NTNU
Norwegian University of
Science and Technology

Multivariate DCC-GARCH Model
-With Various Error Distributions

Elisabeth Orskaug

Master of Science in Physics and Mathematics
Submission date: June 2009
Supervisor: Håvard Rue, MATH
Co-supervisor: Kjersti Aas, Norsk Regnesentral

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

This thesis is concerned with a multivariate GARCH model called Dynamic Conditional Correlation. This model is a tool for forecasting and analyzing volatility of time series when the volatility varies over time. DCC-GARCH with various error distributions have been considered; multivariate Gaussian, Student's t and skew Student's t.

Assignment given: 15. January 2009
Supervisor: Håvard Rue, MATH

Preface

This master thesis, which represents 30 credits, is written during the spring 2009, as a mandatory part of the master's degree in Industrial Mathematics at Department of mathematical sciences, Faculty of IME (Information Technology, Mathematics and Electrical Engineering) at NTNU (The Norwegian University of Science and Technology), Trondheim.

The implementation is done in the statistical software R. The style of writing is at a level that should be understandable to readers who have taken basic statistical courses, in addition to courses in multivariate analysis and time series. However, a stronger background in statistics is highly recommended, to fully understand all the presented matherial.

I have been working with the thesis at Norsk Regnesentral (Norwegian Computing Center), a technical and industrial research institute.

First I would like to thank my supervisor Kjersti Aas, who is Assistant Research Director at Norsk Regnesentral for instruction and guidance. I am also greatful for assistance from my supervisor at NTNU, Håvard Rue who is a professor at Faculty of IME.

Oslo, June 2009

Elisabeth Orskaug

Abstract

The focus in this thesis is on a specific multivariate GARCH model called Dynamic Conditional Correlation (DCC-) GARCH.

GARCH models are tools for forecasting and analyzing volatility of time series when the volatility varies over time.

There exists many multivariate GARCH models. An important goal in constructing multivariate GARCH models is to make them parsimonious enough, but still maintain flexibility. Another aspect is to ensure the conditional covariance matrix to be positive definite. DCC-GARCH model is a generalization of the CCC-GARCH model, which allows the correlation matrix to depend of the time. The DCC-GARCH model have clear computational advantages in that the number of parameters to be estimated in the correlation process is independent of the number of series to be correlated. Thus potentially very large correlation matrices can be estimated.

It have been implemented DCC-GARCH models for different assumptions of the error distribution; assuming Gaussian, Student t- and skew Student t-distribution.

Contents

1 Introduction

If you do business in the financial market it is important to have knowledge about volatility. This is important because volatility is a measure of the risk. If a person is faced with a decision of an investment in the industry, the future forecast will be highly sensitive to the choice of volatility modelling. It is known that volatility varies over time and tends to cluster in periods; small changes tend to be followed by small changes, and large changes by large ones. This phenomenon when the standard deviation varies over time is called heteroscedastisity. Heteroscedasticity means ”fluctuating variance”. In addition, the volatility has shown to be autocorrelated, which means that today’s volatility depends on the past volatility. Considering the fact that the volatility is not directly observable, the need of a good model to predict the future volatilities is essential. One model that has shown to be successful in capturing volatility clustering and predicting future volatilities is the univariate GARCH (Generalized Autoregressive Conditional Heteroscedasticity) model introduced by Bollerslev in 1986 [8].

It is known that financial volatilities move together more or less closely over time across assets and markets. Hence it is essential to take into account the dependence in the comovements of asset returns. One method to estimate the covariance matrix between the assets is to extend the univariate GARCH into a multivariate GARCH model. Extending from univariate to multivariate GARCH opens the door to better decision tools. The main challenge in constructing multivariate GARCH models is to make them parsimonious enough, but still maintain the flexibility.

One approach is to decompose the conditional covariance matrix into conditional standard deviations and a conditional correlation matrix. The first model of this type was the Constant Conditional Correlation (CCC-) model introduced by Bollerslev in 1990 [9]. In this model, the conditional correlation is assumed to be constant over time, and only the conditional standard deviation is time-varying. The assumption that the conditional correlation is constant over time is not always reasonable. In 2001, Engle and Sheppard introduced the DCC-GARCH model [11], which is an extension of the CCC-GARCH model, for which the conditional correlation matrix is designed to vary over the time. In this thesis the implementation of the DCC-GARCH model will be considered, using Gaussian, Student t- and skew Student t-distributed errors.

This thesis is structured as follows: In Chapter 2 the univariate GARCH model will be considered. In Chapter 3 multivariate GARCH models will be discussed in general to give a basic understanding before the DCC-GARCH model will be presented in Chapter 4. How to estimate the parameters of and determine the forecast and Value-at-Risk from the DCC-GARCH model will be considered in Chapters 5 and 6 respectively. In Chapter 7 the the goodness of fit of DCC-GARCH is discussed. In Chapter 8 DCC-GARCH models with different error distributions are fit to real data. Finally, some conclusions are presented in Chapter 9.

2 Univariate GARCH

2.1 Basic idea

The American economist R. F. Engle developed the ARCH model in 1982 [12]. This model captures the tendency of stock prices and other financial variables to move between high volatility and low volatility. Previous research either assumed the volatility to be constant or used simple devices to approximate it. There was need for a better model to measure risk, for example in pricing options and financial derivatives. The ARCH model became an essential tool of modern asset pricing theory and practice. In 2003, Engle was the winner of the "Nobel Memorial Prize in Economic Sciences" (shared with C. Granger) for his work with analyzing time series methods with time-varying volatility (ARCH).

The danish economist T. P. Bollerslev proposed the generalized ARCH model [8], known as GARCH, in 1986 after working with his Ph.D under supervision of Engle. This model is a more generalized model of the ARCH model, and is shown to be more successful in predicting volatilities than the ARCH model.

The main sources used in this chapter are the books [10], [18] and [1].

2.2 The GARCH model

The GARCH(q, p) model is defined as:

$$r_t = \mu_t + a_t \tag{1}$$

$$a_t = h_t^{1/2} z_t \tag{2}$$

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_q a_{t-q}^2 + \beta_1 h_{t-1} + ... + \beta_p h_{t-p} \tag{3}$$

Notation:

r_t :	log return of an asset at time t.
a_t :	mean-corrected return of an asset at time t.
μ_t:	the expected value of the conditional r_t.
h_t:	the square of the volatility, i.e. the conditional variance at time t, conditioned on the history.
$\{z_t\}$:	sequence of independent and identically distributed (iid) standardized, random variables, i.e. E[z_t]=0 and Var[z_t]=1.
$\alpha_0, \alpha_1, ..., \alpha_q$:	parameters of the model.
$\beta_1, ..., \beta_p$:	parameters of the model.
p, q:	order of the GARCH model

In (1) μ_t can be modelled as a time series, e.g. an ARMA model, or just as a constant. In this thesis the modelling of μ_t will not be the focus.

The volatility in (3) can be written as:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

From (3) we see that the conditional variance, h_t, varies over time, dependent on the last squared returns, $\{a_{t-i}^2\}_{i=1}^q$. If a big movement in the market occured yesterday, the day before yesterday or up to q days ago, the effect of this big movement will be shown in an increased volatility. Consequently a_t will tend to be large. This means that a large shock tends to be followed by another large shock. When the volatility is serially dependent, the time series will have periods of high volatility followed by periods of low volatility. This periodical dependence of volatility is called volatility clustering.

a_t is serially uncorrelated even though it is dependent on the last observations, i.e. the dependence is not linear.

It is important to be aware of that even though the volatility is large, this does not necessarily imply that a_t have to be large, it just means that the probability of obtaining a large value of a_t has increased.

An important weakness of the GARCH model is that it does not distinguish between positive and negative movements in the market. This is because the returns, $\{a_{t-i}\}_{i=1}^q$, are squared in (3).

2.3 Properties of the GARCH model

2.3.1 Unconditional distribution of a_t

In this section we derive unconditional mean, variance and autocovariance of a_t

Unconditional mean of a_t

The unconditional mean of a_t is:

$$\begin{aligned} \mathrm{E}[a_t] &= \mathrm{E}[\mathrm{E}\{a_t|F_{t-1}\}] \\ &= \mathrm{E}[\mathrm{E}\{h_t^{1/2} z_t|F_{t-1}\}] \\ &= \mathrm{E}[h_t^{1/2}\mathrm{E}\{z_t|F_{t-1}\}] \\ &= \mathrm{E}[h_t^{1/2}\mathrm{E}\{z_t\}] = 0 \end{aligned} \tag{4}$$

where F_t denotes the information set available at time t, i.e $F_t = \{a_s : s \leq t\}$. Here it is used that h_t and z_t are independent.

In the last equality of (4) we get 0 because $\mathrm{E}[z_t] = 0$ for all t. Hence the unconditional mean of a_t is 0 for all time t.

Unconditional variance of a_t

The unconditional variance of a_t is:

$$\begin{aligned}
\sigma^2 &= \mathrm{Var}[a_t] = \mathrm{E}[a_t^2] \\
&= \mathrm{E}[\mathrm{E}\{a_t^2|F_{t-1}\}] \\
&= \mathrm{E}[\mathrm{E}\{h_t z_t^2|F_{t-1}\}] \\
&= \mathrm{E}[h_t \mathrm{E}\{z_t^2\}] \\
&= \mathrm{E}[h_t] \\
&= \mathrm{E}[\alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_q a_{t-q}^2 + \beta_1 h_{t-1} + ... + \beta_p h_{t-p}] \\
&= \alpha_0 + \sum_{i=1}^{q} \alpha_i \mathrm{E}[a_{t-i}^2] + \sum_{j=1}^{p} \beta_j \mathrm{E}[h_{t-j}] \\
&= \alpha_0 + \sum_{i=1}^{q} \alpha_i \mathrm{Var}[a_{t-i}] + \sum_{j=1}^{p} \beta_j \mathrm{E}[h_{t-j}]
\end{aligned} \tag{5}$$

In (5) it is used that $\mathrm{E}[z_t^2] = \mathrm{Var}[z_t]$=1 since $\mathrm{E}[z_t] = 0$, and that h_t and z_t are independent.

$\mathrm{E}[h_{t-j}] = h_{t-j} = \sigma^2$ for $j = 1, ..., p$ because h_{t-j} is deterministic and equal to the unconditional variance, σ^2.

If we assume $\{a_t\}$ to be a stationary process, we get σ^2 =Var$[a_t]$ =Var$[a_{t-1}] = ...$ =Var$[a_{t-q}]$. Putting this in (5) and solving the equation with respect to σ^2 we get the unconditional variance:

$$\sigma^2 = \frac{\alpha_0}{1 - (\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j)} \tag{6}$$

The assumption that $\{a_t\}$ is stationary is only true when $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$, because the variance have to be positive. If this sum is greater than 1 there is no constant unconditional variance.

Unconditional autocovariance of a_t

The unconditional autocovariance of a_t is:

$$\mathrm{E}[a_{t+k}a_{t-1}] = \mathrm{E}[\mathrm{E}[a_{t+k}a_{t-1}|F_{t-1}]] = \mathrm{E}[a_{t-1}\mathrm{E}[a_{t+k}|F_{t-1}]] = 0 \quad \text{for } k = 0, 1, 2, ... \tag{7}$$

because $\mathrm{E}[a_{t+k}|F_{t-1}] = 0$.

Since $z_t \sim$ IID, the unconditional distribution has the same distribution for all t, i.e. $a_t, ..., a_{t+n}$ has the same distribution as $a_{t+k}, ..., a_{t+n+k}$. That is, the unconditional distribution of a_t is strictly stationary. Since $\mathrm{E}[z_t^2] < \infty$, $\mathrm{E}[a_t^2] < \infty$. Hence the unconditional distribution of $\{a_t\}$ is also weakly stationary.

2.3.2 Excess kurtosis of the GARCH model

The kurtosis of a model is interesting because it gives us an indication of the behaviour of the tails. The excess kurtosis of the normal distribution is 0. Distributions with positive excess kurtosis have heavier tails than the normal distribution. When the excess kurtosis is negative, the distributions are more light-tailed than the normal distribution. In this section the values of the kurtosis for the GARCH model is compared to the values of the kurtosis for the normal distribution.

For simplicity GARCH(1,1) will be considered in this section. The same idea can be applied to GARCH(q,p), $p = 1, 2, ..., q = 1, 2, ...$, as well.

The definition of the excess kurtosis of a_t, K_a, is:

$$\mathrm{K}_a = \frac{\mathrm{E}[a_t^4]}{\mathrm{Var}[a_t]^2} - 3 \tag{8}$$

From (6) we have that $\mathrm{Var}[a_t] = \mathrm{E}[a_t^2] = \alpha_0/[1 - (\alpha_1 + \beta_1)]$.

$\mathrm{E}[a_t^4]$ is:

$$\mathrm{E}[a_t^4] = \mathrm{E}[h_t^2 z_t^4] = \mathrm{E}[z_t^4]\mathrm{E}[h_t^2] = (3 + \mathrm{K}_z)\mathrm{E}[h_t^2], \tag{9}$$

where K_z is the excess kurtosis of z_t. When we assume $z_t \sim \mathrm{N}(0, 1)$, then $\mathrm{K}_z = 0$ and hence $\mathrm{E}[z_t^4] = 3$. This can be calculated by derivating the moment $\mathrm{M}_z(t) = e^{\frac{1}{2}t^2}$ four times, and we find $\mathrm{M}_z^{(4)}(0) = \mathrm{E}[z_t^4] = 3$.

We have to calculate $\mathrm{E}[h_t^2]$:

$$\begin{aligned}
\mathrm{E}[h_t^2] &= \mathrm{E}[(h_t)^2] \\
&= \mathrm{E}[(\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1})^2] \\
&= \mathrm{E}[\alpha_0^2 + \alpha_1^2 a_{t-1}^4 + \beta_1^2 h_{t-1}^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + 2\alpha_0\beta_1 h_{t-1} + 2\alpha_1\beta_1 h_{t-1} a_{t-1}^2] \\
&= \alpha_0^2 + \alpha_1^2\mathrm{E}[a_{t-1}^4] + \beta_1^2\mathrm{E}[h_{t-1}^2] + 2\alpha_0\alpha_1\mathrm{E}[a_{t-1}^2] + 2\alpha_0\beta_1\mathrm{E}[h_{t-1}] + 2\alpha_1\beta_1\mathrm{E}[h_{t-1}a_{t-1}^2]
\end{aligned} \tag{10}$$

Stationarity gives that $\mathrm{E}[h_t^2] = \mathrm{E}[h_{t-1}^2]$. $\mathrm{E}[a_{t-1}^2] = \mathrm{E}[h_{t-1}] = \sigma^2 = \frac{\alpha_0}{1-\alpha_1-\beta_1}$ from (6) and $\mathrm{E}[h_{t-1}a_{t-1}^2] = \mathrm{E}[h_{t-1}^2]$ because $a_{t-1}^2 = h_{t-1}$ since the history up to time $t-1$ is known at time t. $\mathrm{E}[a_{t-1}^4] = (\mathrm{K}_z + 3)\mathrm{E}[h_{t-1}^2]$ from (9). Putting this in (10) and solving for $\mathrm{E}[h_t^2]$ gives:

$$\mathrm{E}[h_t^2] = \frac{\alpha_0^2(1 + \alpha_1 + \beta_1)}{[1 - \alpha_1 - \beta_1][1 - (\mathrm{K}_z + 2)\alpha_1^2 - (\alpha_1 + \beta_1)^2]} \tag{11}$$

Inserting (11) into (9) gives:

$$\mathrm{E}[a_t^4] = \frac{\alpha_0^2(3 + \mathrm{K}_z)(1 + \alpha_1 + \beta_1)}{[1 - \alpha_1 - \beta_1][1 - (\mathrm{K}_z + 2)\alpha_1^2 - (\alpha_1 + \beta_1)^2]} \tag{12}$$

Finally, putting (12) and $\mathrm{Var}[a_t]^2 = (\sigma^2)^2 = (\frac{\alpha_0}{1-\alpha_1-\beta_1})^2$ into (8), the excess kurtosis of a_t is:

$$
\begin{aligned}
\mathrm{K}_a &= \frac{\mathrm{E}[a_t^4]}{\mathrm{Var}[a_t]^2} - 3 \\
&= \frac{\alpha_0^2(3+\mathrm{K}_z)(1+\alpha_1+\beta_1)}{[1-\alpha_1-\beta_1][1-(\mathrm{K}_z+2)\alpha_1^2-(\alpha_1+\beta_1)^2]} \frac{(1-\alpha_1-\beta_1)^2}{\alpha_0^2} - 3 \\
&= \frac{(3+\mathrm{K}_z)[1-(\alpha_1+\beta_1)^2]}{1-(\mathrm{K}_z+2)\alpha_1^2-(\alpha_1+\beta_1)^2} - 3
\end{aligned}
\tag{13}
$$

Hence the excess kurtosis of a_t is dependent on the distribution of z_t.

Consider the case when $z_t \sim \mathbf{N}(0,1)$:

When $z_t \sim \mathrm{N}(0,1)$, then $\mathrm{K}_z = 0$. Then we get from (13):

$$
\mathrm{K}_a^{(g)} = \frac{3[1-(\alpha_1+\beta_1)^2]}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2} - 3 = \frac{6\alpha_1^2}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2} \tag{14}
$$

The superscript (g) is used to denote that z_t is Gaussian distributed.

This result shows that the kurtosis of a_t exists if $1-2\alpha_1^2-(\alpha_1+\beta_1)^2 > 0$. When $\alpha_1 > 0$, $\mathrm{K}_a^{(g)} > 0$ and the distribution of a_t has heavy tails. If $\alpha_1 = 0$, $\mathrm{K}_a^{(g)} = 0$, and then the GARCH(1,1) model does not have heavy tails.

Consider the case when z_t is not Gaussian distributed:

Using the result in (14) we can rewrite the general expression for the excess kurtosis of a_t from (13), and for easier notation put $\mathrm{D}_a = 1-2\alpha_1^2-(\alpha_1+\beta_1)^2$:

$$
\begin{aligned}
\mathrm{K}_a &= \frac{(3+\mathrm{K}_z)[1-(\alpha_1+\beta_1)^2]}{1-(\mathrm{K}_z+2)\alpha_1^2-(\alpha_1+\beta_1)^2} - 3 \\
&= \frac{\mathrm{K}_z[1-2\alpha_1^2-(\alpha_1+\beta_1)^2]+6\alpha_1^2+5\mathrm{K}_z\alpha_1^2}{[1-2\alpha_1^2-(\alpha_1+\beta_1)^2]-\mathrm{K}_z\alpha_1^2} \\
&= \frac{\mathrm{K}_z\mathrm{D}_a+6\alpha_1^2+5\mathrm{K}_z\alpha_1^2}{\mathrm{D}_a-\mathrm{K}_z\alpha_1^2} \\
&= \frac{\mathrm{K}_z+\frac{6\alpha_1^2}{\mathrm{D}_a}+\frac{5\mathrm{K}_z\alpha_1^2}{\mathrm{D}_a}}{1-\frac{\mathrm{K}_z\alpha_1^2}{\mathrm{D}_a}} \\
&= \frac{\mathrm{K}_z+\mathrm{K}_a^{(g)}+\frac{5}{6}\mathrm{K}_z\mathrm{K}_a^{(g)}}{1-\frac{1}{6}\mathrm{K}_z\mathrm{K}_a^{(g)}}
\end{aligned}
\tag{15}
$$

The result in (15) shows that for a GARCH(1,1) model, the coefficient α_1 plays a critical role in determining the tail behaviour of a_t. For instance if $\alpha_1 = 0$, $\mathrm{K}_a^{(g)} = 0$ and $\mathrm{K}_a = \mathrm{K}_z$. Hence the tail behaviour of a_t is similar to that of the standardized error z_t. When $\alpha_1 > 0$, $\mathrm{K}_a^{(g)} > 0$ and the distribution of a_t has heavy tails. Hence the tailheaviness of the distribution of a_t is dependent on the parameters in the model and the distribution of the error z_t. For instance if z_t is Student-t distributed, then $\mathrm{K}_z = 6/(\nu\text{-}4)$ for $\nu > 4$.

This result holds for all GARCH models, but only if the excess kurtosis exists.

2.4 Estimation of the parameters

Estimation of the parameters in the GARCH model can be done by maximum likelihood. The expression of the likelihood is:

$$\begin{aligned}\mathrm{L}(\boldsymbol{\alpha},\boldsymbol{\beta}|a_1,..,a_n) &= f(a_1,...,a_n|\boldsymbol{\alpha},\boldsymbol{\beta})\\ &= f(a_n|F_{n-1},\boldsymbol{\alpha},\boldsymbol{\beta})f(a_{n-1}|F_{n-2},\boldsymbol{\alpha},\boldsymbol{\beta})\cdots f(a_{m+1}|F_m,\boldsymbol{\alpha},\boldsymbol{\beta})\,\cdot\, f(a_1,...,a_m|\boldsymbol{\alpha},\boldsymbol{\beta})\end{aligned}$$

Where n is the sample size, m =max(p,q), $\boldsymbol{\alpha} = (\alpha_0,\alpha_1,...,\alpha_q)^T$, $\boldsymbol{\beta} = (\beta_1,...,\beta_p)^T$ and $f(a_1,...,a_m|\boldsymbol{\alpha},\boldsymbol{\beta})$ is the joint probability density function of $a_1,...,a_m$.

Since the exact form of $f(a_1,...,a_m|\boldsymbol{\alpha})$ is complicated, the conditional likelihood function is usually used instead. It is given by:

$$\mathrm{L}(\boldsymbol{\alpha},\boldsymbol{\beta}|a_{m+1},..,a_n) = f(a_{m+1},...,a_n|\boldsymbol{\alpha},\boldsymbol{\beta},a_1,...,a_m)$$

To determine $f(a_{m+1},...,a_n|\boldsymbol{\alpha},\boldsymbol{\beta},a_1,...,a_m)$, we have to decide a model for z_t. The most common is to assume z_t to be standard Gaussian distributed.

Assuming $\mathbf{z_t} \sim$ N(0,1):

When $z_t \sim$ N(0,1), $a_t|F_{t-1} \sim$ N(0,h_t), since $a_t = \sqrt{h_t}z_t$. This can be calculated from the transformation of variables. In this case the conditional likelihood function is:

$$\mathrm{L}(\boldsymbol{\alpha},\boldsymbol{\beta}|a_{m+1},..,a_n) = f(a_{m+1},...,a_n|\boldsymbol{\alpha},\boldsymbol{\beta},a_1,...,a_m) = \prod_{t=1}^{T}\frac{1}{\sqrt{2\pi h_t}}\exp\left\{-\frac{a_t^2}{2h_t}\right\}$$

where $t = 1,...,T$ denotes the time points in the conditional likelihood function.

Maximizing the conditional likelihood function is equivalent to maximizing its logarithm because ln($\cdot$) is a strictly increasing function. Since the log-likelihood is easier to handle, we prefer this:

$$\begin{aligned}\ln(\mathrm{L}) &= l(a_{m+1},...,a_n|\boldsymbol{\alpha},\boldsymbol{\beta},a_1,...,a_m)\\ &= -\frac{1}{2}\sum_{t=1}^{T}\left[\ln(2\pi)+\ln(h_t)+\frac{a_t^2}{h_t}\right]\\ &= -\frac{1}{2}\sum_{t=1}^{T}\left[\ln(h_t)+\frac{a_t^2}{h_t}\right]+\text{constant}\end{aligned} \tag{16}$$

where $h_t = \alpha_0+\alpha_1a_{t-1}^2+...+\alpha_qa_{t-q}^2+\beta_1h_{t-1}+...+\beta_ph_{t-p}$ has to be evaluated recursively.

Maximizing the log likelihoods

The log-likelihood in (16) can be maximized using numerical optimization methods. The most common method is using a quasi-Newton optimizer.

A common quasi-Newton method for this problem is the Broyden-Fletcher-Goldfarb-Shanno (BFGS). For instance, the statistical software R is using this method when maximizing the log-likelihood.

Distribution of the Maximum Likelihood Estimator

The maximum likelihood estimator can be shown to be approximately normally distributed, with the mean as the true parameter value, $\hat{\boldsymbol{\theta}} \sim \mathrm{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}$$

$\boldsymbol{\Sigma}$ is approximately equal to the inverse of the matrix whose (i, j)th element is:

$$\mathrm{Cov}(\theta_i, \theta_j) = \frac{1}{2} \sum_{t=1}^{T} \frac{1}{h_t^2} \frac{\partial h_t}{\partial \theta_i} \frac{\partial h_t}{\partial \theta_j}$$

Estimation problems

Sometimes there is problem with convergence because the likelihood function becomes flat when the number of parameters is large. It may be that only a local optimum is achieved, and in this case the starting values of the parameters are very important. Different set of estimates may be obtained when the starting values are changed. To secure that we really have found the global maximum, one should run the model with many different starting values. If possible, one should try to use a more parsimonious parametrization of the model if this convergence problem occur.

Another convergence problem may occur if the gradient algorithm used to maximize the likelihood function has hit a boundary. If there are obvious outliers in the data, it is very likely that the optimization algorithm will return the value 0 or 1 for either the $\boldsymbol{\alpha}$ or the $\boldsymbol{\beta}$ parameter (or both). If there is still a problem with convergence after removing the outliers, one should try to change the starting values.

Most univariate models will encounter few convergence problems if the model is well specified for the data, and especially if the number of parameters is low.

