast cannot be directly computed. Engle and Sheppard suggested two different approaches of forecasting the parameters of the DCC-garch model [9]. The first technique makes the assumption that $E_t[\epsilon_{t+1}\epsilon_{t+1}'] \approx \mathbf{Q}_{t+1}$, for $i \in [1, ..., k]$. Using this approximation, the k-step ahead forecast of $\mathbf{Q}_t$ becomes:

$$E_t[Q_{t+k}] = \sum_{i=0}^{r-2} (1-\alpha-\beta)\mathbf{S}(\alpha-\beta)^i + (\alpha+\beta)^{k-1}\mathbf{Q}_{t+1} \tag{38}$$

Where $\mathbf{R}_{t+k}$ still is defined as $\mathbf{R}_{t+k} = (\mathbf{I} \odot \mathbf{Q}_{t+k}^{1/2})\mathbf{Q}_{t+k}(\mathbf{I} \odot \mathbf{Q}_{t+k}^{1/2})$. Another approximation to let $\mathbf{S} \approx \bar{R}$ and $E_t[\mathbf{Q}_{t+1}] \approx E_t[\mathbf{R}_{t+1}]$. Under this latter approximation, $\mathbf{R}_{t+k}$ can be forecasted directly:

$$E_t[R_{t+k}] = \sum_{i=0}^{r-2} (1-\alpha-\beta)\bar{R}(\alpha-\beta)^i + (\alpha+\beta)^{k-1}\mathbf{R}_{t+1} \tag{39}$$

In this thesis the first approach will be applied, where $E[\epsilon_{t+1}\epsilon_{t+1}'] \approx \mathbf{Q}_{t+1}$.

Earlier we defined the conditional covariance matrix of the DCC model as $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$, where $\mathbf{D}_t = \mathrm{diag}(h_{1t}^{1/2}, ..., h_{nt}^{1/2})$. The k-step ahead forecast of $\mathbf{H}_t$ then becomes:

$$\mathbf{H}_{t+k} = \mathbf{D}_{t+k}\mathbf{R}_{t+k}\mathbf{D}_{t+k} \tag{40}$$

Where $\mathbf{D}_{t+k}$ is a diagonal matrix consisting of the underlying univariate forecasts described in (36) and $\mathbf{R}_{t+k}$ is computed as described in (38).

5 Evaluation of estimations

In order to evaluate the goodness of fit for the model the marginal and multivariate distributions has to be calculated separately. The different methods that has been applied will be discussed in this section.

5.1 Goodness of fit for univariate distributions

Visual error evaluation

This is a simple way of getting a good understanding of the errors. A plot of the errors should look random if they are IID.

The Auto Correlation Function

The autocorrelation function for a stochastic process X is defined as [17]:

$$R_X(t) = Cov[X(s), X(s+t)]$$

Given n observations, at least 5% of the lags of the ACF should be outside the 95% confidence interval for ϵ_t to be IID.

Ljung-Box test

The Ljung-Box test statistic evaluates if the data are autocorrelated based on a number of lags h. The test is defines as:

$$\begin{aligned} H_0 &= \gamma_1 = 0 = ... = \gamma_h \\ H_1 &= \gamma_i \neq 0, \quad \text{for } i = 1, ..., h \end{aligned} \tag{41}$$

The Ljung-Box test statistic is defined as [14]:

$$Q = n(n+2) \sum_{t=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

Where n is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k, and h is the number of lags being tested. Q follows under H_0 asymptotically the χ^2_h-distribution. For significance level α, the critical region for rejection of the hypothesis is:

$$Q > \chi^2_{1-\alpha,h}$$

Where $\chi^2_{1-\alpha,h}$ is the $1-\alpha$ quantile oh the χ^2-distribution with h degrees of freedom.

5.2 Goodness of fit for multivariate distributions

It is possible that the univariate distributions are well defined even tho the multivariate distributions are not. In this section two tests for back testing of Value at Risk will be discussed as a way to determine if the tails of a portfolio follows a specified distribution.

Value at Risk

First, let us define Value at Risk (VaR). VaR is a measurement that estimates how much a set of investments might lose given a certain probability during a given time period. Let's call this probability level α and the corresponding loss limit l. As such, l is the VaR at confidence level α.

To apply VaR into a multivariate context, the properties of the underlying portfolio distribution has to be know. In section 3.1 the linearity of the multivariate normal distribution was discussed. The errors ϵ_t were defined as $\epsilon_t \sim MN(0, \mathbf{H}_t)$. For portfolio returns $\mathbf{R}_p$ we get $\mathbf{R}_{p,t} = \mathbf{w}\epsilon_t + \mathbf{w}\mu_t$, where $\mathbf{r}_t$ is the returns and ϵ_t is the corresponding Gaussian errors, we get:

$$\mathbf{R}_{p,t,MN} \sim MN(\mathbf{w}\mu_t, \mathbf{w}'\mathbf{H}_t\mathbf{w}) \tag{42}$$

Similarly, the errors of the multivariate student's t distribution was defined in 3.2 as $\mathbf{r}_t \sim MT(\mu_t, \mathbf{\Omega}_t, \nu)$. We define the portfolio returns $\mathbf{R}_{p,t}$ for the multivariate student's t distribution as:

$$\mathbf{R}_{p,t,MT} \sim MT(\mathbf{w}\mu, \mathbf{w}'\mathbf{H}_t\mathbf{w}, \nu) \tag{43}$$

Where $\mathbf{w}$ will be addressed as equally distributed among the currency pairs.