2.5 Prediction of the volatility

The main use of the GARCH model is to predict future volatilities. More specifically one wants to predict the future volatility k-step ahead, i.e. h_{t+k}. Assume a GARCH(q,p) model. Then we have:

$$
\begin{aligned}
\mathrm{E}[h_{t+k}|F_t] &= \mathrm{Var}[a_{t+k}|F_t] \\
&= \mathrm{E}[a_{t+k}^2|F_t] \\
&= \mathrm{E}[\mathrm{E}\{h_{t+k}z_{t+k}^2|F_{t+k-1}\}|F_t] \\
&= \mathrm{E}[h_{t+k}|F_t] \\
&= \mathrm{E}[\alpha_0 + \alpha_1 a_{t+k-1}^2 + ... + \alpha_q a_{t+k-q}^2 + \beta_1 h_{t+k-1} + ... + \beta_p h_{t+k-p}|F_t] \\
&= \alpha_0 + \alpha_1 \mathrm{E}[a_{t+k-1}^2|F_t] + ... + \alpha_q \mathrm{E}[a_{t+k-q}^2|F_t] + \beta_1 \mathrm{E}[h_{t+k-1}|F_t] + \\
&\quad ... + \beta_p \mathrm{E}[h_{t+k-p}|F_t] \\
&= \alpha_0 + \alpha_1 \mathrm{E}[h_{t+k-1}] + ... + \alpha_q \mathrm{E}[h_{t+k-q}] + \beta_1 \mathrm{E}[h_{t+k-1}] + \\
&\quad ... + \beta_p \mathrm{E}[h_{t+k-p}] \\
&= \alpha_0 + \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)\mathrm{E}[h_{t+k-i}]
\end{aligned} \tag{17}
$$

where

$$\mathrm{E}[h_{t+k}] = a_{t+k}^2 \quad \text{for } k < 0$$

The k-step-ahead conditional variance can be found recursively from this formula, by first computing $\mathrm{E}[h_{t+1}|F_t]$, then $\mathrm{E}[h_{t+2}|F_t]$ and up to $\mathrm{E}[h_{t+k}|F_t]$.

If the forecast horizon, k, increases to infinity, it is reasonable to believe that the prediction of the volatility will converge to the unconditional variance, since the nearest past is unknown. Let us check whether this is true for the simplest model GARCH(1,1). (17) can be rewritten as:

$$
\begin{aligned}
\mathrm{E}[h_{t+k}] &= \alpha_0 + (\alpha_1 + \beta_1)h_{t+k-1} \\
&= \alpha_0 + (\alpha_1 + \beta_1)(\alpha_0 + (\alpha_1 + \beta_1)h_{t+k-2}) \\
&= \alpha_0 + (\alpha_1 + \beta_1)\alpha_0 + (\alpha_1 + \beta_1)^2(\alpha_0 + (\alpha_1 + \beta_1)h_{t+k-3}) \\
&= \alpha_0 \sum_{i=0}^{k-2} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{k-1} h_{t+1} \\
&= \alpha_0 \left(\frac{1 - (\alpha_1 + \beta_1)^{k-1}}{1 - \alpha_1 - \beta_1} \right) + (\alpha_1 + \beta_1)^{k-1} h_{t+1} \quad \rightarrow \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \quad \text{as } k \rightarrow \infty
\end{aligned} \tag{18}
$$

It is used that $\sum_{i=1}^{k-2} (\alpha_1 + \beta_1)^i$ is a geometric series since $(\alpha_1 + \beta_1) < 1$.

This result shows that the GARCH(1,1)-prediction of the volatility in the distant future converges to the stationary variance as calculated in (6). This is also the case in general for GARCH(q,p), i.e.

$$h_{t+k} \rightarrow \frac{\alpha_0}{1 - (\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j)} \quad \text{as } k \rightarrow \infty$$

3 Multivariate GARCH

3.1 Basic idea - why extend from univariate to multivariate?

- In financial econometrics and management understanding, predicting the dependence in the comovements of asset returns is important. For example, asset pricing depends on the covariance of the assets in a portfolio. Hence it is important to consider the comovements in the portfolio.
- Financial volatilities move together more or less closely over time across assets and markets.
- Recognizing this feature through a multivariate model should lead to more relevant empirical models than working with separate univariate models.
- In financial applications, extending from univariate to multivariate modelling opens the door to better decision tools in various areas such as asset pricing models, portfolio selection, hedging, and Value-at-Risk forecasts.

3.2 The multivariate GARCH models

The multivariate GARCH models are defined as:

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{a}_t \tag{19}$$

$$\boldsymbol{a}_t = \boldsymbol{H}_t^{1/2} \boldsymbol{z}_t \tag{20}$$

Notation:

$\boldsymbol{r}_t$:	$n \times 1$ vector of log returns of n assets at time t.
$\boldsymbol{a}_t$:	$n \times 1$ vector of mean-corrected returns of n assets at time t, i.e. E[$\boldsymbol{a}_t$]=0. Cov[$\boldsymbol{a}_t$] $= \boldsymbol{H}_t$.
$\boldsymbol{\mu}_t$:	$n \times 1$ vector of the expected value of the conditional r_t.
$\boldsymbol{H}_t$:	$n \times n$ matrix of conditional variances of $\boldsymbol{a}_t$ at time t.
$\boldsymbol{H}_t^{1/2}$:	Any $n \times n$ matrix at time t such that $\boldsymbol{H}_t$ is the conditional variance matrix of a_t. $\boldsymbol{H}_t^{1/2}$ may be obtained by a Cholesky factorization of $\boldsymbol{H}_t$.
$\boldsymbol{z}_t$:	$n \times 1$ vector of iid errors such that E[$\boldsymbol{z}_t$]=0 and E[$\boldsymbol{z}_t \boldsymbol{z}_t^T$] $= I$.

$\boldsymbol{\mu}_t$ in (19) may be modelled as a constant vector or a time series model. The modelling of $\boldsymbol{\mu}_t$ is however not the focus in this thesis.

As in the univariate case, $\boldsymbol{a}_t$ is uncorrelated in time. However this does not mean that there is no serial dependence, but that the dependence is non-linear.

What remains to be specified is the conditional covariance matrix, $\boldsymbol{H}_t$. There are many possible specifications of $\boldsymbol{H}_t$. The parameters in the conditional covariance matrices increase very rapidly as the dimension of $\boldsymbol{a}_t$ increases.

Since $\boldsymbol{H}_t$ is dependent of the time t, it has to be inverted in each iteration, which makes the computation demanding unless n is small. This creates difficulties in the estimation of the models, and therefore an important goal in constructing the MGARCH models is to make them parsimonious enough, but still maintain the flexibility. Another aspect is to ensure the conditional covariance matrix to be positive definite.

The different specifications of MGARCH models can be divided into four categories as suggested in [17]:

1. **Models of the conditional covariance matrix;** In this class the conditional covariance matrices, $\boldsymbol{H}_t$, are modelled directly. This class includes the VEC and BEKK models. These models were among the first parametric MGARCH models.
2. **Factor models;** The idea of factor models comes from economic theory. In this class the conditional covariance matrices are motivated by parsimony. The process $\boldsymbol{a}_t$ is assumed to be generated by a (small) number of unobserved heteroskedastic factors, hence these models are called factor models. These factors can be studied and one may make assumptions that some characteristics of the data is captured, similar as for principal component analysis. This approach has the advantage that it reduces the dimensionality of the problem when the number of factors relative to the dimension of the return vector $\boldsymbol{a}_t$ is small.
3. **Models of conditional variances and correlations;** The models in this class are built on the idea of modelling the conditional variances and correlations instead of straightforward modelling the conditional covariance matrix. We will consider one specific model of this class in Section 3.3.
4. **Nonparametric and semiparametric approaches;** Models in this class form an alternative to parametric estimation of the conditional covariance structure. The advantage of these models is that they do not impose a particular structure (that can be misspecified) on the data.

The goal of this thesis is to have a closer look at a model called Dynamic Conditional Correlation (DCC-) GARCH which belongs to the category 3 above. However, before we give a closer description of the DCC model in Chapter 4, a more general description of the category 3 models are given in Section 3.3.

3.3 Models of conditional variances and correlations

The models in this class are built on the idea of modelling the conditional variances and correlations instead of straightforward modelling the conditional covariance matrix. The conditional covariance matrix is decomposed into conditional standard deviations and a correlation matrix as:

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t \tag{21}$$

where $\boldsymbol{D}_t = \mathrm{diag}(h_{1t}^{1/2}, ..., h_{nt}^{1/2})$ is the conditional standard deviation, and $\boldsymbol{R}_t$ is the correlation matrix. Models in this class can be classified in two groups; those with a constant correlation matrix and those when the correlation matrix is time-varying.

3.3.1 Constant correlation matrix

Models in this class includes the Constant Conditional Correlation (CCC-) GARCH of Bollerslev [9] and its extensions. The conditional correlation matrix is time invariant, i.e. $\boldsymbol{R}_t = \boldsymbol{R}$. Hence (21) becomes:

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t$$

The correlation matrix, $\boldsymbol{R} = [\rho_{ij}]$, is positive definite with $\rho_{ii} = 1, i = 1, ..., n$. The off-diagonal elements of the conditional covariance matrix, $\boldsymbol{H}_t$, are given by:

$$[\boldsymbol{H}_t]_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij}, \quad i \neq j \tag{22}$$

The process $\{a_{it}\}$ is modelled as univariate GARCH. Hence the conditional variances can be written in a vector form:

$$\boldsymbol{h}_t = \boldsymbol{c} + \sum_{j=1}^{q} \boldsymbol{A}_j \boldsymbol{a}_{t-j}^{(2)} + \sum_{j=1}^{p} \boldsymbol{B}_j \boldsymbol{h}_{t-j} \tag{23}$$

where $\boldsymbol{c}$ is $n \times 1$ vector, $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ are diagonal $n \times n$ matrices, and $\boldsymbol{a}_{t-j}^{(2)} = \boldsymbol{a}_{t-j} \odot \boldsymbol{a}_{t-j}$ is the element–wise product. $\boldsymbol{H}_t$ is ensured positive definite when the elements of $\boldsymbol{c}$ and $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ are positive, since $\boldsymbol{R}$ is positive definite.

There exists also an extended CCC-GARCH model for which $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ do not need to be diagonal.

The estimation of models in this class is computationally attractive because the correlation matrix is constant. However the CCC-GARCH model may be too restrictive in some cases. The model may then be generalized by assuming the correlation matrix to vary with time.

3.3.2 Time-varying correlation matrix

When the correlation matrix, $\boldsymbol{R}_t$, is time-varying, $\boldsymbol{H}_t$ is positive definite if $\boldsymbol{R}_t$ is positive definite at each point in time and the conditional variances, $h_{it}, i = 1, ..., n$ are well-defined. Compared to the CCC-GARCH model, the advantage of numerically simple estimation is lost, as the correlation matrix has to be inverted for each time, t, during every iteration. Several specifications of $\boldsymbol{R}_t$ have been suggested in the literature. In this thesis we will study one specification; the DCC-GARCH model.

4 DCC-GARCH

4.1 Basic idea

The Dynamic Conditional Correlation (DCC-) GARCH belongs to the class "Models of conditional variances and correlations" as discussed in Section 3.3. It was introduced by Engle and Sheppard in 2001 [11]. The idea of the models in this class is that the covariance matrix, $\boldsymbol{H}_t$, can be decomposed into conditional standard deviations, $\boldsymbol{D}_t$, and a correlation matrix, $\boldsymbol{R}_t$. In the DCC-GARCH model both $\boldsymbol{D}_t$ and $\boldsymbol{R}_t$ are designed to be time-varying.

4.2 The DCC-GARCH model

Suppose we have returns, $\boldsymbol{a}_t$, from n assets with expected value 0 and covariance matrix $\mathbf{H}_t$. Then the Dynamic Conditional Correlation (DCC-) GARCH model is defined as:

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{a}_t \tag{24}$$

$$\boldsymbol{a}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t \tag{25}$$

$$\boldsymbol{H}_t = \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t \tag{26}$$

Notation:

$\mathbf{r}_t$: $n \times 1$ vector of log returns of n assets at time t.

$\boldsymbol{a}_t$: $n \times 1$ vector of mean-corrected returns of n assets at time t, i.e. E[$\boldsymbol{a}_t$]=0. Cov[$\boldsymbol{a}_t$] = $\boldsymbol{H}_t$.

$\boldsymbol{\mu}_t$: $n \times 1$ vector of the expected value of the conditional $\boldsymbol{r}_t$.

$\boldsymbol{H}_t$: $n \times n$ matrix of conditional variances of $\boldsymbol{a}_t$ at time t.

$\boldsymbol{H}_t^{1/2}$: Any $n \times n$ matrix at time t such that $\boldsymbol{H}_t$ is the conditional variance matrix of a_t. $\boldsymbol{H}_t^{1/2}$ may be obtained by a Cholesky factorization of $\boldsymbol{H}_t$.

$\boldsymbol{D}_t$: $n \times n$, diagonal matrix of conditional standard deviations of $\boldsymbol{a}_t$ at time t.

$\boldsymbol{R}_t$: $n \times n$ conditional correlation matrix of $\mathbf{a}_t$ at time t.

$\boldsymbol{z}_t$: $n \times 1$ vector of iid errors such that E[$\boldsymbol{z}_t$]=0 and E[$\boldsymbol{z}_t\boldsymbol{z}_t^T$] $= I$.

$\boldsymbol{\mu}_t$ in (24) may be modelled as a constant vector or a time series model. The modelling of $\boldsymbol{\mu}_t$ is however not the focus in this thesis.

The elements in the diagonal matrix $\boldsymbol{D}_t$ are standard deviations from univariate GARCH models.

$$\boldsymbol{D}_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 & \cdots & 0 \\ 0 & \sqrt{h_{2t}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{h_{nt}} \end{bmatrix}$$

where

$$h_{it} = \alpha_{i0} + \sum_{q=1}^{Q_i} \alpha_{iq} a_{i,t-q}^2 + \sum_{p=1}^{P_i} \beta_{ip} h_{i,t-p}$$

Note that the univariate GARCH models can have different orders. Often the simplest model, GARCH(1,1), is adequate. The specification of the univariate GARCH models is not limited to the standard univariate GARCH(p,q) in Chapter 2, but can include any GARCH process with Gaussian distributed errors that satisfies appropriate stationarity conditions that ensures the unconditional variance to exist. In this thesis, however, only the standard univariate GARCH in Chapter 2 will be considered.

$\boldsymbol{R}_t$ is the conditional correlation matrix of the standardized disturbances ϵ_t, i.e:

$$\boldsymbol{\epsilon}_t = \boldsymbol{D}_t^{-1} \boldsymbol{a}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{R}_t)$$

Since $\boldsymbol{R}_t$ is a correlation matrix it is symmetric.

$$\boldsymbol{R}_t = \begin{bmatrix} 1 & \rho_{12,t} & \rho_{13,t} & \cdots & \rho_{1n,t} \\ \rho_{12,t} & 1 & \rho_{23,t} & \cdots & \rho_{2n,t} \\ \rho_{13,t} & \rho_{23,t} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_{n-1,n,t} \\ \rho_{1n,t} & \rho_{2n,t} & \cdots & \rho_{n-1,n,t} & 1 \end{bmatrix}$$

The elements of $\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t$ is:

$$[\boldsymbol{H}_t]_{ij} = \sqrt{h_{it} h_{jt}} \rho_{ij} \tag{27}$$

where $\rho_{ii} = 1$.

As discussed in Section 3.3 there exists different forms of $\boldsymbol{R}_t$. When specifying a form of $\boldsymbol{R}_t$ two requirements have to be considered:

1. $\boldsymbol{H}_t$ has to be positive definite because it is a covariance matrix. To ensure $\boldsymbol{H}_t$ to be positive definite, $\boldsymbol{R}_t$ has to be positive definite ($\boldsymbol{D}_t$ is positive definite since all the diagonal elements are positive).
2. All the elements in the correlation matrix $\boldsymbol{R}_t$ have to be equal to or less than one by definition.

To ensure both of these requirements in the DCC-GARCH model, $\boldsymbol{R}_t$ is decomposed into:

$$\boldsymbol{R}_t = \boldsymbol{Q}_t^{*-1}\boldsymbol{Q}_t\boldsymbol{Q}_t^{*-1} \tag{28}$$

$$\boldsymbol{Q}_t = (1-a-b)\overline{\boldsymbol{Q}} + a\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^T + b\boldsymbol{Q}_{t-1} \tag{29}$$

where $\overline{\boldsymbol{Q}} = \mathrm{Cov}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T] = \mathrm{E}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T]$ is the unconditional covariance matrix of the standardized errors $\boldsymbol{\epsilon}_t$. $\overline{\boldsymbol{Q}}$ can be estimated as [13]:

$$\overline{\boldsymbol{Q}} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t\epsilon_t^T$$

The parameters a and b are scalars, and $\mathbf{Q}_t^*$ is a diagonal matrix with the square root of the diagonal elements of $\boldsymbol{Q}_t$ at the diagonal:

$$\boldsymbol{Q}_t^* = \begin{bmatrix} \sqrt{q_{11t}} & 0 & \cdots & 0 \\ 0 & \sqrt{q_{22t}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{q_{nnt}} \end{bmatrix}$$

$\boldsymbol{Q}_t^*$ rescales the elements in $\boldsymbol{Q}_t$ to ensure the second requirement; $|\rho_{ij}| = |\frac{q_{ijt}}{\sqrt{q_{iit}q_{jjt}}}| \leq 1$. Further $\boldsymbol{Q}_t$ has to be positive definite to ensure $\boldsymbol{R}_t$ to be positive definite.

There are imposed some conditions on the parameters a and b to guarantee $\boldsymbol{H}_t$ to be positive definite. In addition to the conditions for the univariate GARCH model to ensure positive unconditional variances, given in Section 2.3.1, the scalars a and b must satisfy:

$$a \geq 0\ , \quad b \geq 0 \quad \text{and} \quad a+b<1$$

In addition $\boldsymbol{Q}_0$, the starting value of $\boldsymbol{Q}_t$, has to be positive definite to guarantee $\boldsymbol{H}_t$ to be positive definite.

The correlation structure can be extended to the general DCC(M,N)-GARCH model:

$$\boldsymbol{Q}_t = (1 - \sum_{m=1}^{M} a_m - \sum_{n=1}^{N} b_n)\overline{\boldsymbol{Q}}_t + \sum_{m=1}^{M} a_m \mathbf{a}_{t-1}\mathbf{a}_{t-1}^T + \sum_{n=1}^{N} b_n\boldsymbol{Q}_{t-1} \tag{30}$$

In this paper only the DCC(1,1)-GARCH model will be studied. For more details of the general DCC(M,N)-GARCH see [12].

5 Estimation of DCC-GARCH

In this chapter we describe how the parameters of a DCC-GARCH model may be determined. We consider three different distributions for the standardized error $\boldsymbol{z}_t$; the multivariate Gaussian, the multivariate Student's t- and a multivariate skew Student's t-distribution.

5.1 Multivariate Gaussian distributed errors

When the standardized errors, $\boldsymbol{z}_t$, are multivariate Gaussian distributed, the joint distribution of $z_1, ..., z_T$ is:

$$f(\boldsymbol{z}_t) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{n/2}} \exp\{-\frac{1}{2}\boldsymbol{z}_t^T \boldsymbol{z}_t\}$$

since E[$\boldsymbol{z}_t$]=0 and E$[\boldsymbol{z}_t\boldsymbol{z}_t^T] = I$. Here $t = 1, ..., T$ is the time period used to estimate the model.

Using the rule for linear transformation of variables (see e.g. page 13 in [3]), the likelihood function for $\boldsymbol{a}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t$ is:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{n/2}|\boldsymbol{H}_t|^{1/2}} \exp\{-\frac{1}{2}\boldsymbol{a}_t^T \boldsymbol{H}_t^{-1}\boldsymbol{a}_t\} \tag{31}$$

where $\boldsymbol{\theta}$ denotes the parameters of the model. Let the parameters, $\boldsymbol{\theta}$, be divided in two groups; $(\boldsymbol{\phi}, \boldsymbol{\psi}) = (\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_n, \boldsymbol{\psi})$, where $\boldsymbol{\phi}_i = (\alpha_{0i}, \alpha_{1i}, ..., \alpha_{qi}, \beta_{1i}, ..., \beta_{pi})$ are the parameters of the univariate GARCH model for the i^{th} asset series, $i = 1, ..., n$. $\boldsymbol{\psi} = (a, b)$ are the parameters of the correlation structure in (29).

By taking the logarithm of (31) and substituting $\boldsymbol{H}_t = \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t$ we get the log-likelihood:

$$\begin{aligned}
\ln(L(\boldsymbol{\theta})) &= -\frac{1}{2}\sum_{t=1}^{T}\Big(n\ln(2\pi) + \ln(|\boldsymbol{H}_t|) + \boldsymbol{a}_t^T\boldsymbol{H}_t^{-1}\boldsymbol{a}_t\Big) \\
&= -\frac{1}{2}\sum_{t=1}^{T}\Big(n\ln(2\pi) + \ln(|\boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t|) + \boldsymbol{a}_t^T\boldsymbol{D}_t^{-1}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{a}_t\Big) \\
&- \frac{1}{2}\sum_{t=1}^{T}\Big(n\ln(2\pi) + 2\ln(|\boldsymbol{D}_t|) + \ln(|\boldsymbol{R}_t|) + \boldsymbol{a}_t^T\boldsymbol{D}_t^{-1}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{a}_t\Big)
\end{aligned} \tag{32}$$

The estimation of the correctly specified log-likelihood is difficult, and hence the DCC-model was designed to allow for two stage estimation. In the first stage the parameter $\boldsymbol{\phi}$ of the univariate GARCH models are estimated for each asset series. The likelihood used in the first stage results in replacing $\boldsymbol{R}_t$ with the identity matrix $\boldsymbol{I}_n$. In the second stage, the parameter $\boldsymbol{\psi}$ are estimated using the correctly specified log-likelihood in (32), given the parameter $\boldsymbol{\phi}$.

5.1.1 Step one

In the first stage $\boldsymbol{R}_t$ is replaced with the identity matrix $\boldsymbol{I}_n$ in (32), which results in the quasi-likelihood function:

$$
\begin{aligned}
\ln(L_1(\boldsymbol{\phi})) &= -\frac{1}{2}\sum_{t=1}^{T}\Big(n\ln(2\pi) + 2\ln(|\boldsymbol{D}_t|) + \ln(|\boldsymbol{I}_n|) + \boldsymbol{a}_t^T\boldsymbol{D}_t^{-1}\boldsymbol{I}_n\boldsymbol{D}_t^{-1}\boldsymbol{a}_t\Big)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\Big(n\ln(2\pi) + 2\ln(|\boldsymbol{D}_t|) + \boldsymbol{a}_t^T\boldsymbol{D}_t^{-1}\boldsymbol{I}_n\boldsymbol{D}_t^{-1}\boldsymbol{a}_t\Big)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(n\ln(2\pi) + \sum_{i=1}^{n}\left[\ln(h_{it}) + \frac{a_{it}^2}{h_{it}}\right]\right)\\
&= \sum_{i=1}^{n}\left(-\frac{1}{2}\sum_{t=1}^{T}\left[\ln(h_{it}) + \frac{a_{it}^2}{h_{it}}\right] + \text{constant}\right)
\end{aligned}
\tag{33}
$$

Comparing (33) with the log-likelihood (16) in the univariate case we see that the log-likelihood in (33) is the sum of the log-likelihoods of the univariate GARCH equations of n assets, meaning that the parameters of the different univariate models may be separately determined. From this first step, the parameter set $\boldsymbol{\phi} = \boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_n$ is estimated. When $\boldsymbol{\phi}$ is estimated, also the conditional variance h_{it} is estimated for each asset $i = 1, .., n$, and $\epsilon_t = \boldsymbol{D}_t^{-1/2}\boldsymbol{a}_t$ and $\overline{\boldsymbol{Q}} = \mathrm{E}[\epsilon_t\epsilon_t^T]$ can be estimated.

After the first step only the parameters a and b are unknown. These parameters are estimated in the second step.

5.1.2 Step two

In the second step, $\boldsymbol{\psi} = (a, b)$ is estimated using the correctly specified log-likelihood in (32), given the estimated parameters from step one. The second stage quasi-likelihood function is then:

$$
\begin{aligned}
\ln(L_2(\boldsymbol{\psi})) &= -\frac{1}{2}\sum_{t=1}^{T}\Big(n\ln(2\pi) + 2\ln(|\boldsymbol{D}_t|) + \ln(|\boldsymbol{R}_t|) + \boldsymbol{a}_t^T\boldsymbol{D}_t^{-1}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{a}_t\Big)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\Big(n\ln(2\pi) + 2\ln(|\boldsymbol{D}_t|) + \ln(|\boldsymbol{R}_t|) + \boldsymbol{\epsilon}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{\epsilon}_t\Big)
\end{aligned}
\tag{34}
$$

Since $\boldsymbol{D}_t$ is constant when conditioning on the parameters from step one, we can exclude the constant terms and maximize:

$$
\ln(L_2^*(\boldsymbol{\psi})) = -\frac{1}{2}\sum_{t=1}^{T}\Big(\ln(|\boldsymbol{R}_t|) + \boldsymbol{\epsilon}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{\epsilon}_t\Big)
$$

It can be shown under certain conditions that the pseudo-maximum-likelihood method yields consistent and asymptotically normal estimators [11]. A full maximum likelihood estimation is considered in [14] as well as the two-step procedure, and they found that both provided very similar results.

5.2 Multivariate Student's *t*-distributed errors

There exists many candidates for the multivariate generalization of the univariate Student's t-distribution. In this theis the most commonly used distribution is considered.

When the standardized errors, $\boldsymbol{z}_t$, are multivariate Student's t-distributed, the joint density of $z_1, ..., z_T$ is:

$$f(\boldsymbol{z}_t|\nu) = \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})[\pi(\nu-2)]^{n/2}} \left[1 + \frac{\boldsymbol{z}_t^T \boldsymbol{z}_t}{\nu - 2}\right]^{-\frac{n+\nu}{2}}$$

where $\Gamma(\ \cdot\ )$ is the Gamma function.