Kupiec test

The Kupiec test evaluates if the number of failures for a model corresponds to a certain confidence level α [13]. H_0 simply states that we have the expected number of violations given α, and H_α is that the number of violations does not meet the expectations. Further, failures follows a binominal distribution.

The likelihood ratio in the Kupiec test is:

$$L_{KT} = -2log((1-p)^{N-x}p^x) + 2log((1-x)^{N-x}(\frac{x}{N})^x)$$

If H_0 is correct, the test statistic follows asymptotically a χ^2-distribution. If we observe large differences to the expected numbers of violations, the model is not well specified [19].

Christoffersen test

The Christoffersen test evaluates whether the probability of observing an exception on a particular day depends on whether an exception occurred. Unlike the unconditional probability of observing an exception, Christoffersen's test measures the dependency between consecutive days only. Christoffersen's test therefore follows Markov properties. The hypothesis is defined as:

$$H_0 = \text{Failures are IID}$$
$$H_\alpha = \text{Failures are not IID}$$

With the corresponding test statistic:

$$L_{CT} = 2ln(\frac{(1-\pi_{01}^{n_{00}})\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}}{\alpha^x(1-\alpha)^{n-x}})$$

In the test n_{ij} is a binary variable that evaluates failures/non failures.

6 Empirical Application

In order to test the DCC-model, an application to real data was necessary. The chosen data set consists of the following exchange rates:

$$\mathbf{Y}_{t,i} = \{\text{USD/EUR,USD/GBP,USD/JPY}\}$$

Where USD = American Dollar, EUR = Euro, GBP = Great British Pound, JPY = Japanese Yen. The time horizon consists of weekly trading data of opening prizes with start 2005,06,19 to 2021,06,13. This gives the time horizon of $\mathbf{t} = 1, ..., 834$, where $\mathbf{t} = 1$ represents 2005,06,19 and $\mathbf{t} = 834$ represents 2021,06,13. The reason why weekly data was chosen over daily data, is because the number of trading days were not consistent between the currency pairs over this time period. Further the dollar was an arbitrary choice of a reference currency. We define the log returns as:

$$\mathbf{r}_{t,i} = \log\mathbf{Y}_{t,i} - \log\mathbf{Y}_{t-1,i}$$

As stated in (10), $\mathbf{r}_t = \mathbf{y}_t + \mu_t$ where $\mathbf{r}_t$ is the log returns, $\mathbf{y}_t$ are the mean corrected returns and μ_t is the mean vector. Here $\text{E}[\mathbf{y}_t] = 0$, $\text{Cov}[\mathbf{y}_t] = \mathbf{H}_t$

6.1 Data evaluation

By looking at the data in below, it is obvious that each series has volatility clustering. In other words, we see that large changes tends to be followed by large changes (as in late 2008 and early 2020) and small changes are followed by small changes. Given the heteroscedastic properties of the series it indicates that GARCH-models are appropriate to use for this data set.

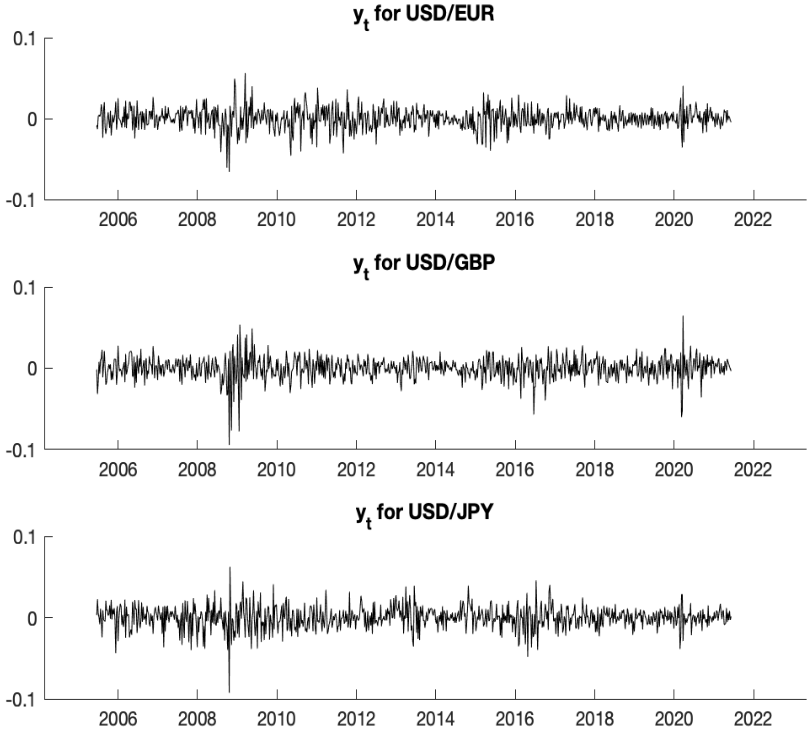

Figure 1: For plots for the mean corrected returns, $\mathbf{y}_t$

Autocorrelation funktion of $\mathbf{y}_t^2$ and $\mathbf{y}_t$

By looking at the autocorrelation function of $\mathbf{y}_t^2$, we see that more than 5% of the lags fall outside the given limit (red lines) and that there is an apparent pattern where the lags are great in the beginning but falls of. Hence $\mathbf{y}_t^2$ is not serially uncorrelated, and as a consequence that means that $\mathbf{y}_t$ is dependent. If we look at the ACF of $\mathbf{y}_t$ we see that less than 5% of the lags fall outside the limit. Hence $\mathbf{y}_t$ is serially uncorrelated. It can be determined that a GARCH process is a good choice for the data set since $\mathbf{y}_t$ is serially uncorrelated, but dependent. This is true for all of the currency pairs.

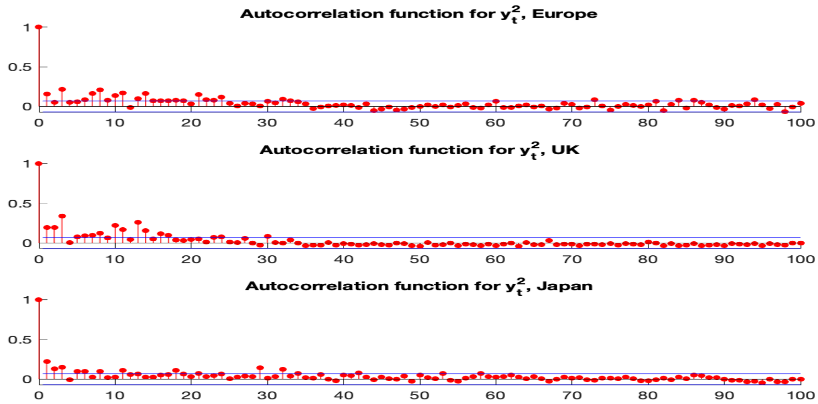

Figure 2: For plots for the ACF, $\mathbf{y}_t^2$

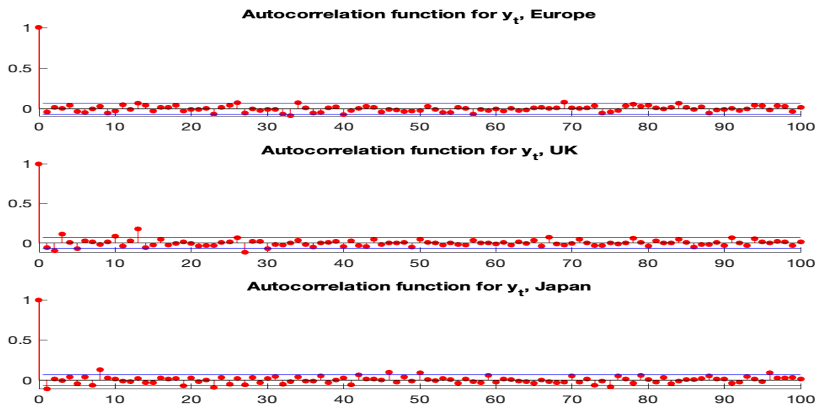

Figure 3: For plots for the ACF, $\mathbf{y}_t$

6.2 Estimations

In this section the estimated parameters of the DCC-model will be presented for the currency pairs $\mathbf{Y}_{t,i}$ = {USD/EUR,USD/GBP,USD,JPY}. For the underlying univariate models, they are all $p = q = 1$.

First stage parameters

The first stage estimation parameters, estimated from $\max\{L_v(\theta)\}$ are the same for both the Gaussian distribution and multivariate student's t distribution. Since both assume normally distributed univariate GARCH-models. The estimated parameters from the first stage are:

Estimated coefficients - First stage			
Currency pair	α_0	α_1	β_1
USD/EUR	1.748365e-07	0.05088643	0.9001501
USD/GBP	1.955484e-07	0.05104649	0.9001768
USD/JPY	2.131842e-07	0.06818003	0.9087671

Second stage parameters

As explained in section 3.1, the second stage estimation is different for the Gaussian and multivariate student's t distribution. For the Gaussian distributed, $\max\{L_c(\hat{\theta}, \phi)\}$ was estimated and for the multivariate student's t distribution, $\max\hat{\phi} = \max\{L_c(\hat{\theta}, \phi, \nu)\}$ was estimated. The second stage parameters are:

Estimated coefficients - Second stage			
Distribution	α_1	β_1	ν
Gaussian	0.04804982	0.8218301	NA
MV Stud. t	0.0394801	0.8614414	9.120951

In the upcoming section, the forecasts of the second stage estimation will be presented. The forecast horizon is $k = 52$, which is a year given the weekly data structure.

6.3 DCC-forecasts

The forecast of $\mathbf{H}_{t+k}$ consists of different parts. As stated earlier, $\mathbf{H}_{t+k} = \mathbf{D}_{t+k}\mathbf{R}_{t+k}\mathbf{D}_{t+k}$, where $\mathbf{D}_{t+k}$ is a diagonal matrix with the univariate forecasts on the diagonal and $\mathbf{R}_{t+k}$ is the conditional correlation matrix.

6.3.1 D_t-forecasts

The $\mathbf{D}_{t+k}$ forecasts were computed based on the first stage of the quasi maximum likelihood estimations. As previously stated, the univariate normal forecasts are used as a base for the second stage estimations for both the Gaussian and multivariate student's t distribution. The black line represents $\mathbf{D}_t$, the red line represents $\mathbf{D}_{t+k}$ and the cyan line is the unconditional variance.