Again, by using the transformation rule, the likelihood function of $\boldsymbol{a}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t$ is:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})[\pi(\nu-2)]^{n/2}|\boldsymbol{H}_t|^{1/2}} \left[1 + \frac{\boldsymbol{a}_t^T \boldsymbol{H}_t^{-1} \boldsymbol{a}_t}{\nu - 2}\right]^{-\frac{n+\nu}{2}}$$

where $\boldsymbol{\theta}$ denotes the parameters of the model.

The log-likelihood is obtained by taking the logarithm and substituting $\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t$:

$$\begin{aligned} \ln(L(\boldsymbol{\theta})) = \sum_{t=1}^{T} \Bigg(&\ln\left[\Gamma(\frac{\nu+n}{2})\right] - \ln\left[\Gamma(\frac{\nu}{2})\right] - \frac{n}{2}\ln\left[\pi(\nu-2)\right] - \frac{1}{2}\ln\left[|\boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t|\right] \\ &- \frac{\nu+n}{2}\ln\left[1 + \frac{\boldsymbol{a}_t^T\boldsymbol{D}_t^{-1}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{a}_t}{\nu-2}\right]\Bigg) \end{aligned} \tag{35}$$

As for the Gaussian, standardized errors in Section 5.1 $\boldsymbol{\theta}$ is divided in two groups; $(\boldsymbol{\phi}, \boldsymbol{\psi}) = (\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_n, \boldsymbol{\psi})$, where $\boldsymbol{\phi}_i = (\alpha_{0i}, \alpha_{1i}, ..., \alpha_{qi}, \beta_{1i}, ..., \beta_{pi})$ are the parameters of the univariate GARCH model for the i^{th} asset series, $i = 1, ..., n$ and $\boldsymbol{\psi} = (a, b, \nu)$.

The optimization of (35) is difficult. Hence, also in this case the parameters are obtained in two steps. In the first step, the parameter $\boldsymbol{\phi}$ is estimated assuming that the standardized errors are Gaussian distributed, while the parameter $\boldsymbol{\psi}$ is estimated in the second step using the correct log-likelihood in 35, given the parameter $\boldsymbol{\phi}$.

5.2.1 Step one

Several authors have shown that the change of the error distribution does not virtually affect the parameters, see e.g. [6] and [19]. Hence the parameters $\boldsymbol{\phi} = \boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_n$ of the univariate GARCH models are fitted using the pseudo-maximum-likelihood; assuming the errors to be Gaussian distributed.

Assuming Gaussian distributed errors, the first stage quasi-likelihood is the same as in (33):

$$\ln(L_1(\boldsymbol{\phi})) = \sum_{i=1}^{n}\left(-\frac{1}{2}\sum_{t=1}^{T}\left[\ln(h_{it}) + \frac{a_{it}^2}{h_{it}}\right] + \text{constant}\right)$$

and the parameter set $\boldsymbol{\phi}_i, i = 1, ..., n$ are estimated assuming univariate GARCH models with Gaussian distributed errors.

The parameters that remain to be estimated are a, b and ν. These are estimated in the second step.

5.2.2 Step two

The parameters $\boldsymbol{\psi} = (a, b, \nu)$ are estimated in the second step using the correctly specified log-likelihood in (35), given the estimated parameters in step one. The second stage quasi-likelihood function is:

$$\begin{aligned}\ln(L_2(\boldsymbol{\psi})) &= \sum_{t=1}^{T}\Bigg(\ln\left[\Gamma(\frac{\nu+n}{2})\right] - \ln\left[\Gamma(\frac{\nu}{2})\right] - \frac{n}{2}\ln\left[\pi(\nu-2)\right] - \frac{1}{2}\ln\left[|\boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t|\right] \\ &- \frac{\nu+n}{2}\ln\left[1 + \frac{\boldsymbol{a}_t^T\boldsymbol{D}_t^{-1}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{a}_t}{\nu-2}\right]\Bigg) \\ &= \sum_{t=1}^{T}\Bigg(\ln\left[\Gamma(\frac{\nu+n}{2})\right] - \ln\left[\Gamma(\frac{\nu}{2})\right] - \frac{n}{2}\ln\left[\pi(\nu-2)\right] - \frac{1}{2}\ln\left[|\boldsymbol{R}_t|\right] \\ &- \ln\left[|\boldsymbol{D}_t|\right] - \frac{\nu+n}{2}\ln\left[1 + \frac{\boldsymbol{\epsilon}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{\epsilon}_t}{\nu-2}\right]\Bigg)\end{aligned}$$

Since $\boldsymbol{D}_t$ is constant when conditioning on the parameters from step one, we can exclude the constant term and maximize:

$$\ln(L_2^*(\boldsymbol{\psi})) = \sum_{t=1}^{T}\left(\ln\left[\Gamma(\frac{\nu+n}{2})\right]-\ln\left[\Gamma(\frac{\nu}{2})\right]-\frac{n}{2}\ln\left[\pi(\nu-2)\right]-\frac{1}{2}\ln\left[|\boldsymbol{R}_t|\right]-\frac{\nu+n}{2}\ln\left[1+\frac{\boldsymbol{\epsilon}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{\epsilon}_t}{\nu-2}\right]\right)$$

5.3 Multivariate skew Student's t-distributed errors

Like for the multivariate Student's t-distribution, there also exists also many candidates for the multivariate skew Student's t-distribution. Here we use Azzalini's skew Student's t-distribution described in [4].

When the standardized errors, $\boldsymbol{z}_t$, are multivariate skew Student's t-distributed, the joint distribution of $z_1, ..., z_T$ is:

$$f(\boldsymbol{z}_t|\nu, \boldsymbol{\varsigma}) = \prod_{t=1}^{T} 2t_d(\boldsymbol{z}_t; \nu, \boldsymbol{\varsigma})T_1\left\{\boldsymbol{\delta}^T\boldsymbol{D}^{-1}(\boldsymbol{z}_t - \boldsymbol{\xi})\left[\frac{\nu+n}{Q_{z_t}+\nu}\right]^{1/2}; \nu+n\right\} \tag{36}$$

where $\boldsymbol{D}$ is the diagonal matrix with the square root of the diagonal elements of $\boldsymbol{\Omega}$ on the diagonal,

$$Q_{z_t} = (\boldsymbol{z}_t - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{z}_t - \boldsymbol{\xi}),$$

$$t_d(\boldsymbol{z}_t; \nu, \boldsymbol{\varsigma}) = \frac{\Gamma(\frac{\nu+n}{2})}{|\boldsymbol{\Omega}|^{1/2} (\pi\nu)^{n/2} \Gamma(\nu/2)} \left[1 + \frac{Q_{z_t}}{\nu}\right]^{-(\nu+n)/2}$$

$T_1(\,\cdot\,; \nu+n)$ denotes the scalar Student's t-distribution with $\nu+n$ degrees of freedom and $\Gamma(\,\cdot\,)$ is the Gamma function. The joint density (36) is well-defined if $\nu > 2$.

We shall denote the skew Student's t-distribution (36):

$$\boldsymbol{Y} \sim St_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta}, \boldsymbol{\nu})$$

Define:

$$\boldsymbol{\varsigma} = \boldsymbol{D}^{-1} \boldsymbol{\delta} \tag{37}$$

[15] shows that Azzalini's skew Student's t-distribution may be standardized to have mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_n$, by letting:

$$\boldsymbol{\Omega} = \begin{cases} \frac{\nu-2}{\nu} \left[\boldsymbol{I}_n + \frac{1}{\boldsymbol{\varsigma}^T\boldsymbol{\varsigma}} \left(-1 + \frac{\pi\Gamma(\frac{\nu}{2})^2(\nu-(\nu-2)\boldsymbol{\varsigma}^T\boldsymbol{\varsigma})}{2\boldsymbol{\varsigma}^T\boldsymbol{\varsigma}(\nu-2)\left[\pi\Gamma(\frac{\nu}{2})^2-(\nu-2)\Gamma(\frac{\nu-1}{2})^2\right]}(-1+K)\right)\boldsymbol{\varsigma}\boldsymbol{\varsigma}^T\right] & \text{for } \boldsymbol{\varsigma} \neq \mathbf{0} \\ \frac{\nu-2}{\nu}\boldsymbol{I}_n & \text{for } \boldsymbol{\varsigma} = \mathbf{0} \end{cases} \tag{38}$$

Here

$$K = \sqrt{1 + \frac{4\nu(\nu-2)\left[\pi\Gamma(\frac{\nu}{2})^2 - (\nu-2)\Gamma(\frac{\nu-1}{2})^2\right]\boldsymbol{\varsigma}^T\boldsymbol{\varsigma}}{\pi\Gamma(\frac{\nu}{2})^2(\nu-(\nu-2)\boldsymbol{\varsigma}^T\boldsymbol{\varsigma})^2}}$$

and

$$\boldsymbol{\xi} = -\sqrt{\frac{\nu}{\pi}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)} \frac{\boldsymbol{\Omega}\boldsymbol{\varsigma}}{\sqrt{1+\boldsymbol{\varsigma}^T\boldsymbol{\Omega}\boldsymbol{\varsigma}}} \tag{39}$$

By using the transformation rule, the likelihood function of $\boldsymbol{a}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t$ is:

$$L(\boldsymbol{\theta}|\boldsymbol{F}_t) = \prod_{t=1}^{T} 2t_d(\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t; \nu, \boldsymbol{\varsigma}) T_1 \left\{\boldsymbol{\delta}^T \boldsymbol{D}^{-1}(\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t - \boldsymbol{\xi}) \left[\frac{\nu+n}{Q_{a_t}+\nu}\right]^{1/2}; \nu+n\right\} \frac{1}{|\boldsymbol{H}_t|^{1/2}}$$

where

$$Q_{a_t} = (\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t - \boldsymbol{\xi}),$$

and $\boldsymbol{\theta}$ denotes the parameters of the model.

We get the log-likelihood by taking the logarithm and substituting $\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t$:

$$\begin{aligned}\ln(L(\boldsymbol{\theta})) = \sum_{t=1}^{T} \Bigg(& \ln(2) + \ln\left[\, t_d(\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t; \nu, \varsigma\,)\right] \\ & + \ln\left[T_1\left\{\boldsymbol{\delta}^T \boldsymbol{D}^{-1}(\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t - \boldsymbol{\xi})\left[\frac{\nu+n}{Q_{a_t}+\nu}\right]^{1/2}; \nu+n\right\}\right] - \frac{1}{2}\ln\left[\,|\boldsymbol{H}_t|\,\right]\Bigg)\end{aligned} \tag{40}$$

$\boldsymbol{\theta}$ is divided in two groups; $(\boldsymbol{\phi}, \boldsymbol{\psi}) = (\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_n, \boldsymbol{\psi})$, where $\boldsymbol{\phi}_i = (\alpha_{0i}, \alpha_{1i}, ..., \alpha_{qi}, \beta_{1i}, ..., \beta_{pi})$ are the parameters of the univariate GARCH model for the i^{th} asset series, $i = 1, ..., n$ and $\boldsymbol{\psi} = (a, b, \nu, \varsigma)$.

The optimization of (40) is difficult. Hence the parameters are estimated in two steps. In the first step, the parameter $\boldsymbol{\phi}$ is estimated assuming that the standardized errors are Gaussian distributed, while the parameter $\boldsymbol{\psi}$ is estimated in the second step using the correct log-likelihood in (40), given the parameter $\boldsymbol{\phi}$.

5.3.1 Step one

The parameters $\boldsymbol{\phi}$ is estimated under the assumption of Gaussian distributed errors as discussed in Section 5.2.1. Hence the first stage quasi-likelihood is:

$$\ln(L_1(\boldsymbol{\phi})) = \sum_{i=1}^{n}\left(-\frac{1}{2}\sum_{t=1}^{T}\left[\ln(h_{it}) + \frac{a_{it}^2}{h_{it}}\right] + \text{constant}\right)$$

The parameters remain to be estimated are a, b, ν, ς. These are estimated in the second step.

5.3.2 Step two

The parameters $\boldsymbol{\psi} = (a,\, b,\, \nu,\, \varsigma)$ are estimated in the second step using the correct specified log-likelihood in (40), given the estimated parameters in step one. The second stage quasi-likelihood function is then:

$$\begin{aligned}\ln(L_2(\boldsymbol{\psi})) = & \sum_{t=1}^{T}\Bigg(\ln(2) + \ln\left[\, t_d(\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t; \nu, \varsigma\,)\right] \\ & + \ln\left[\, T_1\left\{\boldsymbol{\delta}^T \boldsymbol{D}^{-1}(\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t - \boldsymbol{\xi})\left[\frac{\nu+n}{Q_{a_t}+\nu}\right]^{1/2}; \nu+n\right\}\,\right] - \frac{1}{2}\ln\left[\,|\boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t|\,\right]\Bigg) \\ = & \sum_{t=1}^{T}\Bigg(\ln(2) + \ln\left[\Gamma(\frac{\nu+n}{2})\right] - \frac{1}{2}\ln\left[\,|\boldsymbol{\Omega}|\,\right] - \frac{n}{2}\ln\left[\,\pi\nu\,\right] - \ln\left[\Gamma(\frac{\nu}{2})\right] - \frac{\nu+n}{2}\ln\left[1 + \frac{Q_{a_t}}{\nu}\right] \\ & + \ln\left[\, T_1\left\{\boldsymbol{\delta}^T \boldsymbol{D}^{-1}(\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t - \boldsymbol{\xi})\left[\frac{\nu+n}{Q_{a_t}+\nu}\right]^{1/2}; \nu+n\right\}\,\right] - \frac{1}{2}\ln\left[\,|\boldsymbol{R}_t|\,\right] - \ln\left[\,|\boldsymbol{D}_t|\,\right]\Bigg)\end{aligned}$$

$\boldsymbol{D}_t$ is constant when conditioning on the parameters from step one. We exclude the constant terms and maximize:

$$\ln(L_2^*(\boldsymbol{\psi})) = \sum_{t=1}^{T} \Bigg(\ln\left[\Gamma(\frac{\nu+n}{2})\right] - \frac{1}{2}\ln\left[\,|\boldsymbol{\Omega}|\,\right] - \frac{n}{2}\ln\left[\,\pi\nu\,\right] - \ln\left[\Gamma(\frac{\nu}{2})\right] - \frac{\nu+n}{2}\ln\left[1+\frac{Q_{a_t}}{\nu}\right]$$
$$+ \ln\left[\,T_1\left\{\boldsymbol{\delta}^T\boldsymbol{D}^{-1}(\boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t - \boldsymbol{\xi})\left[\frac{\nu+n}{Q_{a_t}+\nu}\right]^{1/2};\nu+n\right\}\,\right] - \frac{1}{2}\ln\left[\,|\boldsymbol{R}_t|\,\right]\Bigg)$$

inserting (38) and (39) for $\boldsymbol{\Omega}$ and $\boldsymbol{\xi}$ respectively.

5.4 Estimation problems

Often in multivariate data the choice of start values is extremely important. When the number of parameters to estimate is large, the likelihood function becomes flat, and there is a great danger of reaching a local optimum. To secure that we really have found a global maximum one should run the estimation with many different starting values. One way to choose the starting values is to make a grid of the possible values the parameters may take, and choose the starting values to be the combination of values that yields the highest likelihood.

Another convergence problem can occur if we have outliers in the data. Then the gradient algorithm used for the maximation may hit a boundary. To deal with this problem one should try to remove the outliers. If there is still a problem with convergence after removing the outliers, one should try to change the starting values.

6 Forecasting and Value-at-Risk using DCC-GARCH

In this chapter we will consider forecasting and Value-at-Risk using the DCC-GARCH model with Gaussian, Student's t- and skew Student's t-distributed errors.

6.1 Forecasting

After the parameters of the model are estimated, we might be interested in determine the forecast of the conditional covariance matrix, $\boldsymbol{H}_{t+k} = \boldsymbol{D}_{t+k}\boldsymbol{R}_{t+k}\boldsymbol{D}_{t+k}$, at time $t+k$ when the history up to time t is known. When forecasting the covariance matrix, the forecasts of $\boldsymbol{D}_{t+k}$ and $\boldsymbol{R}_{t+k}$ can be done separately, see e.g. [16].

6.1.1 Step one; forcasting the conditional variances in $\boldsymbol{D}_{t+k}$

The forecasts of the univariate variances in $\boldsymbol{D}_{t+k} = \mathrm{diag}(\sqrt{h_{1,t+k}}, ..., \sqrt{h_{n,t+k}})$ can be done separately for each of the n assets. As seen in Section 2.5 the k-step ahead forecast for the general GARCH(q,p) will be:

$$\mathrm{E}[h_{i,t+k}|F_t] = \alpha_0 + \sum_{j=1}^{max(p,q)} (\alpha_j + \beta_j)\mathrm{E}[h_{i,t+k-j}|F_t]$$

where

$$\mathrm{E}[h_{i,t+k}|F_t] = a_{i,t+k}^2 \quad \text{for } k < 0,\ i = 1,..,n$$

Often the easiest GARCH(1,1) model is adequate. From (18) the k-step ahead forecast for the GARCH(1,1) is:

$$\mathrm{E}[h_{i,t+k}|F_t] = \sum_{i=0}^{k-2} \alpha_0(\alpha_1+\beta_1)^i + (\alpha_1+\beta_1)^{k-1}\mathrm{E}[h_{i,t+1}|F_t]$$

where

$$\mathrm{E}[h_{i,t+1}|F_t] = \alpha_0 + \alpha_1 a_{i,t}^2 + \beta_1 h_{i,t}$$

The memory will decline with exponential rate $(\alpha_1+\beta_1)$. Compared with empirical studies the GARCH(1,1) model has been criticized to have too short memory, especially with high frequency data [16].

The forecast of the conditional variance is:

$$E[\boldsymbol{D}_{t+k}|F_t] = \mathrm{diag}(\sqrt{\mathrm{E}[h_{1,t+k}|F_t]}, ..., \sqrt{\mathrm{E}[h_{n,t+k}|F_t]})$$

6.1.2 Step two; forecasting the conditional correlation matrix $\boldsymbol{R}_{t+k}$

The elements in the conditional correlation matrix, $\boldsymbol{R}_{t+k}$, are not themselves forecasts, but they are the ratio of the forecast of the conditional covariance to the square root of the product of the forecasts of the conditional variances, i.e. $\hat{\rho}_{ij} = \frac{\hat{q}_{ij}}{\hat{q}_{ii}\hat{q}_{jj}}$, where $\hat{q}_{ij}, \hat{q}_{ii}$ and $\hat{q}_{jj}$ are the forecast elements in $\boldsymbol{Q}_{t+k}$. Thus unbiased forecasts are not easily computed.

The expectation of $\boldsymbol{Q}_{t+k}$ is:

$$\begin{cases} \mathrm{E}[\boldsymbol{Q}_{t+1}|F_t] = (1-a-b)\overline{\boldsymbol{Q}} + a\epsilon_t\epsilon_t^T + b\boldsymbol{Q}_t & \text{for } k=1 \\ \mathrm{E}[\boldsymbol{Q}_{t+k}|F_t] = (1-a-b)\overline{\boldsymbol{Q}} + a\mathrm{E}[\epsilon_{t+k-1}\epsilon_{t+k-1}^T|F_t] + b\mathrm{E}[\boldsymbol{Q}_{t+k-1}|F_t] & \text{for } k>1 \end{cases} \tag{41}$$

where $\mathrm{E}[\epsilon_{t+k-1}\epsilon_{t+k-1}^T|F_t] = \mathrm{E}[\boldsymbol{R}_{t+k-1}|F_t] = \mathrm{E}[\boldsymbol{Q}_{t+k-1}^{*-1}\boldsymbol{Q}_{t+k-1}\boldsymbol{Q}_{t+k-1}^{*-1}|F_t]$.

Since $\mathrm{E}[\boldsymbol{Q}_{t+k-1}^{*-1}\boldsymbol{Q}_{t+k-1}\boldsymbol{Q}_{t+k-1}^{*-1}|F_t]$ is unknown, we cannot directly compute the k-step ahead forecast in (41). However, there exists two methods that approximates this forecast:

1. Method 1; Assumes that $\mathrm{E}[\epsilon_{t+i}\epsilon_{t+i}^T|F_t] \approx \mathrm{E}[\boldsymbol{Q}_{t+i}|F_t]$ for $i = 1, ..., k$
2. Method 2; Assumes that $\overline{\boldsymbol{R}} \approx \overline{\boldsymbol{Q}}$ and $\mathrm{E}[\boldsymbol{R}_{t+i}|F_t] \approx \mathrm{E}[\boldsymbol{Q}_{t+i}|F_t]$ for $i = 1, ..., k$

Method 1

In this method we assume that $\mathrm{E}[\epsilon_{t+i}\epsilon_{t+i}^T|F_t] \approx \mathrm{E}[\boldsymbol{Q}_{t+i}|F_t]$ for $i = 1, ..., k$.

The expectation of $\boldsymbol{Q}_{t+k}$ for $k > 1$ is:

$$\begin{aligned} \mathrm{E}[\boldsymbol{Q}_{t+k}|F_t] &= (1-a-b)\overline{\boldsymbol{Q}} + aE[\epsilon_{t+k-1}\epsilon_{t+k-1}^T] + b\mathrm{E}[\boldsymbol{Q}_{t+k-1}] \\ &\approx (1-a-b)\overline{\boldsymbol{Q}} + (a+b)E[\boldsymbol{Q}_{t+k-1}] \\ &\approx (1-a-b)\overline{\boldsymbol{Q}} + (a+b)\left[(1-a-b)\overline{\boldsymbol{Q}} + (a+b)\mathrm{E}[\boldsymbol{Q}_{t+k-2}]\right] \\ &= (1-a-b)\overline{\boldsymbol{Q}} + (1-a-b)\overline{\boldsymbol{Q}}(a+b) + (a+b)\mathrm{E}[\boldsymbol{Q}_{t+k-2}|F_t] \\ &\approx ... \\ &\approx \sum_{i=0}^{k-2}(1-a-b)\overline{\boldsymbol{Q}}(a+b)^i + (a+b)^{k-1}\mathrm{E}[\boldsymbol{Q}_{t+1}|F_t] \\ &= (1-(a+b)^{k-1})\overline{\boldsymbol{Q}} + (a+b)^{k-1}\mathrm{E}[\boldsymbol{Q}_{t+1}|F_t] \\ &= \hat{\boldsymbol{Q}}_{t+k} \end{aligned}$$

where $\mathrm{E}[\boldsymbol{Q}_{t+1}|F_t] = (1-a-b)\overline{\boldsymbol{Q}} + a\epsilon_t\epsilon_t^T + b\boldsymbol{Q}_t$ from (41).

Then

$$\hat{\boldsymbol{R}}_{t+k} = \mathrm{E}[\boldsymbol{R}_{t+k}|F_t] \approx \hat{\boldsymbol{Q}}_{t+k}^{*-1}\hat{\boldsymbol{Q}}_{t+k}\hat{\boldsymbol{Q}}_{t+k}^{*-1} \tag{42}$$

where $\hat{\boldsymbol{Q}}^*_{t+k}$ is a diagonal matrix with the square root of the diagonal elements of $\hat{\boldsymbol{Q}}_{t+k}$.

A feature to notice is that $\hat{\boldsymbol{Q}}_{t+k}$ decay with ratio $(a+b)$.

Method 2

In this method we assume that $\overline{\boldsymbol{R}} \approx \overline{\boldsymbol{Q}}$ and $\mathrm{E}[\boldsymbol{R}_{t+i}|F_t] \approx \mathrm{E}[\boldsymbol{Q}_{t+i}|F_t]$ for $i = 1, ..., k$.

The expectation of $\boldsymbol{R}_{t+k}$ for $k > 1$ is:

$$\begin{aligned}
\mathrm{E}[\boldsymbol{R}_{t+k}|F_t] &\approx \mathrm{E}[\boldsymbol{Q}_{t+k}|F_t] \\
&= (1-a-b)\overline{\boldsymbol{Q}} + a\mathrm{E}[\boldsymbol{R}_{t+k-1}|F_t] + b\mathrm{E}[\boldsymbol{Q}_{t+k-1}|F_t] \\
&\approx (1-a-b)\overline{\boldsymbol{R}} + (a+b)\mathrm{E}[\boldsymbol{R}_{t+k-1}|F_t] \\
&\approx (1-a-b)\overline{\boldsymbol{R}} + (a+b)\left[(1-a-b)\overline{\boldsymbol{R}} + (a+b)\mathrm{E}[\boldsymbol{R}_{t+k-2}|F_t]\right] \\
&= (1-a-b)\overline{\boldsymbol{R}} + (1-a-b)\overline{\boldsymbol{R}}(a+b) + (a+b)\mathrm{E}[\boldsymbol{R}_{t+k-2}|F_t] \\
&\approx ... \\
&\approx \sum_{i=0}^{k-2}(1-a-b)\overline{\boldsymbol{R}}(a+b)^i + (a+b)^{k-1}\mathrm{E}[\boldsymbol{R}_{t+1}|F_t] \\
&= (1-(a+b)^{k-1})\overline{\boldsymbol{R}} + (a+b)^{k-1}\mathrm{E}[\boldsymbol{R}_{t+1}|F_t]
\end{aligned}$$

where $\mathrm{E}[\boldsymbol{R}_{t+1}|F_t] \approx \hat{\boldsymbol{Q}}^{*-1}_{t+1}\hat{\boldsymbol{Q}}_{t+1}\hat{\boldsymbol{Q}}^{*-1}_{t+1}$, $\hat{\boldsymbol{Q}}_{t+1} = (1-a-b)\overline{\boldsymbol{Q}} + a\epsilon_t\epsilon_t^T + b\boldsymbol{Q}_t$ and $\overline{\boldsymbol{R}} = \overline{\boldsymbol{Q}}^*\overline{\boldsymbol{Q}}\,\overline{\boldsymbol{Q}}^*$, where $\overline{\boldsymbol{Q}}^*$ is a diagonal matrix with the square root of the diagonal elements of $\overline{\boldsymbol{Q}}$ on the diagonal.