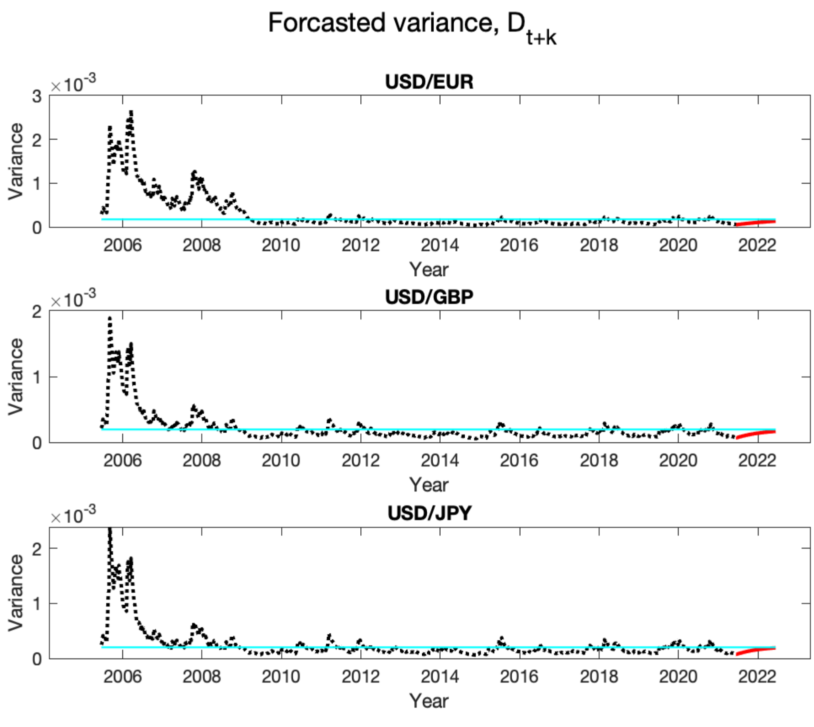

Figure 4: Plots for $\mathbf{D}_{t+k}$, Univariate Normal distribution

6.3.2 R_t-forecasts

The $\mathbf{R}_{t+k}$ forecasts were computed based on the second stage quasi maximum likelihood estimations, as explained in section 3.3. The black line represents $\mathbf{R}_t$, the red line represents $\mathbf{R}_{t+k}$ and the cyan line is the unconditional correlation.

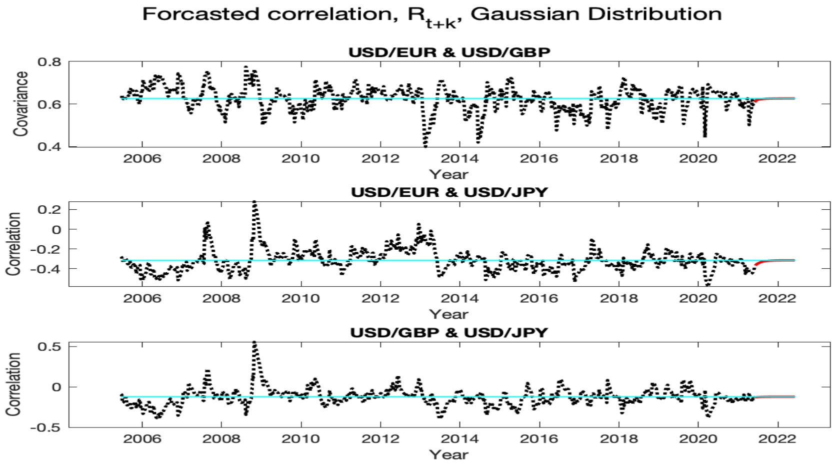

Figure 5: Plots for $\mathbf{R}_{t+k}$, Gaussian distribution

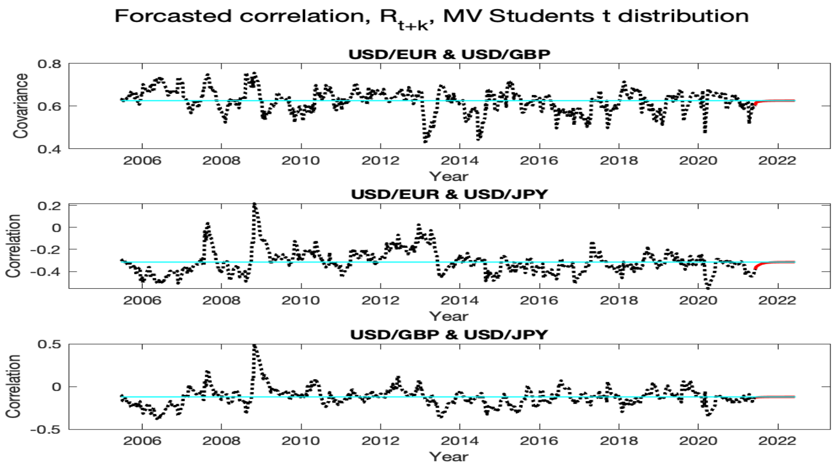

Figure 6: Plots for $\mathbf{R}_{t+k}$, MV Student's t distribution

6.3.3 H_t-forecasts

As $\mathbf{H}_{t+k}$ is the combination of the previously shown plots, those of $\mathbf{D}_{t+k}$ and $\mathbf{R}_{t+k}$. The black line represents $\mathbf{H}_t$, the red line represents $\mathbf{H}_{t+k}$ and the cyan line is the unconditional covariance.