By using the notation $\hat{\boldsymbol{H}}_{t+k} = \mathrm{E}[\boldsymbol{H}_{t+k}|F_t]$, $\hat{\boldsymbol{R}}_{t+k} = \mathrm{E}[\boldsymbol{R}_{t+k}|F_t]$ and $\hat{\boldsymbol{D}}_{t+k} = \mathrm{E}[\boldsymbol{D}_{t+k}|F_t]$ we finally calculate $\hat{\boldsymbol{H}}_{t+k} = \hat{\boldsymbol{D}}_{t+k}\hat{\boldsymbol{R}}_{t+k}\hat{\boldsymbol{D}}_{t+k}$.

A feature to notice is that $\hat{\boldsymbol{R}}_{t+k}$ decay with ratio $(a+b)$.

An empirical study by Engle and Sheppard [11] shows that Method 2 has better bias properties for almost all correlation matrices.

For both methods, the forecast of the conditional correlation matrix, $\hat{\boldsymbol{R}}_{t+k}$, will in the long run converge to the unconditional correlation matrix of the standardized residuals, $\overline{\boldsymbol{Q}}$.

6.2 Value-at-Risk

In financial applications one is often interested in the probability that the return of a portfolio falls below a certain limit, or equivalently, the smallest number l such that the probability that the return is lower than l is α. The number l is denoted Value-at-Risk with confidence level α. Hence to determine VaR one first has to determine the distribution of the portfolio return, $p_t = w^T r_t$. In this section we show this is done for the three distributions of the error; multivariate Gaussian, Student's t and skew Student's t.

6.2.1 Multivariate Gaussian distributed errors

We want to determine the distributions of $\boldsymbol{a}_t$ and p_t. We use that a linear combination of $\boldsymbol{z}_t \sim \mathrm{N}_d(\mathbf{0}, \boldsymbol{I}_n)$, is also Gaussian distributed. Since $\mathrm{E}[\boldsymbol{a}_t]$=0 and $\mathrm{Cov}[\boldsymbol{a}_t] = \boldsymbol{H}_t$, we have that:

$$\boldsymbol{a}_t \sim \mathrm{N}_d(\mathbf{0}, \boldsymbol{H}_t)$$

For the portfolio return, $\boldsymbol{p}_t = \boldsymbol{\omega}^T \boldsymbol{r}_t$ we get the expectation:

$$\begin{aligned} \mathrm{E}[p_t] &= \boldsymbol{\omega}^T \mathrm{E}[\boldsymbol{\mu}_t] + \boldsymbol{\omega}^T \mathrm{E}[\boldsymbol{a}_t] \\ &= \boldsymbol{\omega}^T \mathrm{E}[\boldsymbol{\mu}_t] \end{aligned} \tag{43}$$

and the covariance:

$$\begin{aligned} \mathrm{Cov}[p_t] &= \mathrm{Cov}[\boldsymbol{\omega}^T \boldsymbol{r}_t] \\ &= \mathrm{Cov}[\boldsymbol{\omega}^T (\boldsymbol{\mu}_t + \boldsymbol{a}_t)] \\ &= \mathrm{Cov}[\boldsymbol{\omega}^T \boldsymbol{a}_t] \\ &= \boldsymbol{\omega}^T \mathrm{Cov}[\boldsymbol{a}_t] (\boldsymbol{\omega}^T)^T \\ &= \boldsymbol{\omega}^T \boldsymbol{H}_t \boldsymbol{\omega} \end{aligned} \tag{44}$$

Hence:

$$p_t \sim N_1(\boldsymbol{w}^T \mathrm{E}[\boldsymbol{\mu}_t], \boldsymbol{w}^T \boldsymbol{H}_t \boldsymbol{w})$$

6.2.2 Multivariate Student's t-distributed errors

To determine the distribution of $\boldsymbol{a}_t$ and p_t, we use that a linear combination of $\boldsymbol{z}_t \sim t_{\nu,d}(\mathbf{0}, \boldsymbol{I}_n)$, is also Student's t-distributed. Since $\mathrm{E}[\boldsymbol{a}_t]$=0 and $\mathrm{Cov}[\boldsymbol{a}_t] = \boldsymbol{H}_t$, we have that:

$$a_t \sim t_{\nu,d}(\mathbf{0}, \boldsymbol{H}_t)$$

The expectation and the covariance of p_t is the same as given in (43) and (44), respectively.

Hence:

$$p_t \sim t_\nu(\boldsymbol{w}^T \mathrm{E}[\boldsymbol{\mu}_t], \boldsymbol{w}^T \boldsymbol{H}_t \boldsymbol{w})$$

6.2.3 Multivariate skew Student's t-distributed errors

To determine the distribution of $\boldsymbol{a}_t$ and p_t, we use the following result:

Let $\boldsymbol{X}_t \sim St_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta}, \nu)$ and $\boldsymbol{A}$ be an $m \times n$ constant, matrix of rank m. Then a linear combination $\boldsymbol{Y}_t = \boldsymbol{b} + \boldsymbol{A}\boldsymbol{X}_t$ is $St_d(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Omega}}, \tilde{\boldsymbol{\delta}}, \tilde{\nu})$. The parameters $\tilde{\xi}, \tilde{\Omega}, \tilde{\delta}, \tilde{\nu}$ is determined by Azzalini and Capitanio [4]:

$$\begin{aligned}
\tilde{\boldsymbol{\xi}} &= \boldsymbol{b} + \boldsymbol{A}\boldsymbol{\xi} \\
\tilde{\boldsymbol{\Omega}} &= \boldsymbol{b} + \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}^T \\
\tilde{\nu} &= \nu \\
\tilde{\boldsymbol{\delta}} &= \frac{\tilde{\boldsymbol{D}}\tilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{B}^T\boldsymbol{\delta}}{[1 + \boldsymbol{\delta}^T(\boldsymbol{C} - \boldsymbol{B}\tilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{B}^T)\boldsymbol{\delta}]^{1/2}}
\end{aligned} \tag{45}$$

where

$$\boldsymbol{B} = \boldsymbol{D}^{-1}\boldsymbol{\Omega}\boldsymbol{A}^T \quad \text{and} \quad \boldsymbol{C} = \boldsymbol{D}^{-1}\boldsymbol{\Omega}\boldsymbol{D}^{-1}$$

$\boldsymbol{D}$ is the diagonal matrix with the square root of the diagonal elements of $\boldsymbol{\Omega}$ on the diagonal as described in Section 5.3, and $\tilde{\boldsymbol{D}}$ is the diagonal matrix with the square root of the diagonal elements of $\tilde{\boldsymbol{\Omega}}$ on the diagonal.

When $\boldsymbol{X}_t = \boldsymbol{z}_t$, the distribution of $\boldsymbol{Y}_t = \boldsymbol{a}_t$ is:

$$\boldsymbol{a}_t \sim St_d(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Omega}}, \tilde{\boldsymbol{\delta}}, \tilde{\nu})$$

by setting $\boldsymbol{b} = 0$ and $\boldsymbol{A} = \boldsymbol{H}_t^{1/2}$ in the expression for $\tilde{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{\Omega}}$ in (45).

And when $\boldsymbol{X}_t = \boldsymbol{a}_t$, the distribution of $\boldsymbol{Y}_t = \boldsymbol{p}_t$ is:

$$p_t \sim St_1(\xi^*, \Omega^*, \delta^*, \nu^*)$$

by setting $b = \boldsymbol{\omega}^T\boldsymbol{c} + \boldsymbol{\omega}^T\boldsymbol{K}\boldsymbol{r}_{t-1}$ and $\boldsymbol{A} = \boldsymbol{\omega}^T$ in the expression for ξ^* and Ω^* in (45).

7 Goodness of Fit

To check whether the fitted DCC-GARCH model is appropriate we may check the goodness of the errors $\boldsymbol{z}_t$ separately for each asset series and the goodness of the multivariate fit.

7.1 Goodness of fit of marginals

There are several ways to check if the model fits the data. First we will consider the fit of the marginals. If the model is suitable, the standardized errors, z_t, should be iid. This can be checked by several different tests:

1. **Plot of the errors**

   The errors should look random, if they are iid.

2. **The sample autocorrelation function**

   The 95% confidence interval for the acf of z_t can be computed as $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$, where n is the number of observations. The acf should be outside this interval for 5% of the lags if z_t is iid, and the lags that fall out should be random.

3. **Ljung-Box test**

   The Ljung-Box test checks whether the data are autocorrelated based on a number of lags, m. We want to test whether the autocorrelations, $\gamma_1, ..., \gamma_m$, of z_t is 0 or not. The test can be defined as:

   H_0: $\gamma_1 = ... = \gamma_m = 0$
   H_a: At least one $\gamma_i \neq 0, i = 1, ..., m$

   The test statistic is:

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

   where n is the sample size, $\hat{\rho}_i$ is the sample correlation of z_t^2, at lag i, and m is the number of lags being tested. When n is large, then Q_m is asymptotically distributed as a chi-squared distribution with m degrees of freedom under the null hypothesis. Then for a significance level α, we reject H_0 if

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

   where $\chi^2_{1-\alpha,m}$ is the α-quantile of the chi-square distribution with m degrees of freedom.

   If we accept H_0, we do not reject the hypothesis that the errors are random. In practice, the selection of the number of lags, m, may affect the performance of Q_m. Therefore are often several values of m tested.

4. **Turning point test**

   If $z_1, ..., z_n$ is a sequence of standardized errors, we say that there is a turning point at time i, $1<i<n$, if $z_{i-1} < z_i$ and $z_i > z_{i+1}$, or if $z_{i-1} > z_i$ and $z_i < z_{i+1}$. If T is the number of turning points of an iid sequence of length n, then the probability of a turning point at time i is $2/3$, and the expected value of T is:

$$\mathrm{E}[T] = 2(n-2)/3$$

Further it can be shown that the variance is:

$$\mathrm{Var}[T] = (16n-29)/90$$

If n is large, T will be approximately N(E[T], Var[T]).

5. **Difference-sign test**

   For this test we count the number of times, S, the differenced series $z_i - z_{i-1} > 0$. For an iid sequence the expected number of S is

   $$\mathrm{E}[S] = (n-1)/2$$

   and the variance can be shown to be

   $$\mathrm{E}[S] = (n+1)/12$$

   For large n, S is approximately N(E[S],Var[S]).

6. **Q-Q plot**

   Q-Q plot is a graphical method to check whether a data set is from a given distribution. One plots the assumed distribution on the horizontal axis and the quantiles of the data set on the vertical axis. If the data set is from the assumed distribution, then the plot will approximately be a straight line, especially near the center. If one has significant deviations from linearity, the null hypothesis of the assumed distribution for the data set is rejected.

7.2 Goodness of multivariate fit

In Section 7.1 we have described how to validate univariate fit. The multivariate distribution does not have to fit well, even though the marginals do. Hence it is important also to validate the multivariate fit. Assessing multivariate fit is however difficult, and there exists only a few statistical goodness of fit in the literature. The first test considered, the Baringhaus-Franz multivariate test, is an in-sample test, while the last considered, the Kupiec and Christoffersen's Markov test, are out-of-sample tests.

1. **Baringhaus-Franz multivariate test**

   The test described in Baringhaus and Franz [5] checks whether to datasets, $\boldsymbol{X}$ and $\boldsymbol{Y}$, are identically distributed or not. The test can be defined as:

   H_0: "$\boldsymbol{X}$ is distributed as $\boldsymbol{Y}$"
   H_a: "$\boldsymbol{X}$ is not distributed as $\boldsymbol{Y}$"

   The test statistic is:

$$T = \frac{mn}{m+n}\left\{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}||\boldsymbol{X}_i - \boldsymbol{Y}_j|| - \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}||\boldsymbol{X}_i - \boldsymbol{X}_j|| - \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}||\boldsymbol{Y}_i - \boldsymbol{Y}_j||\right\}$$

where $||\cdot||$ is the Euclidian distance.

When using this test for validation of the goodness of fit for the DCC-GARCH model, $\boldsymbol{X}_1, ..., \boldsymbol{X}_m$ are vectors of errors, $\boldsymbol{z}_t$, of length d, while $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ are vectors of samples from a given distribution, e.g. multivariate Gaussian, Student's t or skew Student's t.

The critical point is obtained by bootstrapping this statistic with a 95% confidence level. If the observed test statistic T falls inside the confidence interval, the null-hypothesis that the DCC-GARCH model explaines the data well is accepted.

2. **Backtesting of VaR; Kupiec test and Christoffersen's Markov test**

   Kupiec test and Christoffersen's Markov test can also be used as a measure of the multivariate goodness of fit using backtesting for VaR.

   To evaluate the goodness of the VaR estimates, backtesting can be used. We calculate the percentage of times that the observed portfolio returns, p_t, fall below the VaR estimate, and compare that number to the confidence level used. If the observed value of p_t fall outside the confidence interval a violation is said to occur.

   If the model is correct specified there are two properties that must be satisfied:

   (a) The total number of violations must be equal to the expected.

   (b) The violations must be independently distributed over time.

   A test proposed by Kupiec [7] consider the first property, and the Markov test by Christoffersen [2] consider the last property. Christoffersen [2] have also presented a test that consider both properties. However if this test fails, we do not know what part that has failed. Therefore we will use the Kupiec test and Christoffersen's Markov test instead.

The Kupiec test

We want to determine which of the distributions; multivariate Gaussian, Student's t and skew Student's t, that is best by considering the number of violations. This can by done by the Kupiec test.

The number of violations follows a binomial distribution:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where n is the length of the sample, x is the number of violations and p is the probability of getting a violation.

The Kupiec test can be defined by:

H_0: The expected proportion of violations is equal to α
H_a: The expected proportion of violations is not equal to α

Under the null hypothesis the test statistic is:

$$\begin{aligned}\text{Kupiec} &= 2\ln\left\{\frac{\binom{n}{x}p_{obs}^{x}(1-p_{obs})^{n-x}}{\binom{n}{x}\alpha^{x}(1-\alpha)^{n-x}}\right\}\\ &= 2\ln\left[\left(\frac{x}{n}\right)^{x}\left(1-\frac{x}{n}\right)^{n-x}\right]-2\ln\left[\alpha^{x}(1-\alpha)^{n-x}\right]\sim\chi^{2}(1),\quad \text{as } n\to\infty\end{aligned}$$

Here $p_{obs} = \frac{x}{n}$, is the estimated probability, and α is the probability of getting a violation for a given confidence level.

If the estimated probability, p_{obs}, is above the significance level (usually 5%), we accept the model. If the estimated probability is below the significance level, we reject the model and conclude that it is not correct.

Christoffersen's Markov test

Christoffersen's Markov test checks whether violations are independently distributed over time. The test can be defined by:

H_0: The violations are independently distributed over time.
H_a: The violations are not independently distributed over time.

Under the null hypothesis and the test statistic is:

$$\begin{aligned}\text{Christoffersen} &= 2\ln\left\{\frac{(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}}{\alpha^{x}(1-\alpha)^{n-x}}\right\}\\ &= 2\ln\left[(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}\right]-2\ln\left[\alpha^{x}(1-\alpha)^{n-x}\right]\\ &\sim\chi^{2}(1),\quad \text{as } n\to\ \infty\end{aligned}$$

Here n_{ij} is the number of times there is a transition from i to j in a sequence of zeros and ones corresponding to non-violations and violations, i.e. $i, j = 0, 1$. The corresponding probabilities are $\pi_{ij} = n_{ij}/\sum_j n_{ij}$.

Christoffersen's Markov test considers only the dependence from one day to the next, i.e. we assume a Markov property. This is a weakness of this test.

8 Use of DCC-GARCH to Real Data

In this example we will try to fit DCC-GARCH to a three-dimensional data set consisting of European, American and Japanese stocks assuming three different error distributions; multivariate Gaussian, Student's t and skew Student's t. The data consists of 4062 time points.

K. Aas, I. H. Haff and X. K. Dimakos at the Norwegian Computing Center have previously used a CCC-GARCH model to fit these stocks [15]. As discussed in Chapter 3, CCC-GARCH is a simple version of the DCC-GARCH, where the conditional correlation matrix $\boldsymbol{R}_t$ is assumed constant, i.e. $\boldsymbol{R}_t = \boldsymbol{R}$. It is interesting to see whether the more general DCC-GARCH model is a significantly better model for these data. Not all tests have been done with the CCC-GARCH model, we only have results for the Kupiec test. To model the geometric returns, $\boldsymbol{r}_t = \log \boldsymbol{Y}_t - \log \boldsymbol{Y}_{t-1}$, where $\boldsymbol{Y}_t$ is the vector of index values at time t, K. Aas et al. [15] proposed the model:

$$\boldsymbol{r}_t = \boldsymbol{c} + \boldsymbol{K}\boldsymbol{r}_{t-1} + \boldsymbol{a}_t$$

where $\boldsymbol{c} = (c_{eur}, c_{usa}, c_{jpy})$, $\mathrm{E}[\boldsymbol{a}_t] = 0$ and $\mathrm{Cov}[\boldsymbol{a}_t] = \boldsymbol{H}_t$. They modelled $\boldsymbol{H}_t$ assuming the CCC-GARCH model.

To ensure that $\boldsymbol{a}_t$ is the same in the CCC-GARCH and the DCC-GARCH models, we will in this example use the same $\boldsymbol{c}$ and $\boldsymbol{K}$ as in [15], but model $\boldsymbol{a}_t$ with DCC-GARCH instead of CCC-GARCH. The elements in the diagonal matrix $\boldsymbol{D}_t$ are modelled with a univariate GARCH(1,1) model.

The $\boldsymbol{K}$-matrix is on the form:

$$\boldsymbol{K} = \begin{bmatrix} 0 & d_{eur} & 0 \\ 0 & 0 & 0 \\ e_{jpy} & d_{jpy} & 0 \end{bmatrix}$$

From the $\boldsymbol{K}$-matrix we see that the European return depends on the American return the previous day, and the Japanese return depends on both the European and the American returns the previous day. In this way the Japanese returns also depends indirectly on the American return two days earlier. Because of the time zones, the US market is the latest to close within the day. This explains why it affects other markets the next calendar day.

8.1 Data summary statistics

The three series of $\boldsymbol{a}_t$ are shown in Figure 1. We see that all the three series have volatility clustering, i.e. periods with high volatility and periods with low volatility, which indicates that a GARCH model can be used to fit the data.

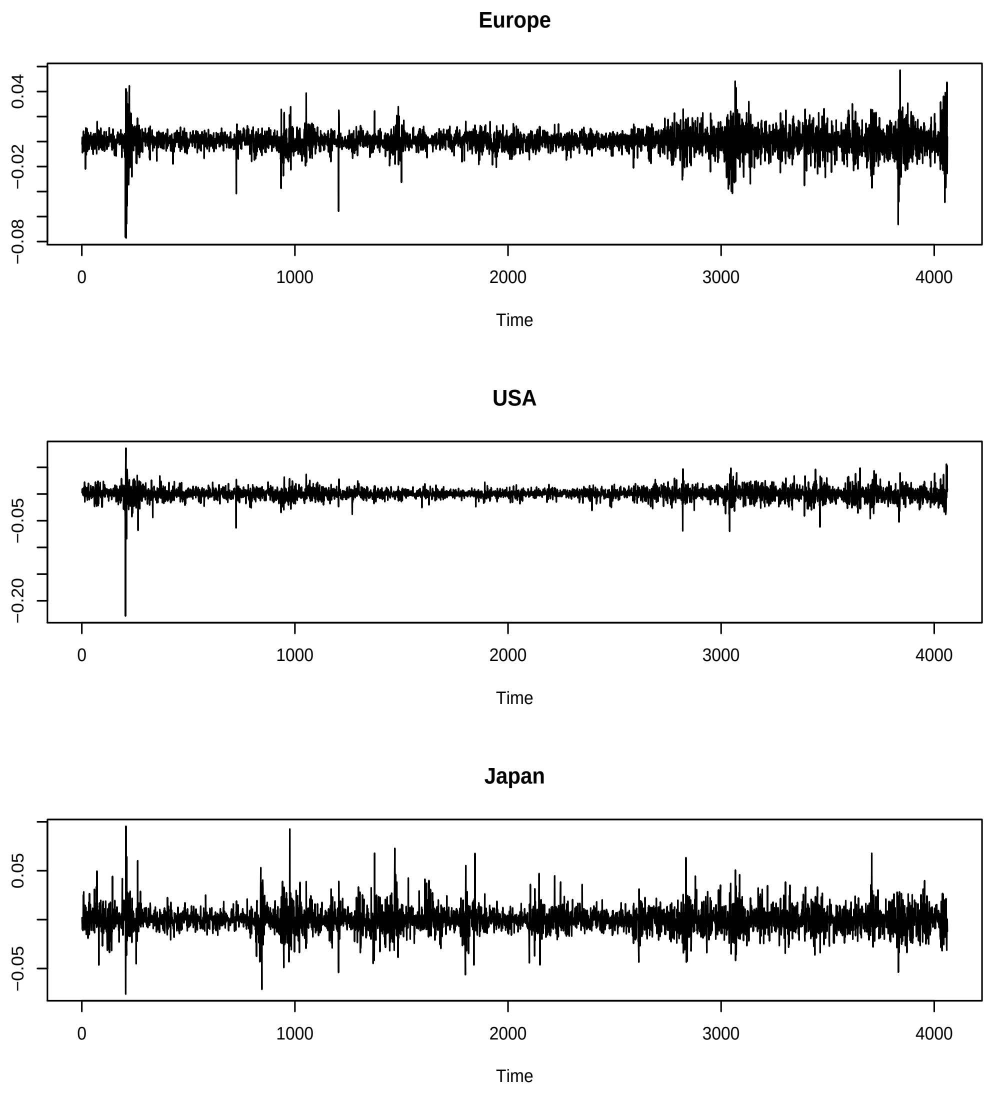

Figure 1: American, European and Japanese a_t series of the period January 1, 1987 to May 28, 2002.

Figure 2 shows the autocorrelation function of $\boldsymbol{a}_t$ and Figure 3 shows the autocorrelation function of $\boldsymbol{a}_t^2$. If $\boldsymbol{a}_t$ is serially uncorrelated, 5% of the lags in the acf-plot in Figure 2 is expected to fall outside the limits (the blue dotted lines). With 100 lags only 5 lags are expected to fall outside. In Figure 2 we see that more than 5 lags fall outside for all the series, but the lags that fall outside does not make a pattern, i.e it is random which lag that fall outside, and most of the lags that fall outside are just barely outside the limits. Hence we can conclude that $\boldsymbol{a}_t$ is approximately uncorrelated. But the lags that fall outside the limits in the acf of $\boldsymbol{a}_t^2$ in Figure 3 do make a pattern. We see that the first lags are greatest, and then the acf decreases. Because the number of lags that fall outside is large and the lags that fall outside make a pattern, $\boldsymbol{a}_t^2$ is not uncorrelated. If $\boldsymbol{a}_t$ is serially independent, $\boldsymbol{a}_t^2$ should be uncorrelated, but it is not, which means that $\boldsymbol{a}_t$ is dependent. Hence a GARCH model is a good choice of modelling $\boldsymbol{a}_t$, because $\boldsymbol{a}_t$ is uncorrelated, but dependent.

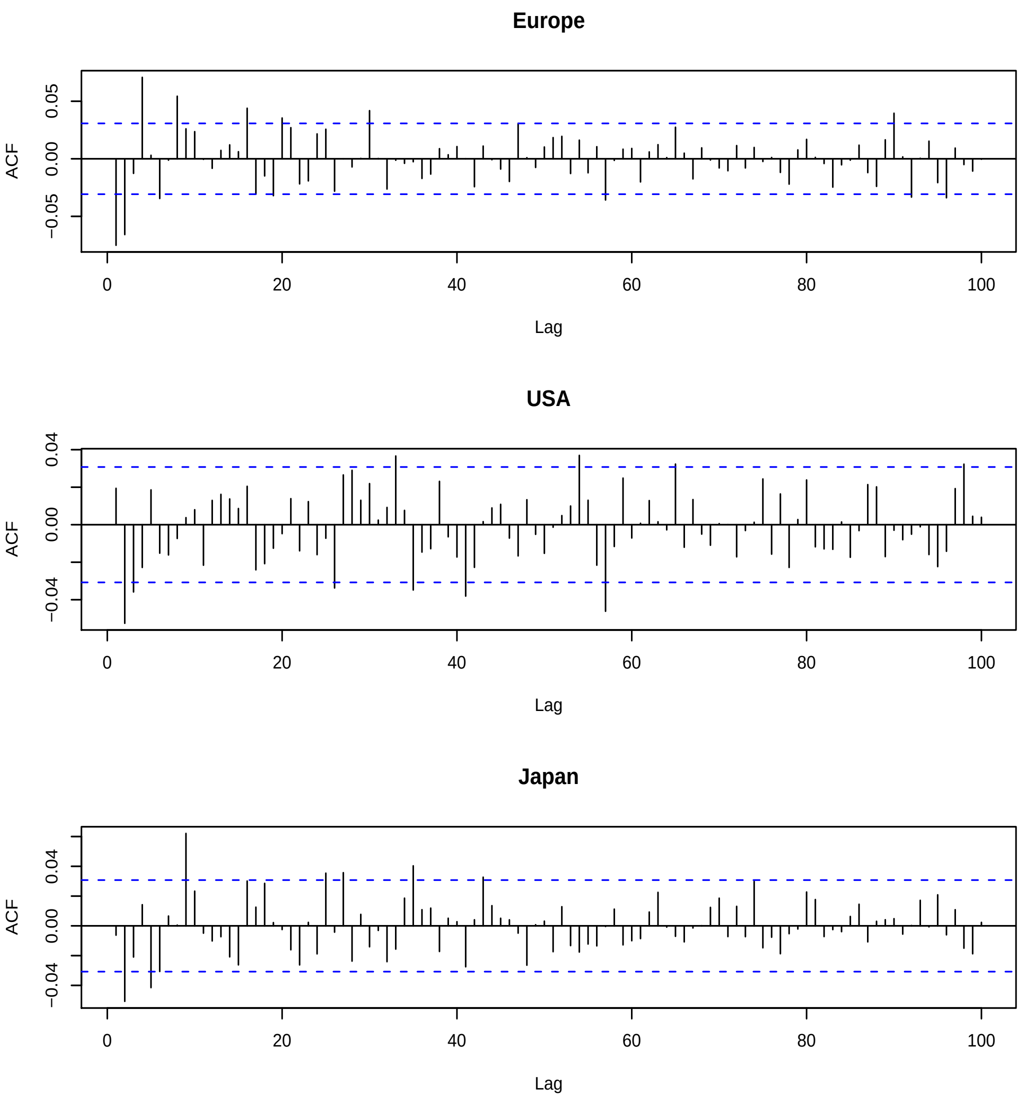

Figure 2: The acf-plots for the American, European and Japanese a_t series.