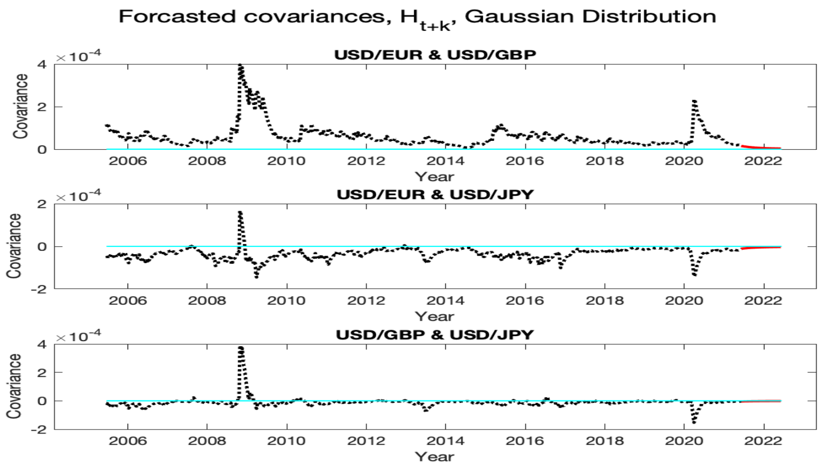

Figure 7: Plots for $\mathbf{H}_{t+k}$, Gaussian distribution

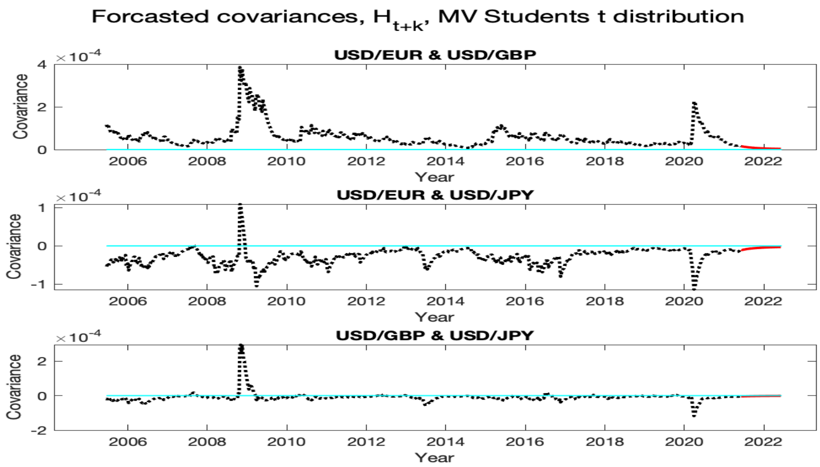

Figure 8: Plots for $\mathbf{H}_{t+k}$, MV Student's t distribution

7 Goodness of fit

7.1 Marginal distributions

7.1.1 Visual error evaluation

The errors are calculated from the conditional covariance matrix. The errors are calculated from $\epsilon_t = \mathbf{H}_t^{-1/2}\mathbf{y}_t$. It is not possible to draw any conclusions by looking at the error plots, since they are very similar for both distributions.