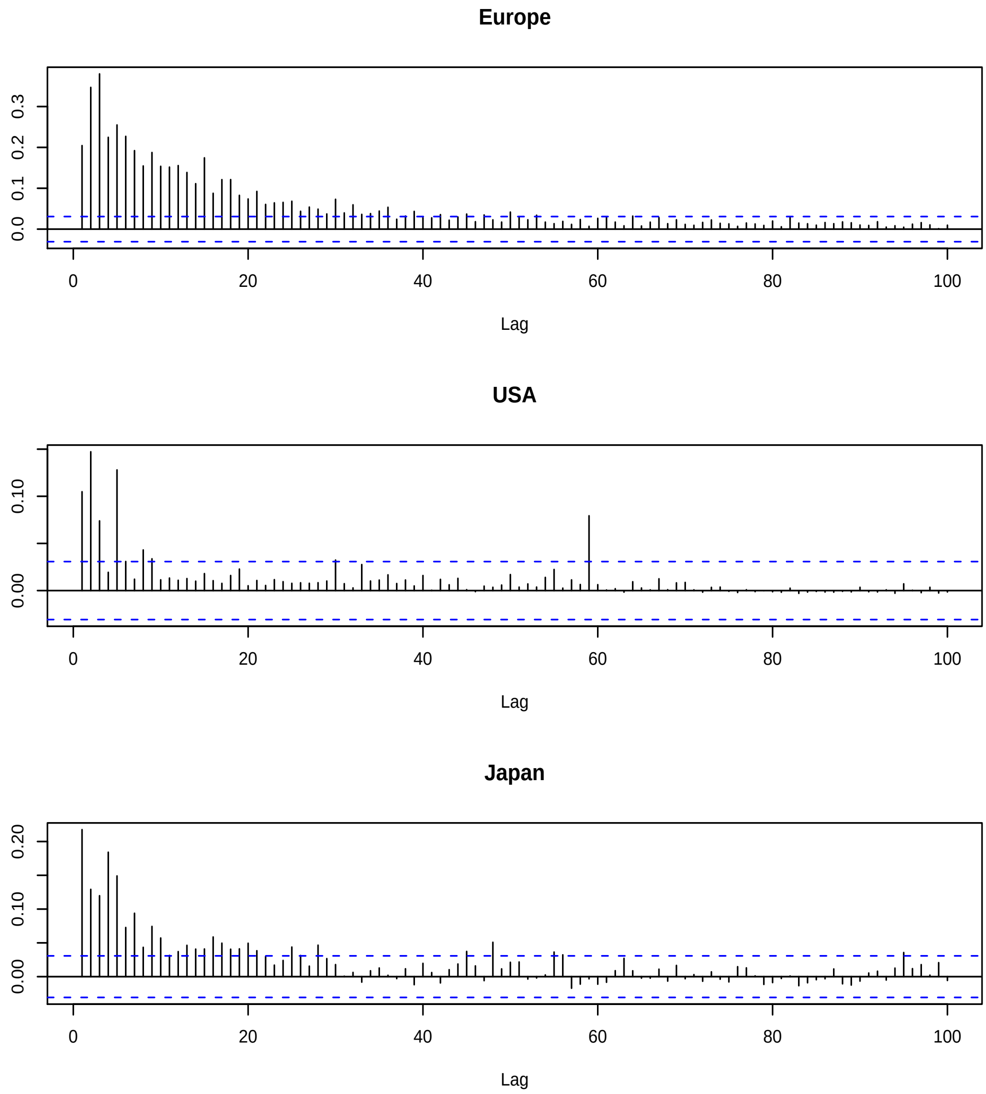

Figure 3: The acf-plots for the American, European and Japanese a_t^2 series.

The excess kurtosis of the European, American and Japanese a_t series are 8.15, 52.2 and 5.30 respectively. Since all excess kurtosis are greater than 0, the distribution of $\boldsymbol{a}_t$ has heavy tails, which is a requirement for the GARCH models as described in Section 2.3.2.

8.2 Parameter estimation

Starting values for the parameters are chosen by calculating the likelihood by different combinations of values of parameters.

Starting value for the matrix $\boldsymbol{Q}_t$ is set to $\boldsymbol{Q}_0 = \overline{\boldsymbol{Q}}$.

The $\boldsymbol{K}$-matrix used is:

$$\boldsymbol{K} = \begin{bmatrix} 0 & 0.362 & 0 \\ 0 & 0 & 0 \\ 0.178 & 0.323 & 0 \end{bmatrix}$$

and $\boldsymbol{c} = (c_{eur}, c_{usa}, c_{jpy}) = (0.000122, 0.000316, -0.0000260)$.

8.2.1 Gaussian distributed errors

First we fit $\boldsymbol{a}_t$ assuming Gaussian distributed errors, $\boldsymbol{z}_t$. When $\boldsymbol{z}_t \sim \mathrm{N}(0,I)$, $a_t \sim \mathrm{N}(0,\boldsymbol{H}_t)$.

The estimated parameters from step one are given in Table 1.

Table 1: Parameters from step 1 when assuming Gaussian distributed errors.

	α_0	α_1	β_1
Europe	2.28e-06	0.116	0.854
USA	1.39e-06	0.091	0.903
Japan	3.99e-06	0.117	0.863

We see that α_0 is small for all the assets. α_1 and β_1 are about the same values for the three different assets.

The estimated parameters from step two are $a = 0.0115$ and $b = 0.948$.

8.2.2 Student's t-distributed errors

Since we, as described in Section 5.2.1, assume Gaussian distributed errors in step one, the estimated parameters in this step are exactly the same as the parameters given in Table 1.

The estimated parameters from step two are $a = 0.00716$, $b = 0.963$ and $\nu = 6.75$. The estimated parameters a and b are close to the parameters estimated with the Gaussian distribution for the errors. The parameter b is a bit larger, and a a bit smaller.

8.2.3 Skew Student's t-distributed errors

Since we, as described in Section 5.2.1, assume Gaussian distributed errors in step one, the estimated parameters in this step are exactly the same as the parameters given in Table 1.

The estimated parameters from step two are $a = 0.00867$, $b = 0.677$, $\nu = 6.68$ and $\boldsymbol{\varsigma} = [-0.604, -0.532, -0.0169]$.

By inserting ν and $\boldsymbol{\varsigma}$ into (38) we compute the dispersion matrix $\boldsymbol{\Omega}$:

$$\boldsymbol{\Omega} = \begin{bmatrix} 0.805 & 0.092 & 0.003 \\ 0.092 & 0.781 & 0.003 \\ 0.003 & 0.003 & 0.701 \end{bmatrix}$$

We see that the elements on the diagonal in the $\boldsymbol{\Omega}$-matrix is close to 1, and small on the off-diagonal.

The skewness vector $\boldsymbol{\delta} = [-0.542, -0.470, -0.0142]$ is obtained by putting $\boldsymbol{\varsigma}$ and $\boldsymbol{D}$ calculated from $\boldsymbol{\Omega}$ into (37).

Since the skewness vector is not equal to the 0-vector, the distribution of the returns is not symmetric.

Finally, inserting ν, $\boldsymbol{\varsigma}$ and $\boldsymbol{\Omega}$ in (39), we get the location vector $\boldsymbol{\xi} = [0.385, 0.339, 0.0108]$.

8.3 Forecasting

When forecasting the covariance matrix, $\boldsymbol{H}_{t+k}$, k steps ahead, the forecasts of $\boldsymbol{D}_{t+k}$ and $\boldsymbol{R}_{t+k}$ can be done separately, as described in Section 6.1. First we will consider the two different methods of forecasting.

8.3.1 Difference between Method 1 and Method 2 of forecasting

As described in Section 6.1, the correlation matrix $\boldsymbol{R}_{t+k}$ may be forcasted using two different methods. We will consider both methods.

A DCC-GARCH model with Gaussian distributed errors, $\boldsymbol{z}_t$, is used to fit the whole data set, $t = 1, ..., 4062$. Forecasts of the off-diagonal of $\boldsymbol{R}_{4062+k}$, $k = 1, ..., 365$, is shown in Figure 4 for both Method 1 (black line) and Method 2 (red line). The blue line is the unconditional correlation matrix, $\overline{\boldsymbol{R}}$. The numbers 1, 2 and 3 stands for Europe, USA and Japan, respectively. For example $\boldsymbol{R}_{4062+k}[1, 2]$ is the correlation between Europe and USA. The diagonal of $\boldsymbol{R}_{t+k}$ is by definition 1 for both methods.

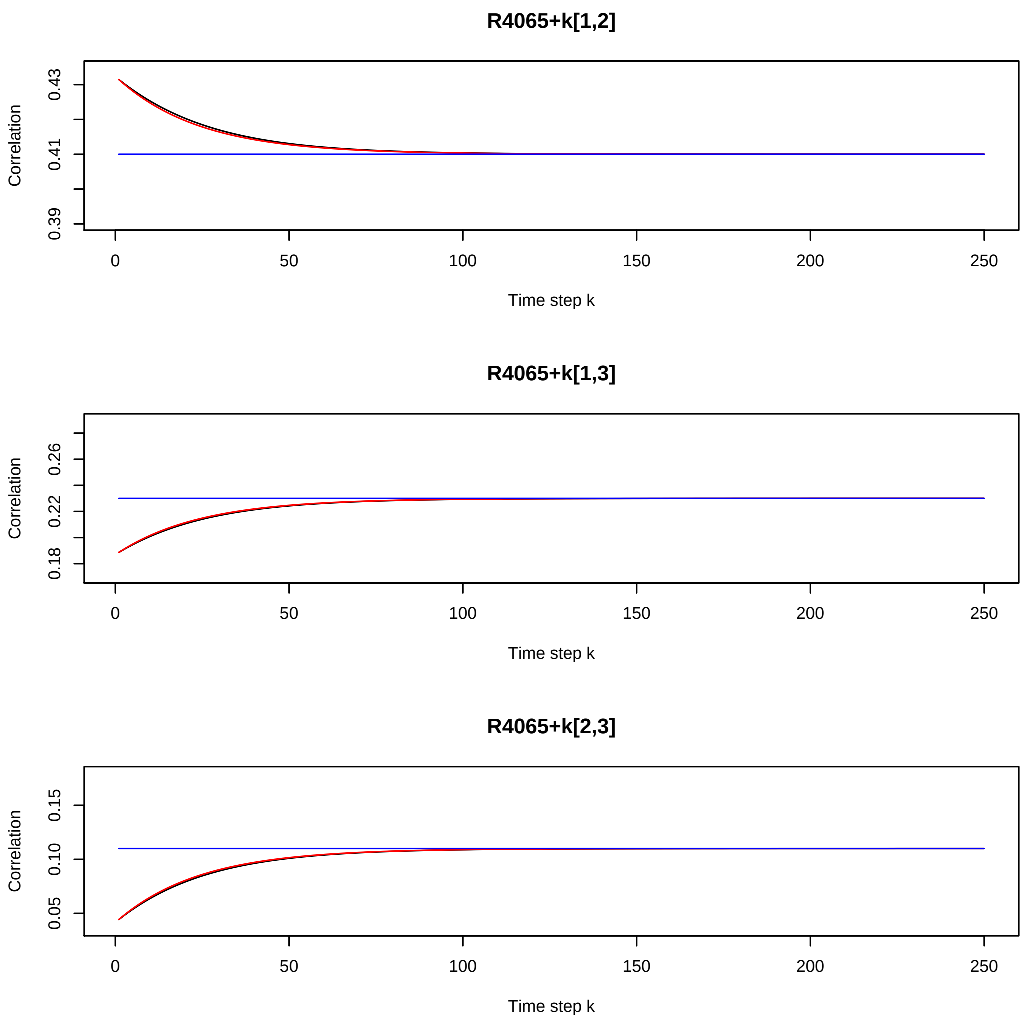

Figure 4: 365 forecasts of the Gaussian model, both Method 1 and Method 2. Black: Method 1, Red: Method 2.

If we look close at Figure 4, we see that Method 2 (red line) is a bit closer to the unconditional correlation (blue line) than Method 1 (black line), but the difference is very small. In this example we see that the forecasts need approximately 80 time steps to converge to the unconditional correlation for both methods.

There are no big difference in the plots of the forecasts of the DCC-GARCH with Gaussian and Student's t-distributed errors, hence the plots of forecasts when we assume Student's t-distributed errors is not shown. This is because the estimated value of a and b are quite similar.

The forecasts of $\boldsymbol{R}_{4062+k}$, $k = 1, ..., 365$, of the DCC-GARCH with skew Students's t-distributed errors is shown in Figure 5. Comparing these forecasts with the forecasts made for the Gaussian assumption in Figure 4, we see that the forecasts of the skew Student's t-distributed errors converge faster to the unconditional value of $\boldsymbol{R}_{4062+k}$. It only need approximately 15 time points to converge. This is because the estimated value of b is smaller for the DCC-GARCH with skew Student's t-distributed errors than for the model with Gaussian distributed errors. As mentioned in Section 6.1, $\boldsymbol{R}_{t+k}$ decay with ratio $(a + b)$.

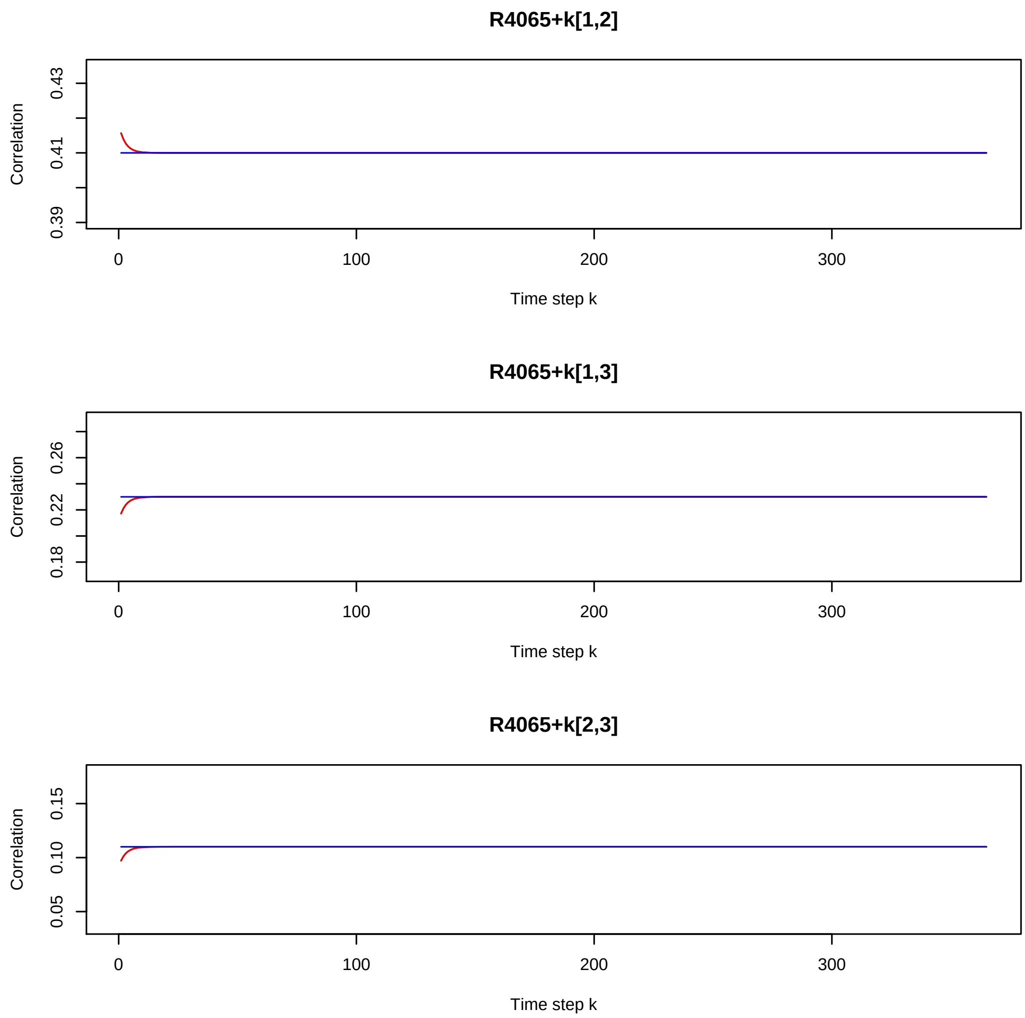

Figure 5: 365 forecasts of the skew Student's t model, both Method 1 and Method 2. Black: Method 1, Red: Method 2.

8.3.2 Forecasts with Method 2

The forecasts of $\boldsymbol{H}_{t+k}$ is calculated as explained in Section 6.1. When forecasting the covariance matrix, $\boldsymbol{H}_{t+k}$, k step ahead, the forecasts of $\boldsymbol{D}_{t+k}$ and $\boldsymbol{R}_{t+k}$ may be done separately. We choose Method 2 when forecasting because Method 2 has shown to have better bias properties for almost all correlations as mentioned in Section 6.1.2.

A DCC-GARCH model with assumption of Gaussian, Student's t- and skew Student's t-distributed error, $\boldsymbol{z}_t$, is used to fit the whole data set, $t = 1, ..., 4062$.

Forecasts of $\boldsymbol{D}_{4062+k}$

Forecast of $\boldsymbol{D}_{4062+k}$, $k = 1, ..., 365$, with $\boldsymbol{D}_t, t = 1, ..., 4062$, fit to the data is shown in Figure 6. The black lines are $\boldsymbol{D}_t, t = 1, ..., 4062$, fit to the data and the red lines are the forecasts $\boldsymbol{D}_{4062+k}, k = 1, ..., 365$. The green, horizontal lines are the unconditional variance calculated from (6) with parameters given in 8.2.1. $\boldsymbol{D}_{t+k}$ is the same for all the three distributions since we assume Gaussian distributed errors.

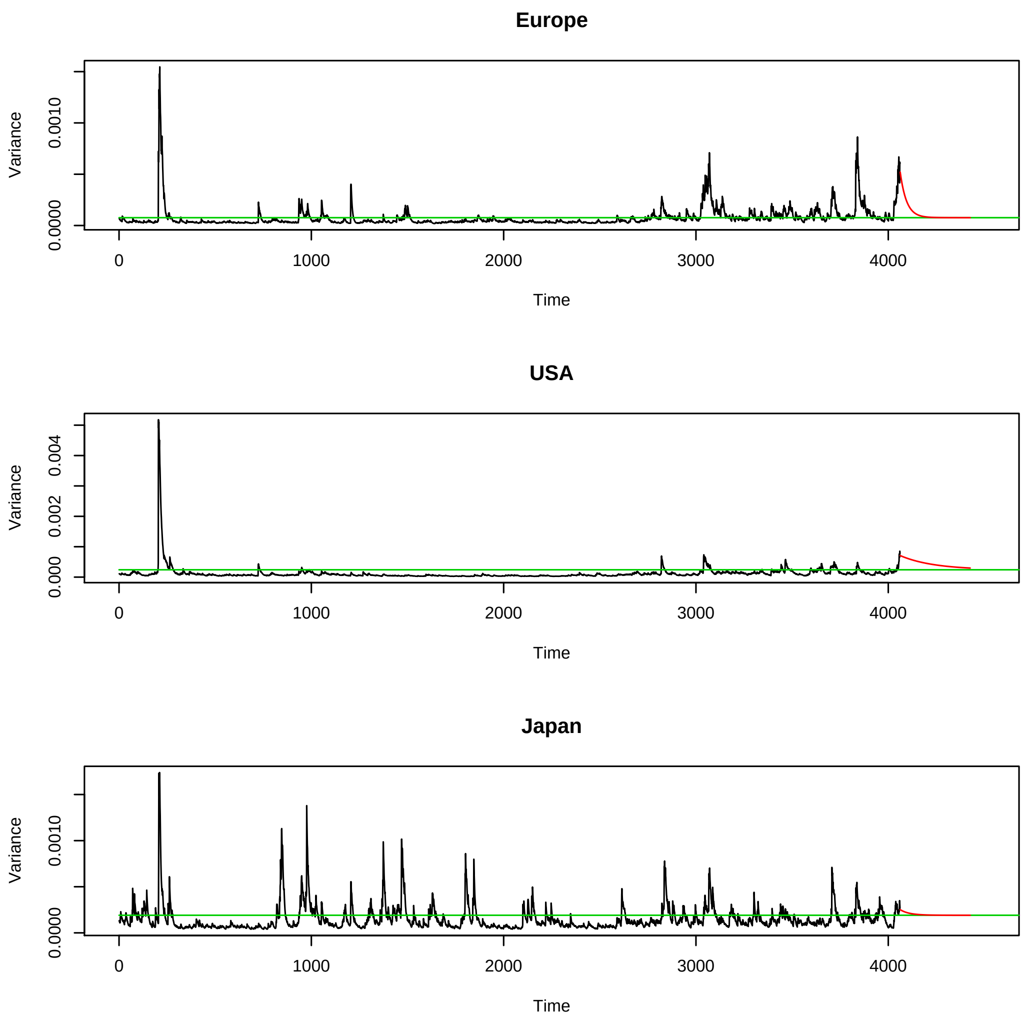

Figure 6: Plot of Dt from the data with 365 forecasts. Gaussian distributed error

The forecasts for USA use longer time to converge to the unconditional variance than for

Europe and Japan, since the estimated values of $\alpha_1 + \beta_1$ in Section 8.2.1 is larger for USA. As mentioned in Section 6.1, the memory of $\boldsymbol{D}_{t+k}$ decline with exponential rate $(\alpha_1 + \beta_1)$. Further, the estimated values of $\alpha_1 + \beta_1$ is larger for the Japanese data than for the European data, hence the European data converges fastest to the unconditional variance.

Forecasts of $\boldsymbol{R}_{4062+k}$

Forecasts of $\boldsymbol{R}_{4062+k}$, $k = 1, ..., 365$, with $\boldsymbol{R}_t, t = 1, ..., 4062$, fit to the data is shown in Figure 7. The diagonal of $\boldsymbol{R}_{4062+k}$, $k = 1, ..., 365$ is not shown, since it is 1 by definition. The black, red and light blue lines are $\boldsymbol{R}_t, t = 1, ..., 4062$ fit to the data with the Gaussian, Student's t- and skew Student's t-distributed errors, respectively. However, the difference between $\boldsymbol{R}_t, t = 1, ..., 4062$ fit to the data for the three distributions are are not easy to distinguish in this figure. The green, horizontal lines are the unconditional correlations. The grey, dark red and blue lines are the forecasts $\boldsymbol{R}_{4062+k}, k = 1, ..., 365$ assuming Gaussian, Student's t- and skew Student's t-distribution, respectively. It is also hard to distinguish the forecasts for the three distributions in this figure. But we see that all forecasts

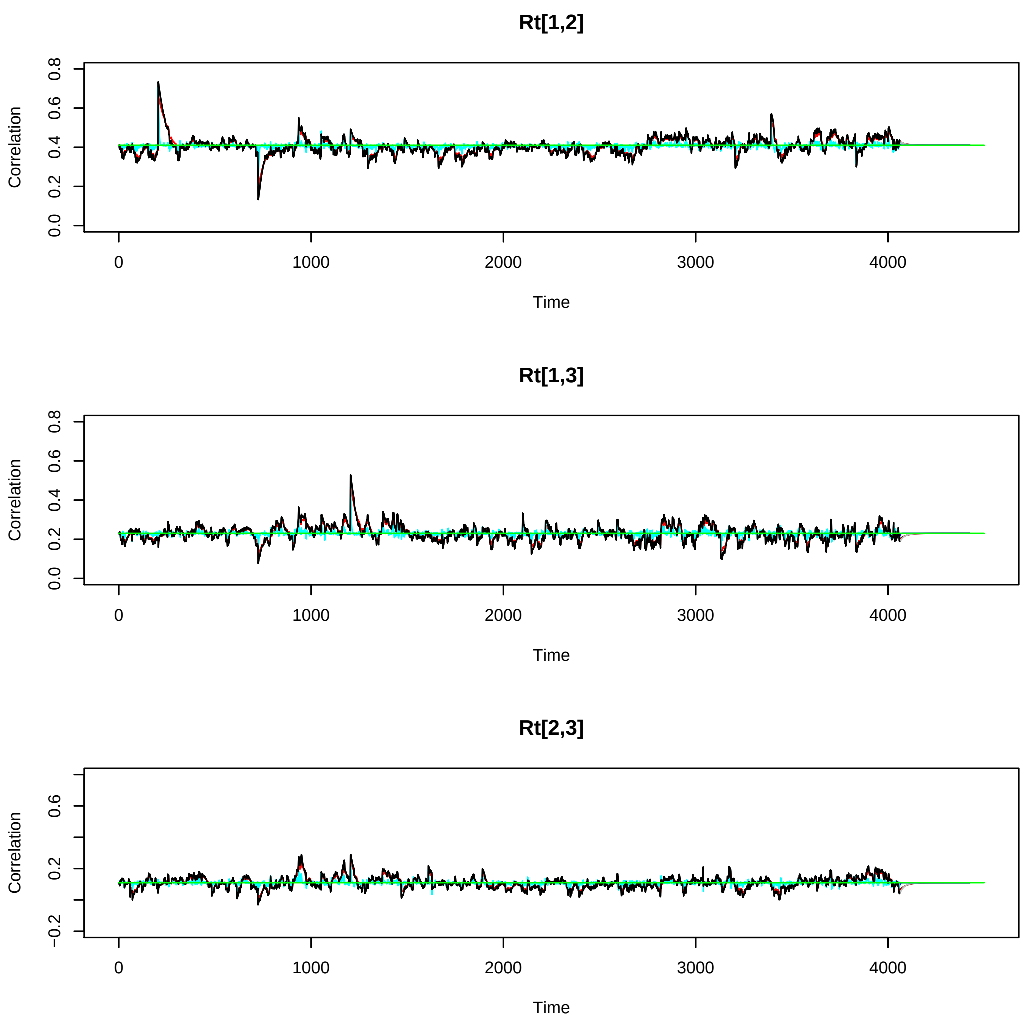

Figure 7: Plot of $\boldsymbol{R}_t$ from the data with 365 forecasts.

reach the unconditional correlations (green lines) when $k \to \infty$.

To more easily distinguish the forecasts, the forecasts $\boldsymbol{R}_{4062+k}$, $k = 1, ..., 100$, is shown in Figure 8 with the last 30 points of $\boldsymbol{R}_t$ fit to the data. In all the three plots the model with skew Student's t-distributed errors (blue lines) reach the unconditional correlations (green lines) faster than the model with Gaussian (grey lines) and Student's t-distributed errors (dark red lines), and the model with Student's t-distributed errors (dark red lines) reach the unconditional correlations (green lines) faster than the model with Gaussian distributed errors (grey lines). As mentioned earlier, this is caused by the difference in the estimated values of a and b.

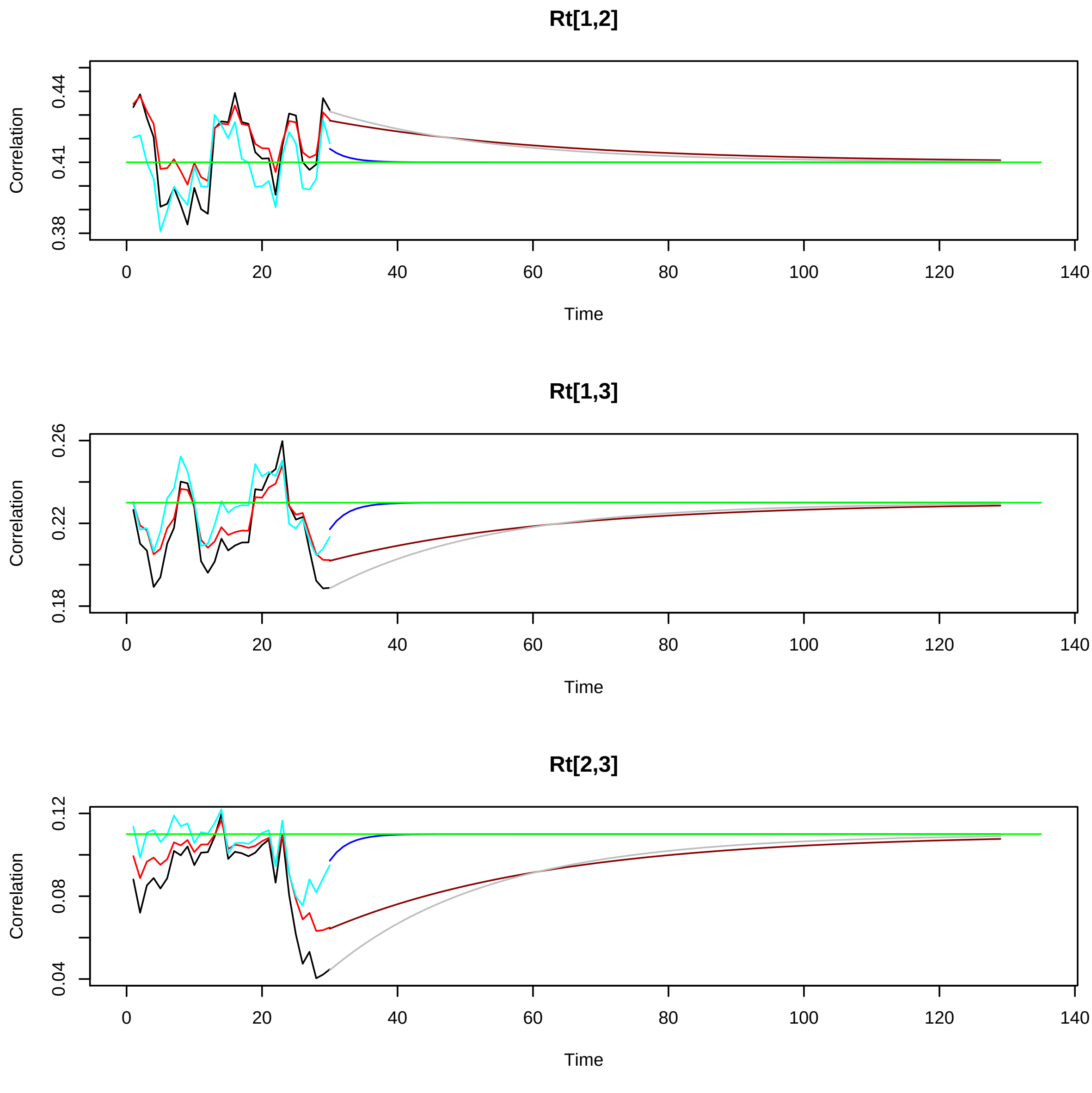

Figure 8: Plot of $\boldsymbol{R}_t$ from the last 30 datapoints with 100 forecasts.

Forecasts of $\boldsymbol{H}_{4062+k}$

The forecasts of the diagonal of $\boldsymbol{H}_{t+k}$ are the same for the model with Gaussian, Student's t- and skew Student's t-distributed errors.

Since $\boldsymbol{H}_{t+k} = \boldsymbol{D}_{t+k}\boldsymbol{R}_{t+k}\boldsymbol{D}_{t+k}$, the diagonal elements of $\boldsymbol{H}_{t+k}$ are:

$$\boldsymbol{H}_{t+k}[i,i] = \boldsymbol{D}_{t+k}[i,i]^2\boldsymbol{R}_{t+k}[i,i] = \boldsymbol{h}_{i,t+k}$$

Since $\boldsymbol{R}_{t+k}$ is 1 on the diagonal, the diagonal elements of $\boldsymbol{H}_{t+k}$ depends only of the elements of $\boldsymbol{D}_{t+k}$. Since the elements of $\boldsymbol{D}_{t+k}$ is the same for the model with Gaussian, Student's t and skew Student's t-distributed errors, the diagonal elements of $\boldsymbol{H}_{t+k}$ is the same as $\boldsymbol{D}_{t+k}$ and shown in Figure 6.

Forecasts of $\boldsymbol{H}_{4062+k}$, $k = 1, ..., 365$, with $\boldsymbol{H}_t, t = 1, ..., 4062$, fit to the data is shown in Figure 9. In Figure 9 the black, red and light blue lines are $\boldsymbol{H}_t, t = 1, ..., 4062$ fit to the data with Gaussian, Student's t- and skew Student's t-distributed errors, respectively. The grey, dark red and blue lines are the forecasts $\boldsymbol{H}_{4062+k}, k = 1, ..., 365$ assuming Gaussian, Student's t- and skew Student's t-distribution, respectively. The green, horizontal lines

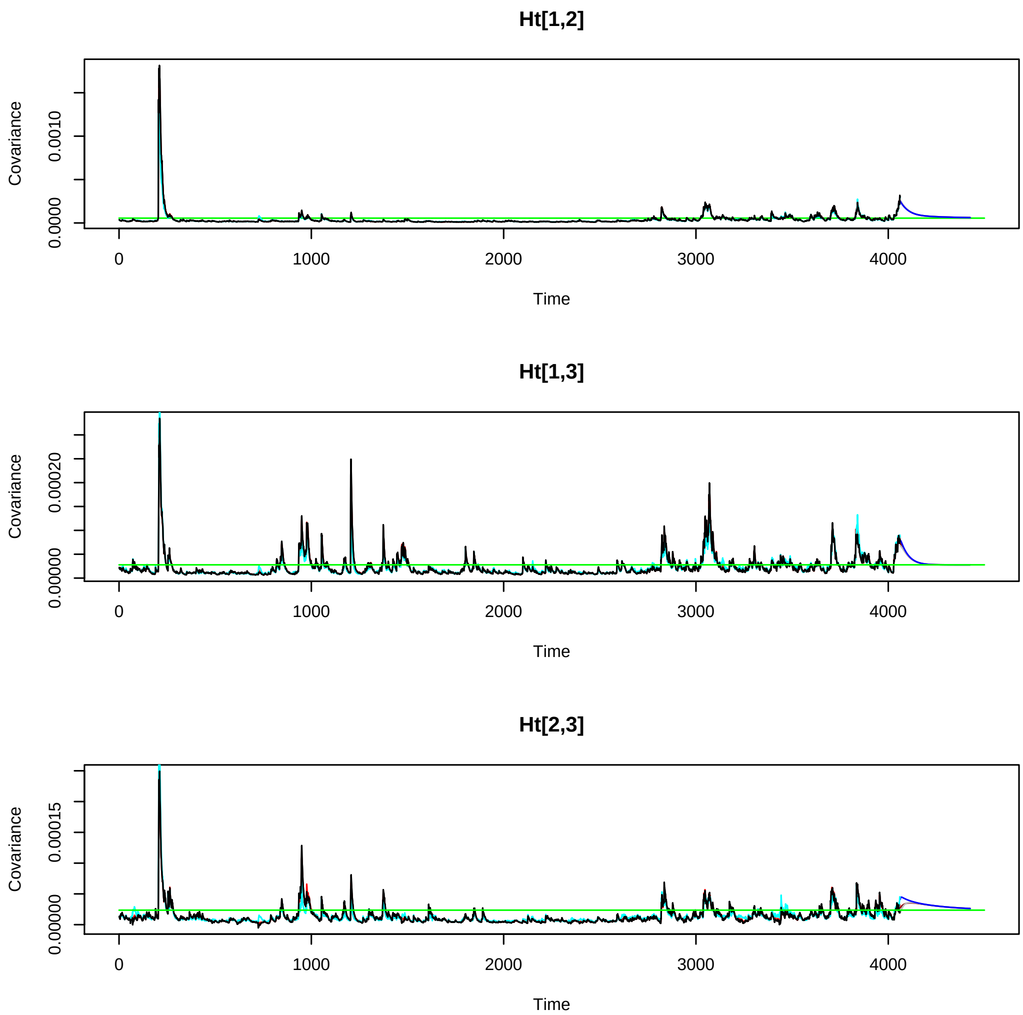

Figure 9: Plot of $\boldsymbol{H}_t$ from the data with 365 forecasts.

are the unconditional covariance, $\overline{\boldsymbol{H}} = \overline{\boldsymbol{D}}\ \overline{\boldsymbol{R}}\ \overline{\boldsymbol{D}}$, where $\overline{\boldsymbol{D}}$ is the unconditional standard deviation, and $\overline{\boldsymbol{R}}$ is described in Section 6.1.2. It is hard to distinguish the forecasts for the three distributions in this figure. However, we see that all forecasts reach the unconditional covariances (green lines) when $k \to \infty$.

To more easily distinguish the forecasts, $\boldsymbol{H}_{4062+k}$, $k = 1, ..., 100$, is shown in Figure 10 with the last 30 points of $\boldsymbol{H}_t$ fit to the data. The forecasts use many time points to reach the unconditional covariances (green lines), even 100 time points is not enough.

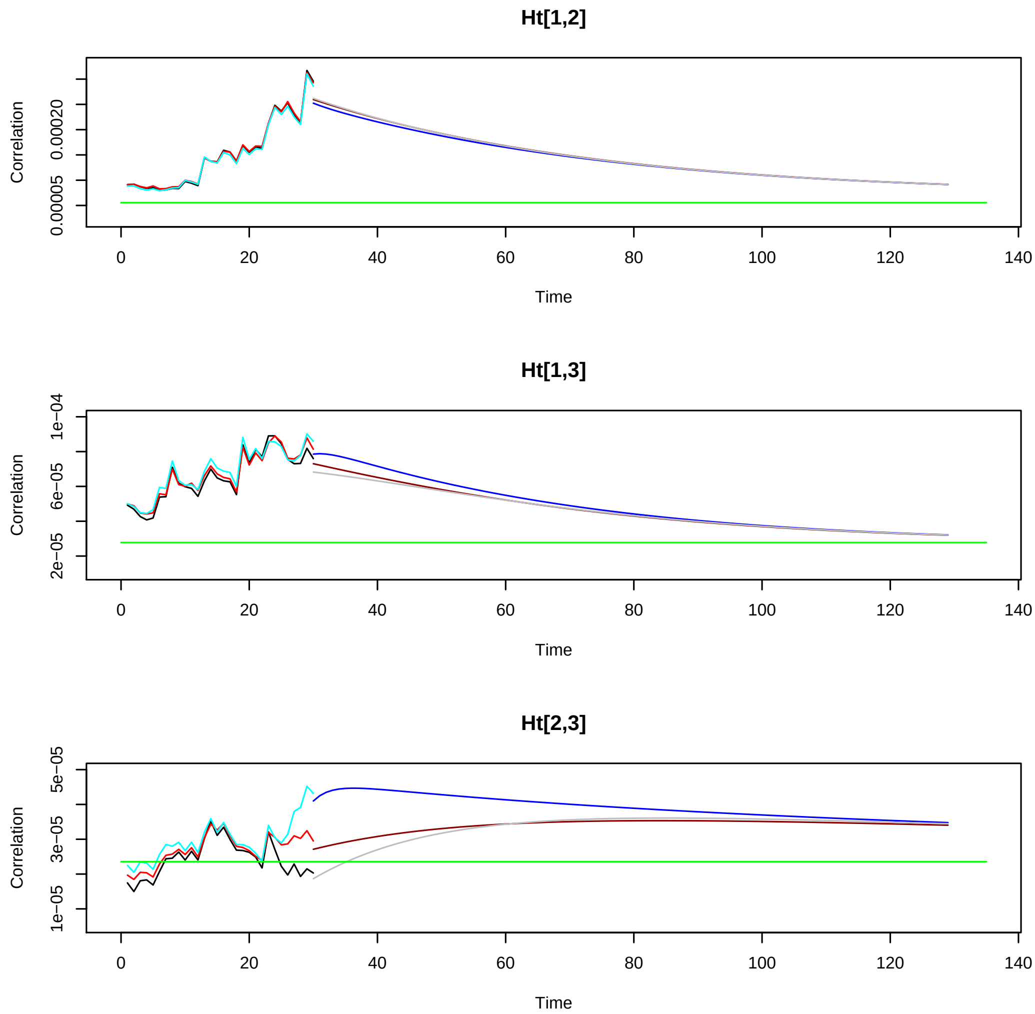

Figure 10: Plot of $\boldsymbol{H}_t$ from the last 30 datapoints with 100 forecasts.

8.4 Goodness of fit

We test the goodness of fit using the methods described in Chapter 7.

8.4.1 Goodness of marginal fits

In this section we will check whether the errors, z_t, for each of the three time series; Europe, USA and Japan is iid or not. The errors is jointly calculated from $\boldsymbol{z}_t = \boldsymbol{H}_t^{-1/2}\boldsymbol{a}_t$.

1. **Plot of the standardized errors, $\boldsymbol{z}_t$**

   The standardized errors, $\boldsymbol{z}_t$, are shown in Figure 11 for the three time series, Europe, USA and Japan and the three different distributions; Gaussian, Student's t and skew Student's t. There are no distinct difference between the errors of the three different error distributions. The plot of the European and American errors have some large

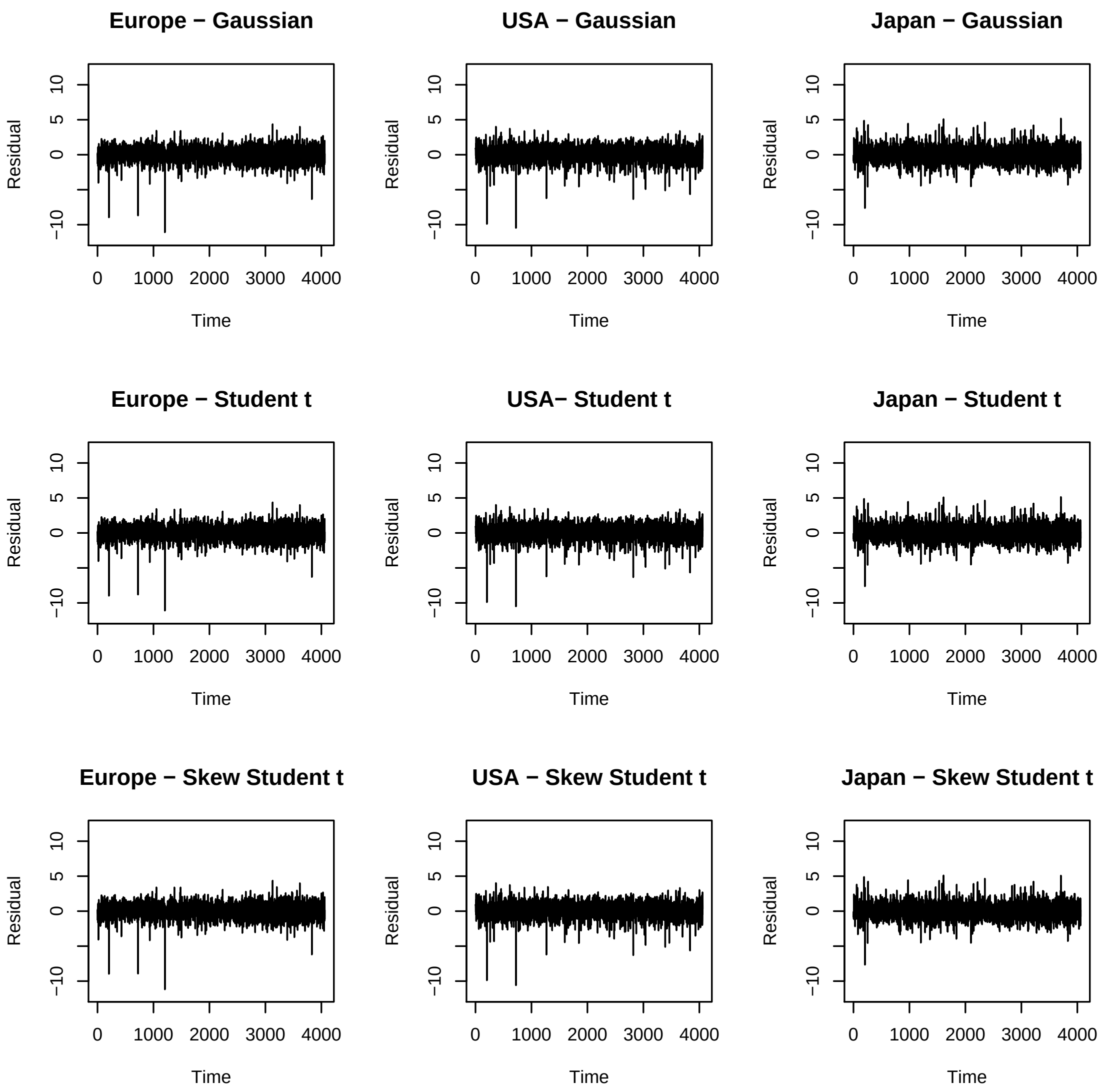

Figure 11: Plot of the errors

negative values, and do not quite look like white noise. It seems like the models have not explained all the variation in the data, especially in the negative direction for these two series. The Japanese errors on the other hand, look random and iid distributed.

2. **The sample autocorrelation function**

The autocorrelation function for the errors, z_t, of the Gaussian, Student's t- and skew Student's t-distribution is computed for 100 lags. It is expected that 5 acf-values should fall outside the 95% confidence-limits. The number of acf-values that falls outside is binomial distributed with number of trials equal to 100 and probability 5%.

Europe

The acf-plot for the European errors is shown in Figure 12. The acf-plot for the three different distributions; Gaussian, Student's t and skew Student's t looks similar, and there are only a small difference. For the Gaussian and the Student's t-distribution, 2 of the acf-values fall outside the limits, i.e. the blue dotted line. For the skew

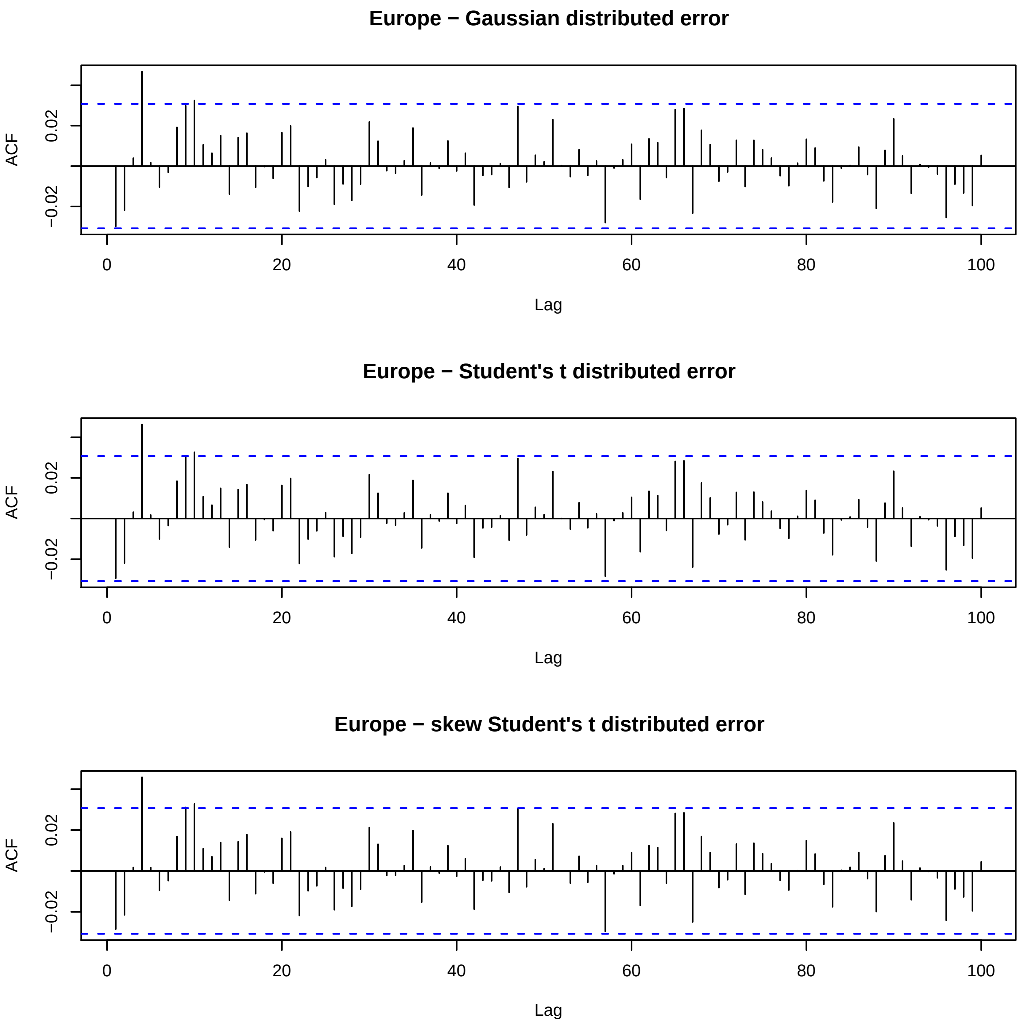

Figure 12: Acf of the European errors

Student's t-distibution 3 of the acf-values fall outside. The acf-values that fall outside the limits are the same for the different distributions. The acf-value that in addition falls outside for the skew Student's t-distribution is just barely outside, and barely inside for the two other distributions. The p-values is 0.25, 0.25, 0.49 for Gaussian, Student's t- and skew Student's t-distribution, respectively. Hence we can accept that the European errors are random from this test at level 5%.

USA

The acf-plot for the American errors, z_t, is shown in Figure 13. 9 of the acf-values fall outside the limits for all the three distributions. The acf-values that fall outside the limits are the same for the different distributions. The p-value of 9 acf-values to fall outside is 0.10. Hence we can accept that the American errors are random from this test at level 5%.