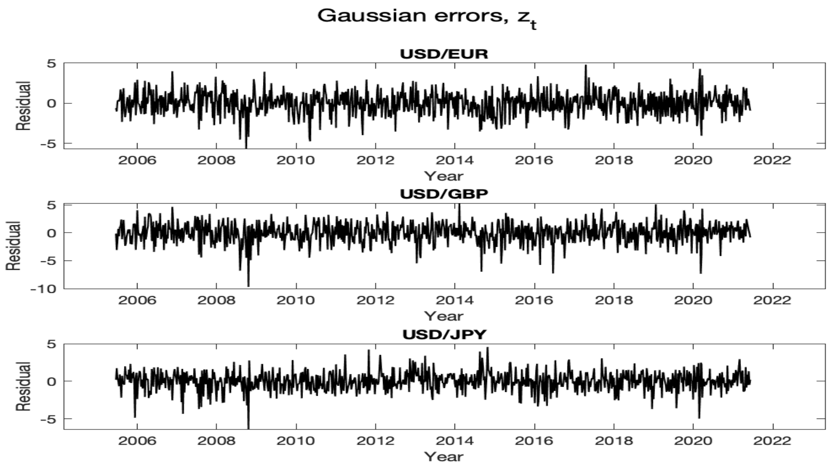

Figure 9: Plots for ϵ_t, Gaussian errors

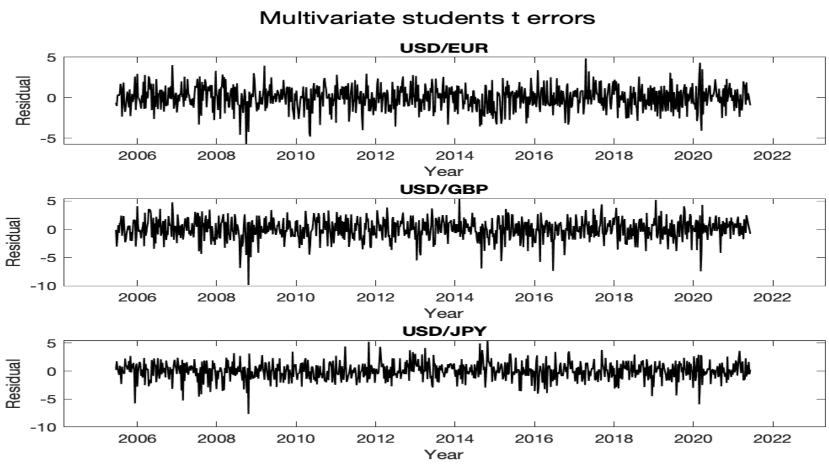

Figure 10: Plots for ϵ_t, MV Student's t errors

7.1.2 The auto correlation function

Below are the autocorrelation functions of the different error series ϵ_t. 5% of the lags are expected to fall outside the limits (the blue lines), but less than 5% are doing so. We can draw the conclusion that the errors for the two distributions are random at a 5%-level, for all of the series.

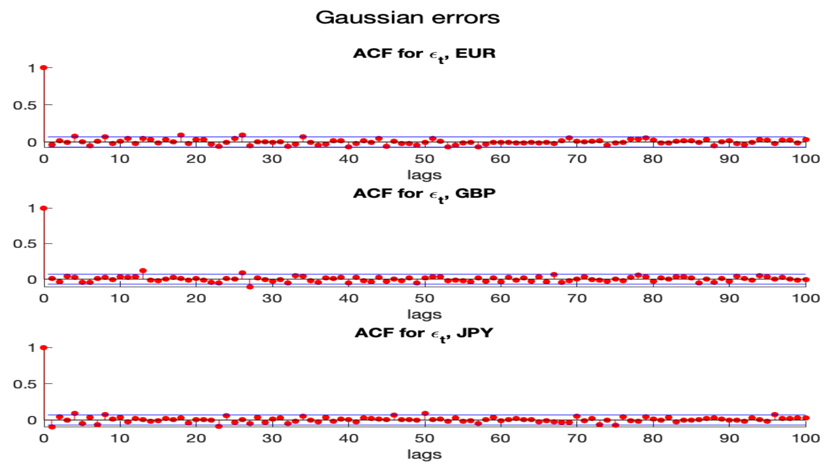

Figure 11: ACF for ϵ_t, Gaussian errors

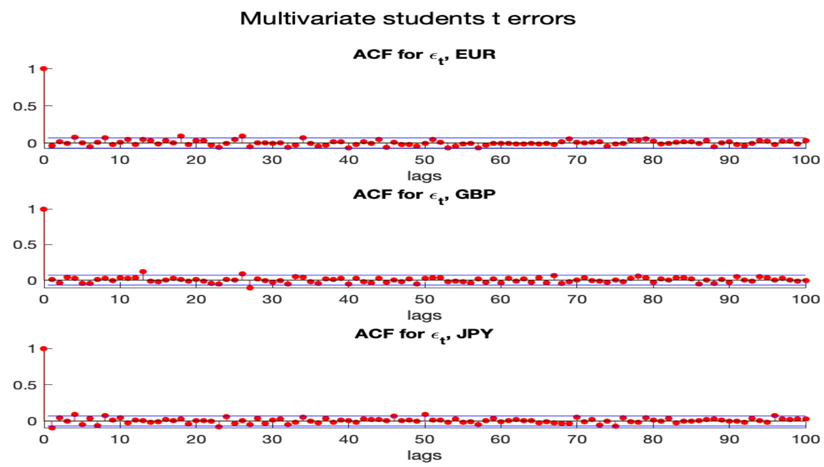

Figure 12: ACF for ϵ_t, MV Student's t errors

7.1.3 Ljung Box test

The plots below shows the test statistic of the Ljung Box test as a function of lag $= 1, .., 50$. The test statistic concludes that for all of the series, for both distributions, that for some lags they are correlated, and for some lags that they are uncorrelated. It is therefore not possible to draw any conclusion from this test. The red line represents the 5% level of the test statistic. A summary of the Ljung Box test can be found in Appendix A.