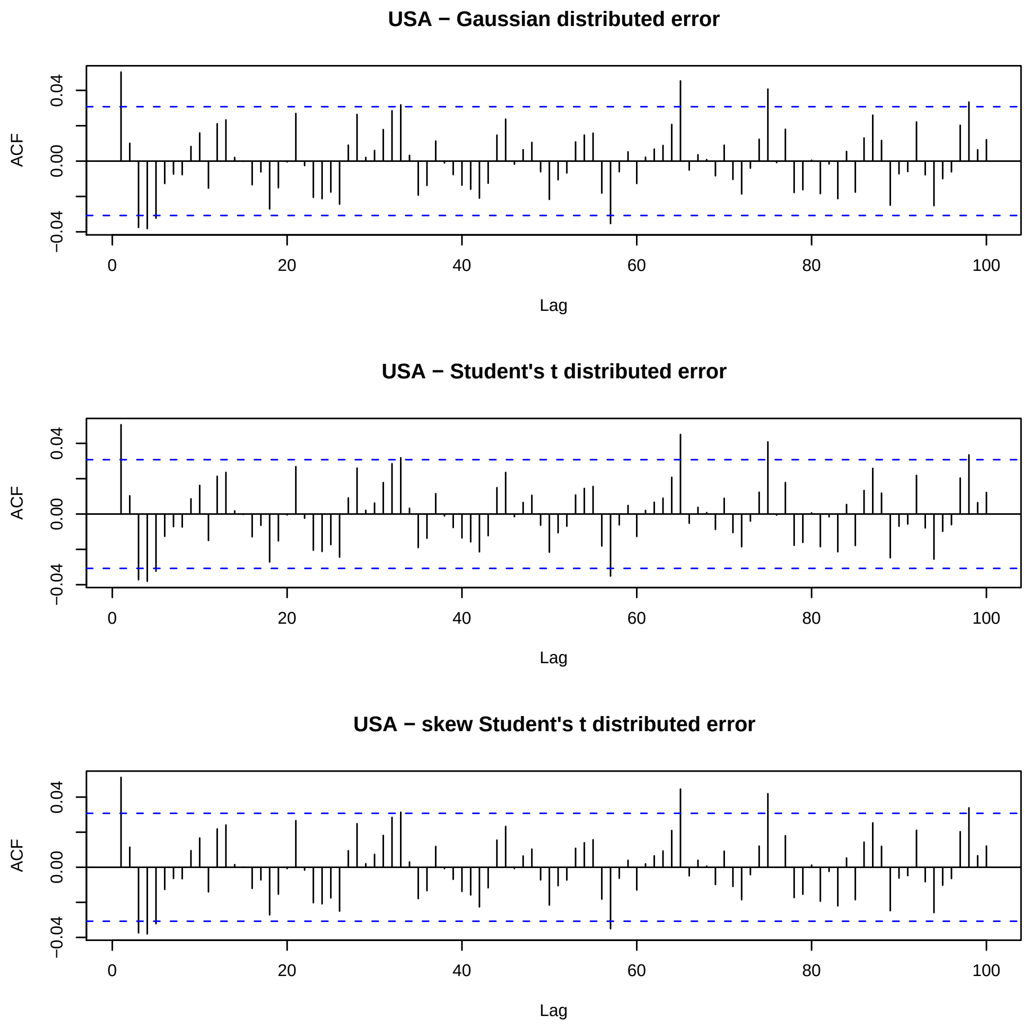

Figure 13: Acf of the American errors

Japan

The acf-plot for the Japanese errors, z_t, is shown in Figure 14. 3 of the acf-values fall outside the limits for the Gaussian and the Student's t-distribution. For the skew Student's t-distibution 4 of the acf-values fall outside. The acf-values that fall outside the limits are the same for the different distributions. The acf-value that in addition falls outside for the skew Student's t-distribution is just barely outside, and barely inside for the two other distributions. The p-values is 0.49, 0.49, 0.82 for Gaussian, Student's t- and skew Student's t-distribution, respectively. Hence we can accept that the Japanese errors are random from this test at level 5%.

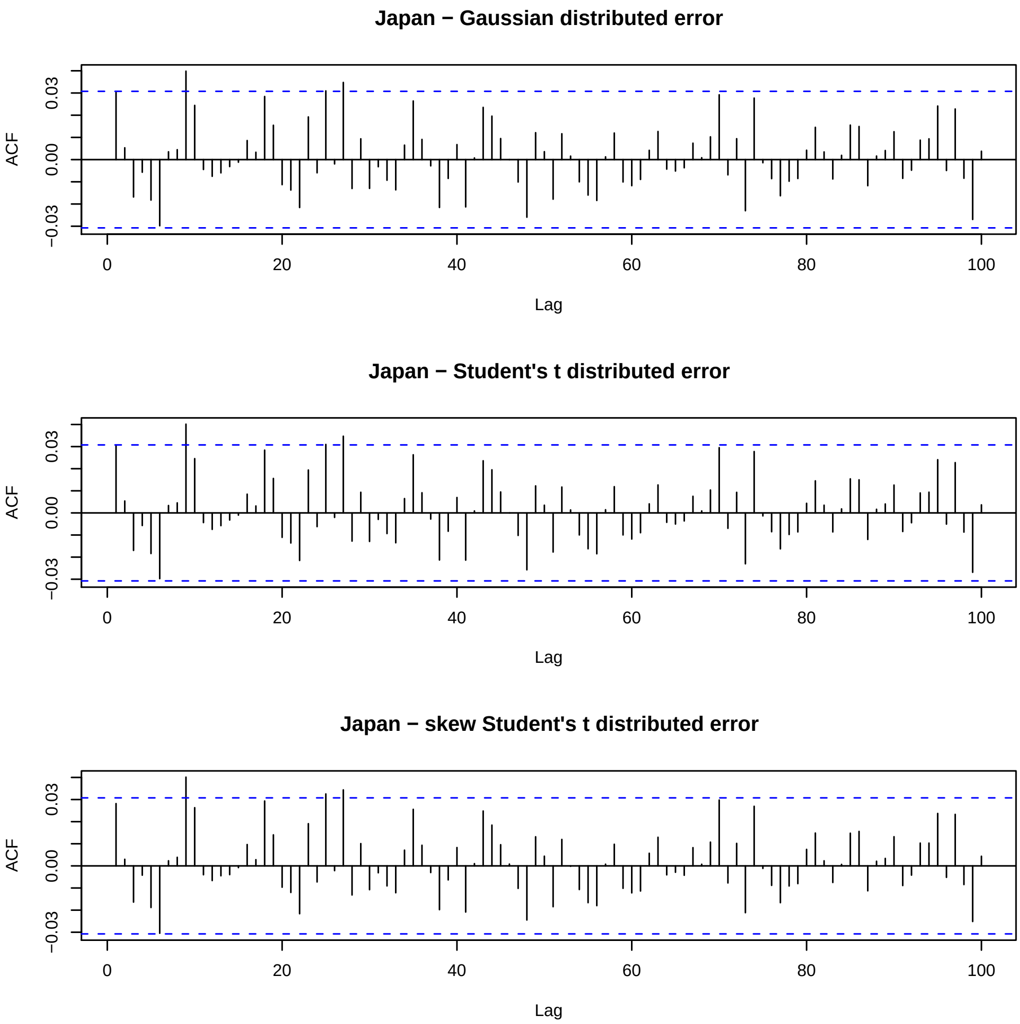

Figure 14: Acf of the Japanese errors

From this test, the skew Student's t-distributions performes a bit better than the Gaussian and Student's t-distribution. There was no different between the Gaussian and the Student's t-distribution. The p-values for the Japanese errors is highest, and indicates that the model is better fit for the Japanese data than the European and American data for all the three distributions. The results from the acf-plots indicates that the model is less well fit for the American data.

3. **Ljung-Box test**

The results of the Ljung-Box test is shown in Figure 15. The red line indicates the 5% level. There is no visible difference between the three distributions in this test.

This test accept the hypothesis that the errors, z_t, are uncorrelated for some of the lags for the European series. Hence there is no clear conclusion whether the European errors are uncorrelated or not from this test. For the American series on the other hand, this test concludes that the errors of the American series is not uncorrelated, since the hypothesis of uncorrelateness is rejected for all lags tested. The hypothesis of uncorrelateness is accepted for almost all the lags for the Japanese errors. Here the errors of the Japanese data seems to be uncorrelated.

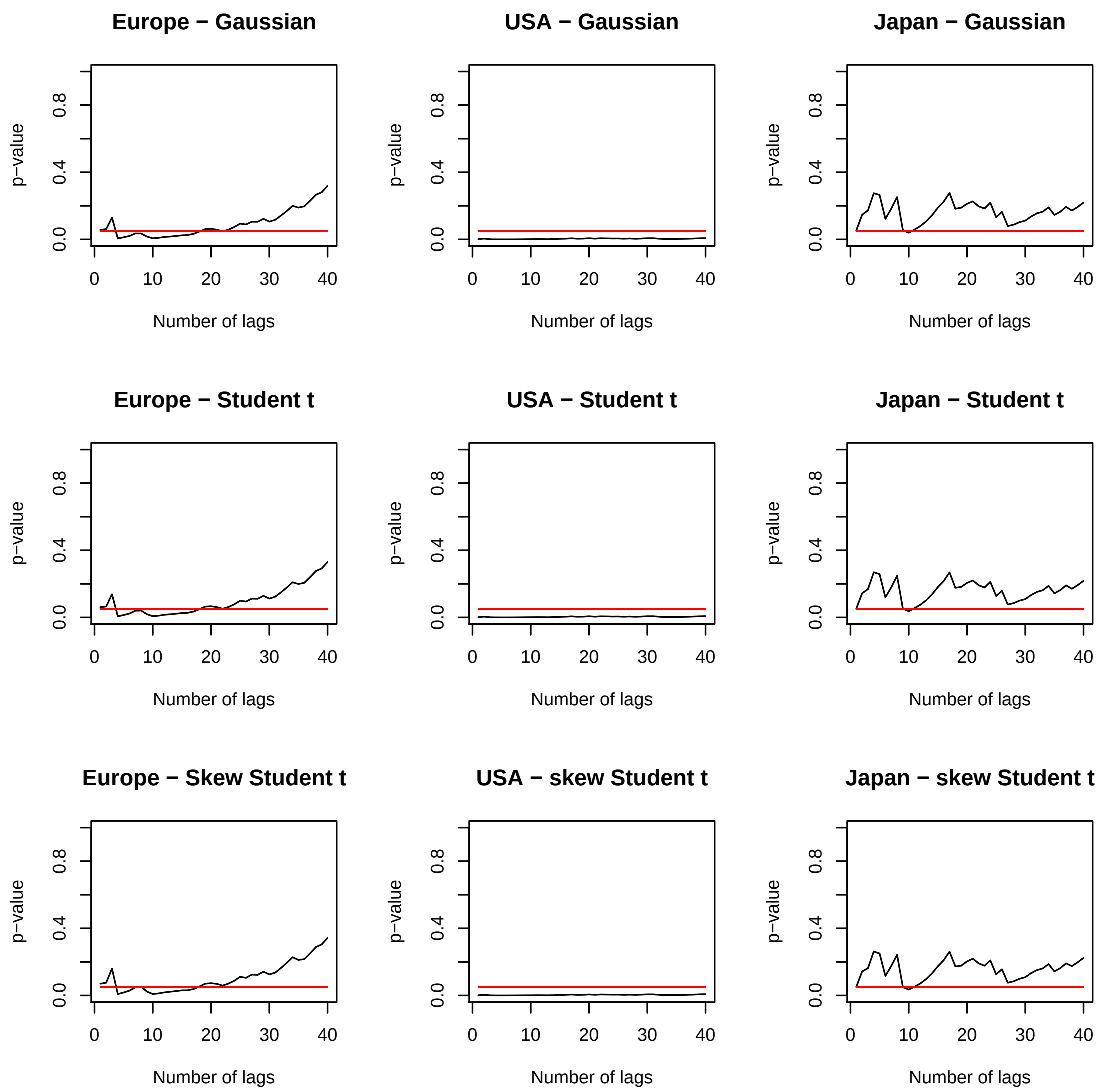

Figure 15: p-values from the Ljung-Box statistics

4. **Turning point test**

The expected number of turning points, T, for an iid sequence of size $n = 4061$ is:

$$\mathrm{E}[T] = 2(n-2)/3 = 2(4061-2) = 2706$$

and the variance is:

$$\mathrm{Var}[T] = (16n-29)/90 = (16 \cdot 4061 - 29)/90 = 721.6$$

Since n is large, T will be approximately N(E[T], Var[T]). A 95% confidence interval for T is [2653, 2759].

Europe

For the European errors, z_t, T_{Eur} is 2697, 2691 and 2701 for the Gaussian, Student's t- and skew Student's t-distribution respectively. T_{Eur} is inside both the 95% confidence interval for T for all the three distributions. Hence we accept the hypothesis that the European errors, z_t, is iid for all three distributions.

USA

For the American residulas, z_t, T_{USA} is 2689, 2687 and 2683 for the Gaussian, Student's t- and skew Student's t-distribution respectively. T_{USA} is inside the 95% confidence interval for T for all the three distributions. Hence we accept the hypothesis that the American errors, z_t, is iid for all three distributions.

Japan

For the Japanese errors, z_t, T_{Jap} is 2718 for all the three distributions. T_{Jap} is inside the 95% confidence interval for T. Hence we accept the hypothesis that the Japanese errors, z_t, is iid for all three distributions.

5. **Difference-sign test**

The expected number of times when the differenced series $z_i - z_{i-1} > 0$, S, of size $n = 4060$ is iid is:

$$\mathrm{E}[S] = (n-1)/2 = (4060-1)/2 = 2029.5$$

and the variance is:

$$\mathrm{Var}[S] = (n+1)/12 = (4060+1)/12 = 338.4$$

Since n is large, S will be approximately N(E[S], Var[S]). A 95% confidence interval for S is [1994, 2066].

Europe

For the European errors, z_t, S_{Eur} is 2018, 2018 and 2017 for the Gaussian, Student's t- and skew Student's t-distribution respectively. S_{Eur} is inside both the 95% confidence interval for S for all the three distributions. Hence we accept the hypothesis that the European errors, z_t, is iid for all three distributions.

USA

For the American errors, z_t, T_{USA} is 1974, 1974 and 1967 for the Gaussian, Student's t- and skew Student's t-distribution respectively. S_{USA} is not inside the 95% confidence interval for S for any of the three distributions. Hence this test concludes that the American errors, z_t, is not iid. This test indicates that our models have not explained all the dependence in the data.

Japan

For the Japanese errors, z_t, T_{Jap} is 2010, 2011 and 2010 for the Gaussian, Student's t- and skew Student's t-distribution respectively. S_{Jap} is inside the 95% confidence interval for S for all the three distributions. Hence we accept the hypothesis that the Japanese errors, z_t, is iid for all three distributions.

6. **Q-Q plot**

Europe

The upper panel of Figure 16 shows that the European errors, z_t, are more heavy-tailed than the Gaussian distribution in both tails. The left tail is more heavy than the right tail, so the European data seems to be skew.

The middle panel of Figure 16 shows that the Student's t-distribution gives a better fit for the European errors. However, the Student's t-distribution also seems to underestimate the left tail. The right tail is now slightly overestimated. Because the Student's t is symmetric it can not take into account different tails.

Finally, the lower panel in Figure 16 shows that the skew Student's t-distribution gives the best fit to the errors, although even this distribution underestimates the left tail. The right tail is here good explained.

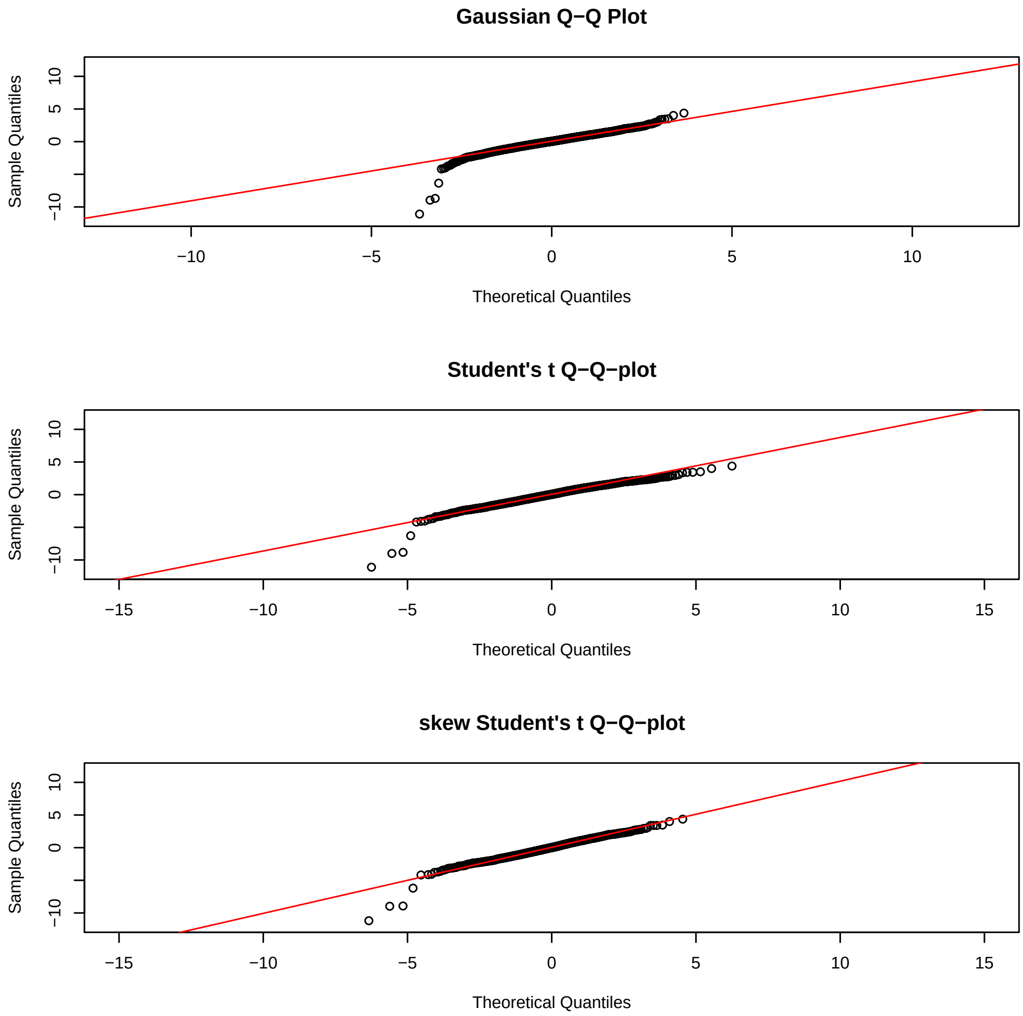

Figure 16: QQ-plot of the European errors, z_t.

USA

The upper panel of Figure 17 also shows like the upper panel of Figure 16 that the American errors, z_t, are more heavy-tailed than the Gaussian distribution in both tails. The left tail is more heavy than the right, which indicates that also the American data is skew.

The middle panel of Figure 17 shows that the Student's t-distribution gives a better fit for the American errors. However, the Student's t-distribution also seems to underestimate the left tail, and the right tail is slightly overestimated.

The lower panel of Figure 17 shows that the skew Student's t-distribution gives the best fit for the American data. The right tail is here good explained. However, the left tail is still underestimated.

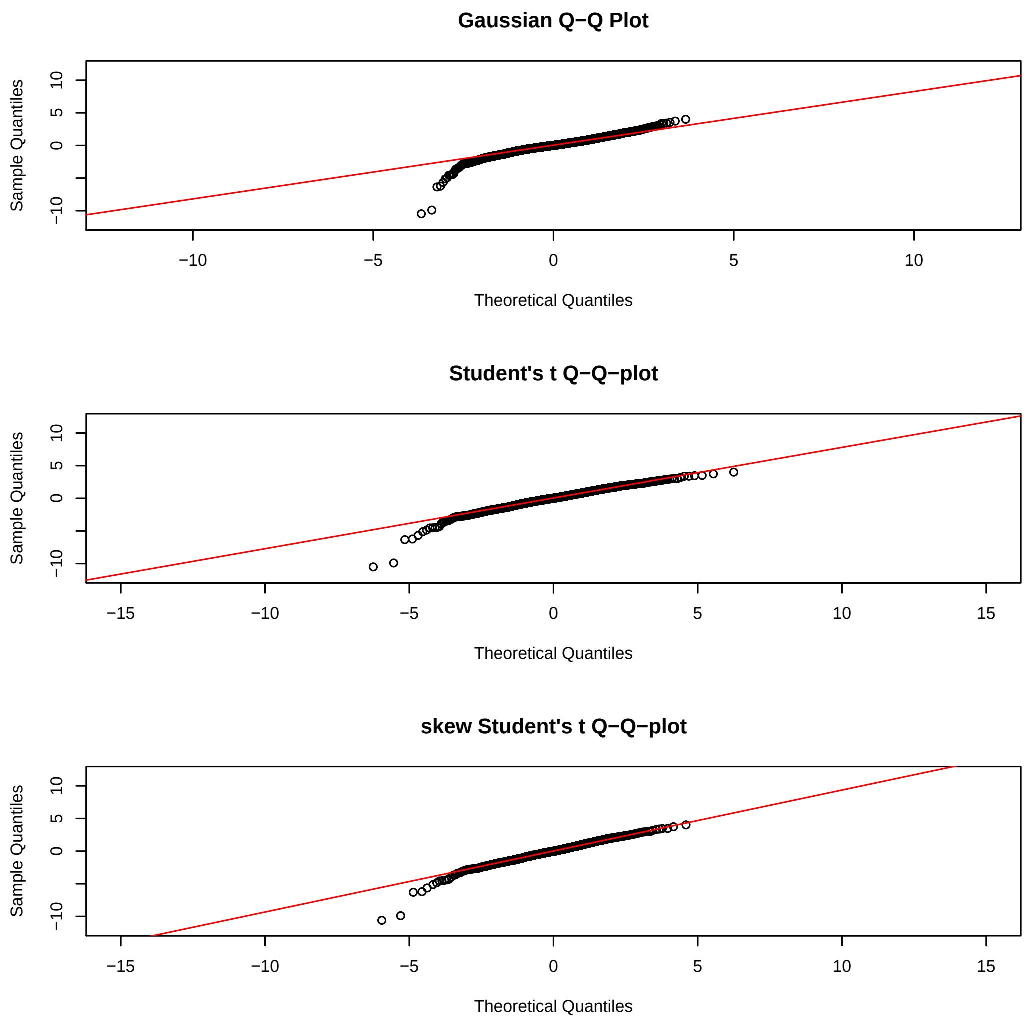

Figure 17: QQ-plot of the American errors, z_t.

Japan

The upper panel of Figure 18 also shows that for the Japanese errors, z_t, are more heavy-tailed than the Gaussian distribution in both tails. It is not clear which tail is the most heaviest from this plot, as seen for the upper panel in Figures 16 and 17.

The middle panel of Figure 18 is similar to the middle panels of Figures 16 and 17, it shows that the Student's t-distribution gives a better fit for the Japanese errors. However, the Student's t-distribution also seems to underestimate the left tail, and the right tail is slightly overestimated.

The lower panel of Figure 17 shows that the skew Student's t-distribution gives the best fit for the Japanese data in both tails. Both tails seems to be a bit underestimated, however, the points do not deviate much from the red line. The DCC-GARCH model with skew Student's t-distributed errors seems to better fit the Japanese data, than the European and American data.

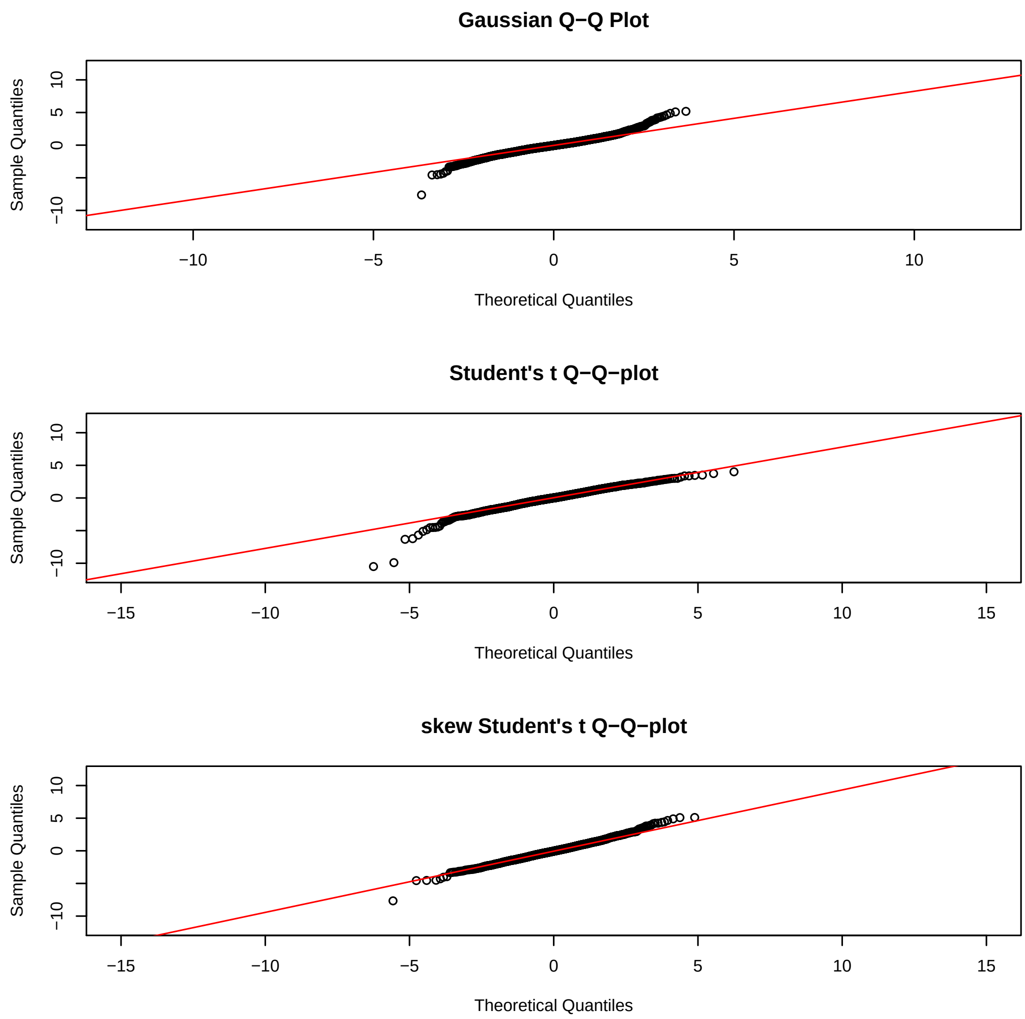

Figure 18: QQ-plot of the Japanese errors, z_t.