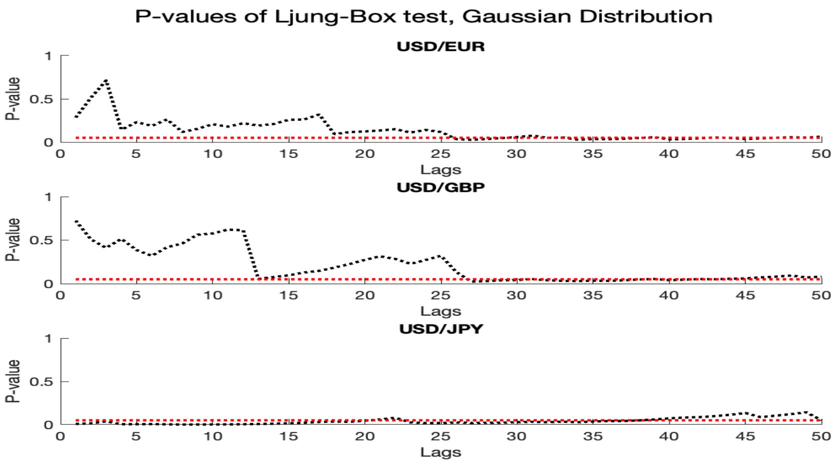

Figure 13: Ljung Box test, Gaussian errors

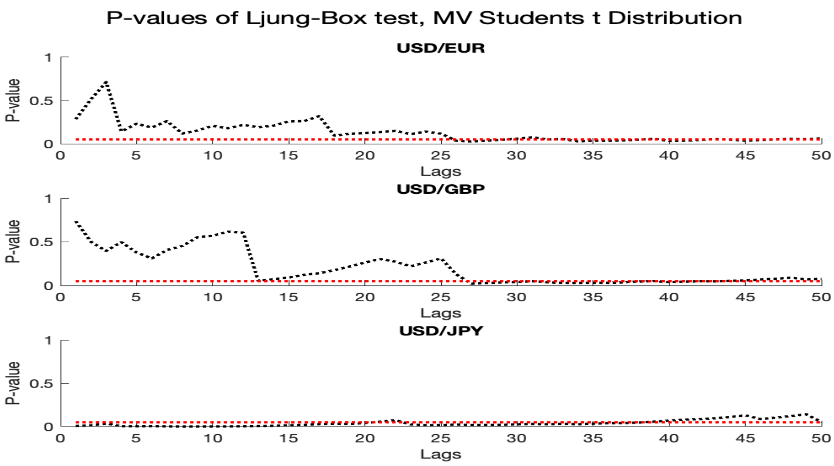

Figure 14: Ljung Box test, MV Student's t errors

7.2 Multivariate goodness of fit

This section will evaluate VaR-backtesting with the Kupiec test and Christoffersen markov test. Two test periods for VaR-backtesting has been arbitrary chosen. The fitting set is where the model trains according to the real data. The forecast period is the period that is forecasted basted on the fitted test set, which will be compared to the real data. The test periods are:

Backtesting of VaR		
Test	Fitting set	Forecast
1	1:499	500:700
2	1:599	600:800

With the corresponding DCC-GARCH parameters:

Backtesting of VaR, DCC-parameters				
Test	Distribution	α_1	β_1	ν
1	Gaussian	0.09303988	0.8146408	NA
1	MV Student t	0.06075544	0.899408	12.83456
2	Gaussian	0.1060632	0.8480775	NA
2	MV Student t	0.08379241	0.8847507	7.815305

7.2.1 Violations

Assuming a binominal distribution given the VaR-level, the expected number of violations becomes:

Backtesting of VaR - Violations							
Test	Distribution	1%	5%	10 %	90 %	95%	99%
1	Gaussian	1	8	22	12	5	3
1	MV Student t	1	9	21	13	5	3
2	Gaussian	0	7	16	20	13	4
2	MV Student t	0	7	16	21	15	5
Expected	Binominal	2	10	20	20	10	2

For the first test set, we see that the multivariate student's t distribution performs slightly better than the Gaussian distribution. For the second set, the Gaussian distribution performs slightly better than the MV Student's t distribution.

7.2.2 Kupiec test

The Kupiec test statistic were computed for the different test sets. At a 5% significance level, the Gaussian distribution rejected the test statistic two times and the multivariate student's t distribution rejects the test statistic one time. Suggesting the multivariate student's t distribution fits slightly better to the test sets.

Backtesting of VaR - Kupiec test statistics							
Test	Distribution	1%	5%	10 %	90 %	95%	99%
1	Gaussian	0.432	0.502	0.642	0.043	0.074	0.508
1	MV student t	0.432	0.742	0.815	0.080	0.074	0.508
2	Gaussian	0.045	0.305	0.330	1.000	0.351	0.211
2	MV student t	0.045	0.305	0.330	0.815	0.130	0.211

7.2.3 Christoffersen test

The Christoffersen test statistic were also computed for the different test sets. At a 5% significance level, both the Gaussian and the multivariate student's t distribution rejects the test statistic two times. Since both distributions perform equally for this test, it is not possible to draw any conclusions from it.

Backtesting of VaR - Christoffersen test statistics							
Test	Distribution	1%	5%	10 %	90 %	95%	99%
1	Gaussian	0.428	0.027	0.131	0.833	0.762	0.075
1	MV student t	0.428	0.027	0.004	0.660	0.540	0.075
2	Gaussian	0.428	0.298	0.616	0.794	0.761	0.022
2	MV student t	0.428	0.158	0.456	0.834	0.762	0.214

8 Discussion and conclusion

The purpose of this thesis was to evaluate how to fit a DCC-GARCH model and how to use it for forecasting. The were a few chosen approaches in this thesis which would affect the results if made differently. For instance, the mean vector μ_t was modelled as a constant vector but could have been modelled as an ARMA series. Also, as explained earlier, Engle and Sheppard proposed two methods of forecasting for $\mathbf{R}_t$. Only one of the two methods was considered in this thesis. It is also possible to make the same calculations with different lag orders of the univariate and multivariate GARCH models. All of these limitations were made to narrow down the scope of the thesis.

By looking at the results, all test series was shown to be random at a 5% level by looking at the auto correlation functions for the marginal distributions. The Ljung Box tests didn't give any proper explanation of whether or not the series were correlated or uncorrelated.

Given the multivariate evaluation of the tails for the two test sets, the multivariate student's t distribution showed a more accurate number of violations for the first test set. The Gaussian distribution showed a better accuracy for the second set, by looking at the number of violations. The Kupiec test showed a slightly better fit for the multivariate student's t distribution than the Gaussian distribution. By looking at the Christoffersen test, it was not possible to draw any conclusions from it since both distributions performed equally.

The overall performance of both distributions are surprisingly similar, considering that the student's t distribution is widely known to be representing financial data better than the normal distribution. Due to it having heavier tails. However, the similarity of the two different distributions can be explained by the fact that the second stage estimations resulted into high values for ν. The two distributions performed equally for all of the tests but one - the Kupiec test, where the multivariate student's t distribution showed better results given the test statistic. Hence it would be fair to say that the multivariate student's t distribution performer slightly better than the Gaussian distribution.

9 References

References

[1] Richard T. Baillie and Tim Bollerslev. "Prediction in dynamic models with time-dependent conditional variances". In: *Journal of Econometrics* 52.1-2 (Apr. 1992), pp. 91–113. ISSN: 0304-4076. DOI: 10.1016/0304-4076(92)90066-Z.

[2] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics* 31.3 (Apr. 1986), pp. 307–327. ISSN: 0304-4076. DOI: 10.1016/0304-4076(86)90063-1.

[3] Tim Bollerslev. "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model". In: *The Review of Economics and Statistics* 72.3 (Aug. 1990), p. 498. DOI: 10.2307/2109358.

[4] Tim Bollerslev, Robert F. Engle, and Jeffrey M. Wooldridge. "A Capital Asset Pricing Model with Time-Varying Covariances". In: *Journal of Political Economy* 96.1 (Feb. 1988), pp. 116–131. ISSN: 0022-3808. DOI: 10.1086/261527.

[5] Kris Boudt et al. "Multivariate GARCH models for large-scale applications: A survey". In: *Handbook of Statistics* 41 (Jan. 2019), pp. 193–242. ISSN: 0169-7161. DOI: 10.1016/BS.HOST.2019.01.001.

[6] Julien Chevallier. "Time-varying correlations in oil, gas and CO 2 prices: an application using BEKK, CCC and DCC-MGARCH models". In: (2011). DOI: 10.1080/00036846.2011.589809.

[7] Robert Engle. "Dynamic Conditional Correlation A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models". In: (2002). ISSN: 1537-2707. DOI: 10.1198/073500102288618487.

[8] Robert F Engle and Kenneth F Kroner. "Multivariate Simultaneous Generalized Arch". In: 11.1 (1995), pp. 122–150.

[9] Robert F Engle and Kevin Sheppard. "NBER WORKING PAPER SERIES THEORETICAL AND EMPIRICAL PROPERTIES OF DYNAMIC CONDITIONAL CORRELATION MULTIVARIATE GARCH". In: (2001).

[10] Robert F. Engle. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4 (July 1982), p. 987. ISSN: 00129682. DOI: 10.2307/1912773.

[11] Robert F. Engle, Victor K. Ng, and Michael Rothschild. "Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills". In: *Journal of Econometrics* 45.1-2 (July 1990), pp. 213–237. ISSN: 0304-4076. DOI: 10.1016/0304-4076(90)90099-F.

[12] Robert F. Engle and Kevin Sheppard. "Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH". In: (Oct. 2001). DOI: 10.3386/W8554.

[13] Paul H. Kupiec. "Techniques for Verifying the Accuracy of Risk Measurement Models". In: *The Journal of Derivatives* 3.2 (Nov. 1995), pp. 73–84. ISSN: 1074-1240. DOI: 10.3905/JOD.1995.407942.

[14] G. M. LJUNG and G. E. P. BOX. "On a measure of lack of fit in time series models". In: *Biometrika* 65.2 (Aug. 1978), pp. 297–303. ISSN: 0006-3444. DOI: 10.1093/BIOMET/65.2.297.

[15] *Long, X. and Ullah, A. (2005): Nonparametric and... - Google Scholar.*

[16] Daniel B. Nelson and Charles Q. Cao. "Inequality constraints in the univariate garch model". In: *Journal of Business and Economic Statistics* 10.2 (1992), pp. 229–235. DOI: 10.1080/07350015.1992.10509902.

[17] Lennart Råde and Bertil Westergren. "Mathematics Handbook, BETA". In: (2014).

[18] Annastiina Silvennoinen. "Multivariate GARCH models". In: (2008).

[19] Pavel Stoimenov. "Philippe Jorion, Value at Risk, 3rd Ed: The New Benchmark for Managing Financial Risk". In: *Stat Papers* 52 (2011), pp. 737–738. DOI: 10.1007/s00362-009-0296-7.

[20] Roy Van Der Weide. "GO-GARCH: A multivariate generalized orthogonal GARCH model". In: *Journal of Applied Econometrics* 17.5 (Sept. 2002), pp. 549–564. DOI: 10.1002/JAE.688.

[21] Johannes Venter and Pieter de Jongh. "Risk estimation using the normal inverse Gaussian distribution". In: *The Journal of Risk* 4.2 (Mar. 2001), pp. 1–23. DOI: 10.21314/JOR.2002.059.

10 Appendix A

Summary of the Ljung Box test

Summary of the Ljung Box test			
Currency pair	Distribution	Correlated lags	Uncorrelated lags
USD/EUR	Gaussian	16	34
USD/EUR	MV student t	16	34
USD/GBP	Gaussian	13	37
USD/GBP	MV student t	16	34
USD/JPY	Gaussian	36	14
USD/JPY	MV student t	36	14

TRITA-SCI-GRU 2021:412