8.4.2 Goodness of multivariate fit

The tests in Section 7.2 will be considered. In this section we will check the errors, $\boldsymbol{z}_t$ jointly.

1. **Baringhaus-Franz multivariate test**

   The test statistic is used to test if the errors, $\boldsymbol{z}_t$, are Gaussian, Student's t- or skew Student's t-distributed with the estimation parameters found in Section 8.2. The dimension of $\boldsymbol{z}_t$ is 4061×3.

   Gaussian distributed errors

   First we will test whether $\boldsymbol{z}_t$ is from the multivariate Gaussian distribution. We test $\boldsymbol{z}_t$ against a data set of simulated multivariate standardized Gaussian variates of the same dimension as $\boldsymbol{z}_t$.

   With 1000 bootstrap replicates, the critical point is estimated to be 1.98. The observed statistic is $T = 9.15$, which gives a p-value of 0.00. Hence the hypothesis that $\boldsymbol{z}_t$ is distributed as multivariate standardized Gaussian is rejected.

   Student's t-distributed errors

   We then test whether $\boldsymbol{z}_t$ is from the multivariate Student's t-distribution. We test $\boldsymbol{z}_t$ against a multivariate data set of simulated standardized Student's t-variates of the same dimension as $\boldsymbol{z}_t$ with the ν-parameter estmated in Section 8.2.2.

   With 1000 bootstrap replicates, the critical point is estimated to be 2.00. The observed statistic is $T = 7.3$, which gives a p-value of 0.00. Hence the hypothesis that $\boldsymbol{z}_t$ is distributed as multivariate standardized Student's t is rejected.

   Skew Student's t-distributed errors

   Finally, we test whether $\boldsymbol{z}_t$ is from the multivariate skew Student's t-distribution. We test $\boldsymbol{z}_t$ against a data set of simulated multivariate standardized skew Student's t-variates of the same dimension as $\boldsymbol{z}_t$ with the parameters estimeated in Section 8.2.3. How to simulate from a multivariate skew Student's t-distribution is described in Azzalini [4] on page 15.

   With 1000 bootstrap replicates the critical point is estimated to be 1.98. The observed statistic is $T = 2.30$, which gives a p-value of 0.012. At the 5% level the hypothesis that $\boldsymbol{z}_t$ is distributed as multivariate standardized Student's t is rejected, but at the 1% level the hypothesis is accepted. A great difference between real data and simulated data is that with real data it is often hard to obtain a p-value larger than 1%. Obtaining a p-value larger than 1% is actually not that bad.

2. **Backtesting of VaR; Cupiec test and Christoffersen's Markov test**

   This will be discussed in Section 8.5.

8.5 1-day ahead Value-at-Risk

In this section VaR is computed at different confidence levels as described in Section 6.2. The confidence levels used are q ∈ {0.005,0.01,0.05,0.95,0.99,0.995}.

If the observed value of p_t fall outside the confidence interval a violation is said to occur, i.e. if VaR_t^q for q ∈ {0.005,0.01,0.05} is greater than the observed value of p_t this day, or VaR_t^q for q ∈ {0.95,0.99,0.995} is less than the observed value of p_t this day, we have a violation.

To measure how well the VaR-forecasts are, we extract six training and test sets for six corresponding experiments, as shown in Table 2. Based on each training set, a model is estimated. Then this model is used to forecast 1-day ahead VaR at different confidence levels for each day in the test set.

Table 2: Training and test sets

Experiment	Training set		Test set	
no	First day	Last day	First day	Last day
1	1	1000	1001	1500
2	500	1500	1501	365
3	1000	365	2001	3650
4	1500	3650	3651	3000
5	365	3000	3001	3500
6	3650	3500	3501	4000

The test procedure is as follows:

- Estimate the parameters of the DCC-GARCH model based on the *training set*.
- For each day t in the *test set*:
  1. Compute the 1-day ahead forecast, $\boldsymbol{H}_{t+1}$, given information up to time t.
  2. Compute the observed portfolio value of $p_t = \boldsymbol{w}^T \boldsymbol{\mu}_t$. We follow Aas et al. [15] and choose $\boldsymbol{\mu}_t = (\boldsymbol{c} + \boldsymbol{K}\boldsymbol{r}_{t-1})$ where $\boldsymbol{r}_{t-1}$ is the observed return at time point $t-1$.
  3. Calculate the expected value of p_t; $\mathrm{E}[p_t]$, and the variance of p_t; $\mathrm{Var}[p_t]$. Then compute the VAR_t^q for the portfolio return p_t at confidence level q at time t.
  4. Calculate the number of violations at each quantile level.

All six experiments were run using the different error distributions; multivariate Gaussian, Student's t and skew Student's t.

For each training set the $\boldsymbol{K}$-matrix and $\boldsymbol{c}$ is given in Table 3.

Table 3: $\boldsymbol{K}$-matrix and $\boldsymbol{c}$ for the training sets.

Experiment	c_1	c_2	c_3	d_{eur}	d_{jpy}	e_{jpy}
1	-0.000041	0.000221	-0.000036	0.390	0.388	0.162
2	-0.000030	0.000442	-0.000705	0.395	0.253	0.109
3	0.000275	0.000452	-0.000295	0.350	0.325	0.180
4	0.000309	0.000447	0.000003	0.368	0.228	0.205
5	0.000442	0.000909	-0.000524	0.394	0.298	0.127
6	0.000577	0.000824	-0.000403	0.379	0.310	0.145

Gaussian distributed errors

The estimated parameters when assuming a multivariate Gaussian error distribution for step two of the estimation procedure are given in Table 4, for the six different experiments. The estimated values of b differ a lot for the different experiments. The value of a does not differ that much for the different experiments.

Table 4: Estimated parameters for the DCC-GARCH when the errors are multivariate standardized Gaussian distributed.

Experiment	a	b
1	0.0224	0.923
2	0.0407	0.482
3	0.0333	0.229
4	0.0155	2.88e-06
5	0.0037	0.987
6	0.0046	0.954

If we consider experiment 4, the parameter b is estimated to be approximately 0. The parameter a is also close to 0. Consider the case when both $a \to 0$ and $b \to 0$. Then we get the following expression for $\boldsymbol{Q}_t$:

$$\boldsymbol{Q}_t = (1 - a - b)\overline{\boldsymbol{Q}} + a\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^T + b\boldsymbol{Q}_{t-1} \approx \overline{\boldsymbol{Q}} \tag{46}$$

We see that in this case, $\boldsymbol{Q}_t$ will be approximately constant, and equal to $\overline{\boldsymbol{Q}}$. In experiment 4 we nearly have this case. Figure 19 shows $\boldsymbol{Q}_t$ for the whole dataset. $\boldsymbol{Q}_t$ for experiment 4 is between the blue lines, $t = 1500, ..., 3650$. We see that in this period $\boldsymbol{Q}_t$ is close to being constant, there are no big peaks. Hence it seems logical for b to be estimated as approximately 0 for this period.

Student's t-distributed errors

The estimated parameters when assuming a multivariate Student's t error distribution for step two of the estimation procedure are given in Table 5, for the six different experiments. The estimated values of a and b are close to the ones for the Gaussian distribution in Table 4. There are however some differences in the values of a in experiments 1 and 4. We

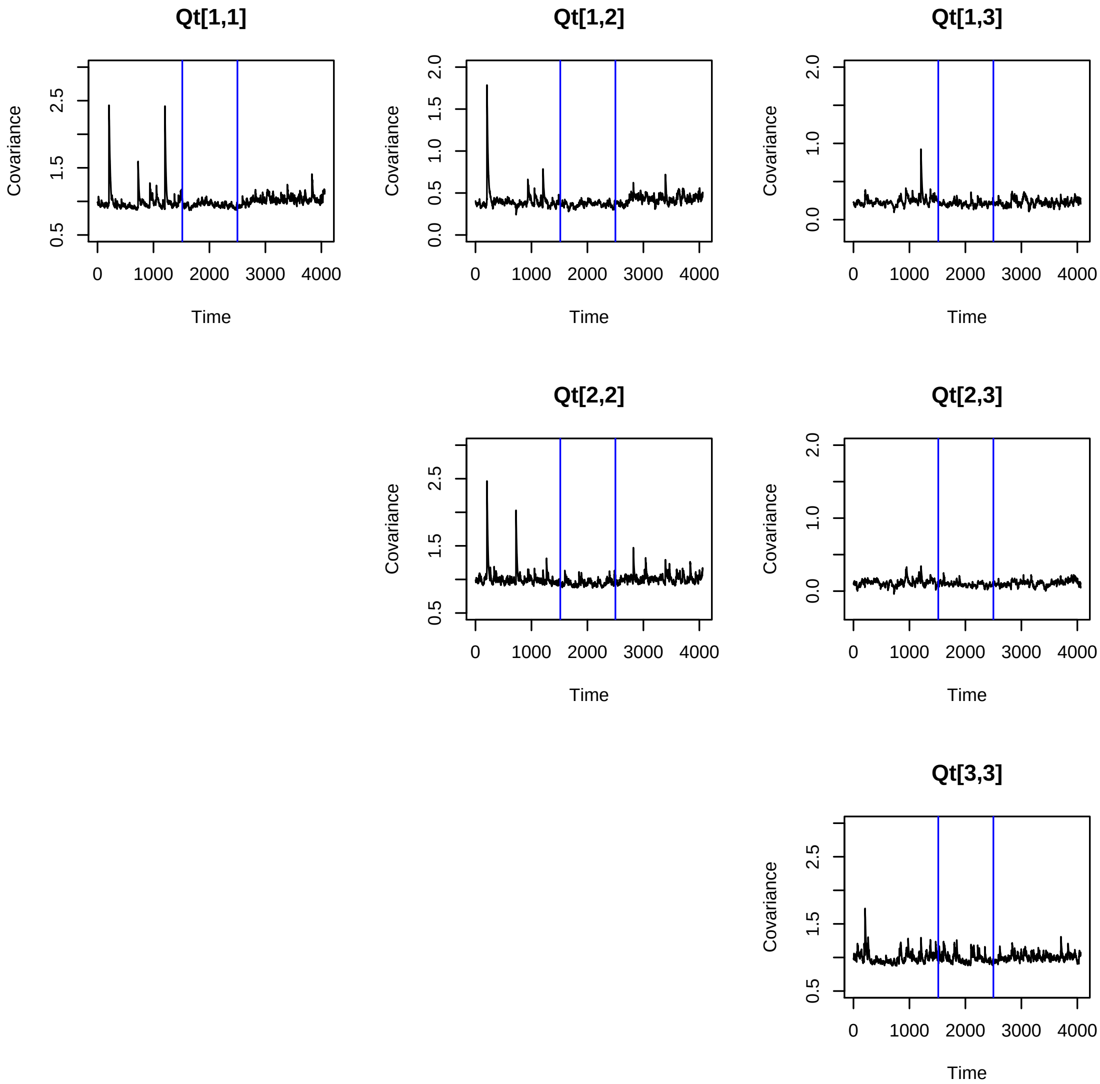

Figure 19: Plot of $\boldsymbol{Q}_t$ for the whole dataset when $\boldsymbol{z}_t$ is Gaussian distributed. $\boldsymbol{Q}_t$ for experiment 4 is between the blue lines.

Table 5: Estimated parameters for the DCC-GARCH when the errors are multivariate standardized Student's t-distributed.

Experiment	a	b	ν
1	0.0061	0.951	5.53
2	0.0478	0.407	5.50
3	0.0374	0.486	6.24
4	0.0033	1.19e-04	7.71
5	0.0034	0.989	8.37
6	0.0066	0.947	8.87

see that a is estimated even smaller for the Student's t-distribution than for the Gaussian distribution. Moreover, there is also a difference in b in experiment 3.

Skew Student's t-distributed errors

The estimated parameters when assuming a multivariate skew Student's t error distribution for step two of the estimation procedure are given in Table 6, for the six different experiments. The estimated values of ν is close to those for the Student's t-distribution in Table 5. There are two big differences between Tables 5 and 6. In experiment 4, b is close to 0 for the Student's t-distribution, but close to 1 for the skew Student's t-distribution. If we look at the expression of $\boldsymbol{Q}_t$, we get with a approximately equal to 0:

$$\begin{aligned}
\boldsymbol{Q}_t &= (1-a-b)\overline{\boldsymbol{Q}} + a\epsilon_{t-1}\epsilon_{t-1}^T + b\boldsymbol{Q}_{t-1} \\
&\approx (1-b)\overline{\boldsymbol{Q}} + b\boldsymbol{Q}_{t-1} \\
&= (1-b)\overline{\boldsymbol{Q}} + b(1-b)\overline{\boldsymbol{Q}} + b^2\boldsymbol{Q}_{t-2} \\
&= \dots \\
&= \sum_{i=0}^{t-1} b^i(1-b)\overline{\boldsymbol{Q}} + b^t\boldsymbol{Q}_0
\end{aligned} \tag{47}$$

When b is close to 1, $\boldsymbol{Q}_t$ will be approximately equal to $\boldsymbol{Q}_0 = \overline{\boldsymbol{Q}}$, because the first term, $\sum_{i=0}^{t-1} b^i(1-b)\overline{\boldsymbol{Q}}$, will be approximately 0.

Hence we see that even in this case, $\boldsymbol{Q}_t$ will be approximately equal to $\overline{\boldsymbol{Q}}$, as for the case when both $a \to 0$ and $b \to 0$ in (46). Hence when $a \to 0$, there is no big difference between b close to 1 or 0.

In experiment 1 there is also a big difference in b. $\boldsymbol{Q}_t$ corresponding to the Student's t-distribution and the skew Student's t in experiment 1 is shown in Figures 20 and 21, respectively. $\boldsymbol{Q}_t$ of the skew Student's t-distribution seems to be closest to the corresponding period at time 1 to 1000 in figure 19 of $\boldsymbol{Q}_t$ of the whole dataset. The peaks in Figure 21 seems to be closer to the peaks in Figure 19. E.g. in both cases; Student's t and skew Student's t, there is a peak around time 200 in four of the plots. The peak is larger for the skew Student's t-distribution. In figure 19 we see that there is a big peak for $\boldsymbol{Q}_t$ also around time 200, at around same values as for $\boldsymbol{Q}_t$ for the skew Student's t-distribution.

Table 6: Estimated parameters for the DCC-GARCH when the errors are skew multivariate standardized Student's t-distributed. The $\boldsymbol{\Omega}$-matrix can be computed from the other parameters using (38).

Experiment	a	b	ν	δ_1	δ_2	δ_3	ξ_1	ξ_2	ξ_3
1	0.0197	1.45e-03	5.64	-0.794	-0.770	-0.129	0.501	0.488	0.0910
2	0.0140	0.427	5.60	-0.415	-0.283	-0.110	0.298	0.207	0.0825
3	0.0121	0.558	6.28	-0.266	0.130	0.152	0.198	-0.098	-0.114
4	0.000533	0.9995	7.77	-0.129	-0.456	0.301	0.0995	0.335	-0.228
5	0.00273	0.997	8.28	0.0836	-0.649	0.0273	-0.0650	0.460	-0.0213
6	0.00501	0.957	8.82	-0.353	-0.600	-0.0448	0.267	0.431	0.0350

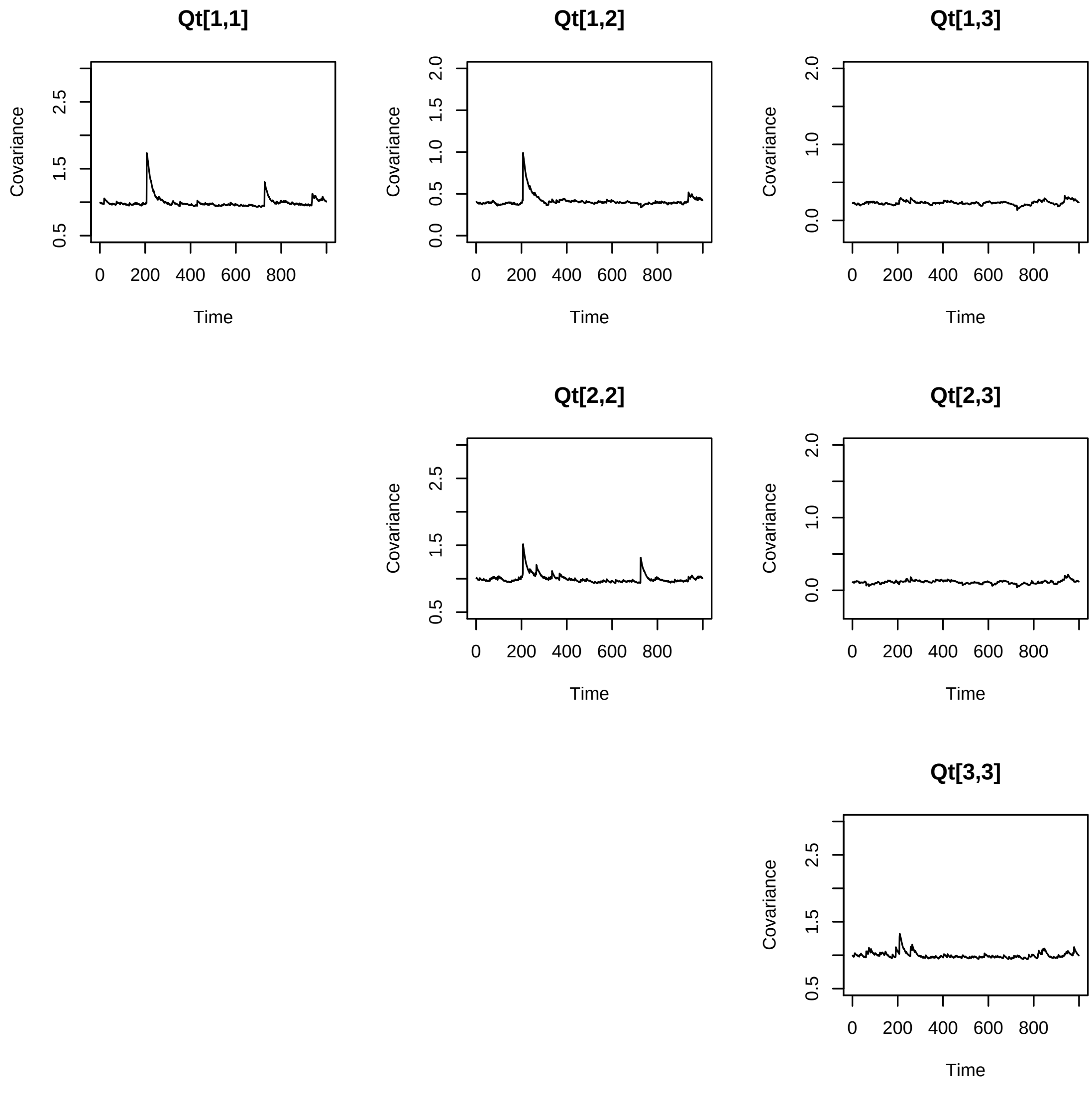

Figure 20: Plot of $\boldsymbol{Q}_t$ for experiment 1 when $\boldsymbol{z}_t$ is Student's t-distributed.

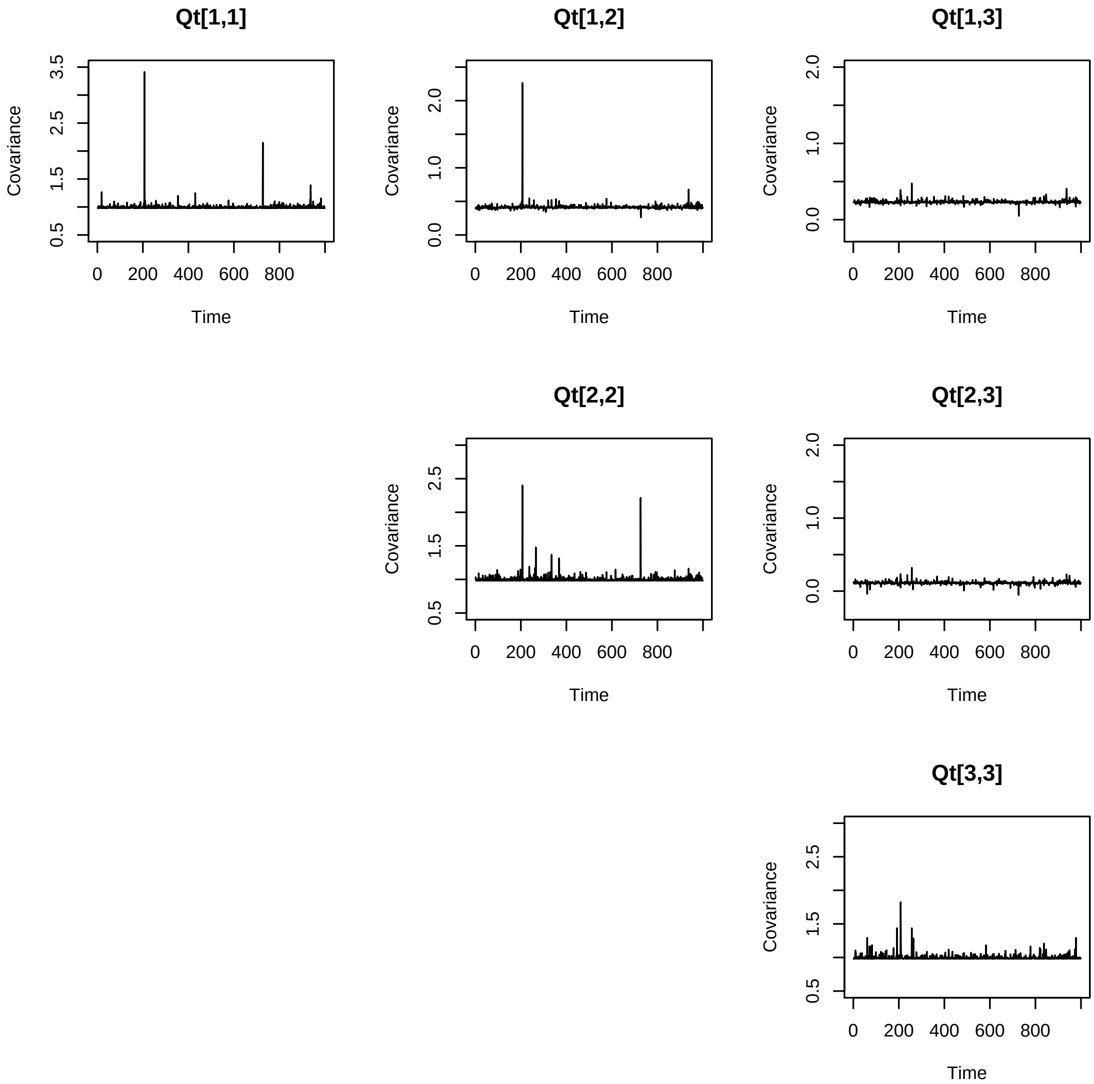

Figure 21: Plot of $\boldsymbol{Q}_t$ for experiment 1 when $\boldsymbol{z}_t$ is skew Student's t-distributed.

8.5.1 Number of violations

The number of violations is counted for each of the error distributions; Gaussian, Student's t and skew Student's t. Both Method 1 and Method 2 of forecasting provide the same result.

In addition to the number of violations for the DCC-GARCH model, the number of violations for the CCC-GARCH model is also given here.

Gaussian distribution

The number of violations of VaR of DCC-GARCH and CCC-GARCH for multivariate Gaussian distributed errors are given in Table 7.

Table 7: Number of violations of VaR of DCC-GARCH and CCC-GARCH for multivariate Gaussian distributed errors.

Multivariate Gaussian

			CCC			
Experiment no	0.5%	1%	5%	99.5%	99%	95%
1	7	11	25	4	5	21
2	2	5	8	4	7	15
3	2	2	16	1	2	13
4	10	12	28	9	12	37
5	6	7	27	1	2	24
6	6	14	34	3	6	21
Expected number	2.5	5	25	2.5	5	25
Full period	33	51	138	22	34	131
Expected number	15	30	150	15	30	150

			DCC			
Experiment no	0.5%	1%	5%	99.5%	99%	95%
1	4	8	20	4	4	16
2	4	7	10	6	8	18
3	2	8	19	2	4	17
4	9	12	26	7	11	33
5	6	9	33	2	3	24
6	6	10	31	1	6	17
Expected number	2.5	5	25	2.5	5	25
Full period	31	54	139	22	36	125
Expected number	15	30	150	15	30	150

Student's t-distribution

The number of violations of VaR of DCC-GARCH and CCC-GARCH for multivariate Student's t-distributed errors are given in Table 8.

Table 8: Number of violations of VaR of DCC-GARCH and CCC-GARCH for multivariate Student's t-distributed errors.

Multivariate Student's t

			CCC			
Experiment no	0.5%	1%	5%	99.5%	99%	95%
1	5	9	27	3	4	21
2	0	2	8	4	4	16
3	2	2	17	1	1	13
4	6	10	30	4	10	38
5	4	6	29	1	1	26
6	3	11	36	0	4	22
Expected number	2.5	5	25	2.5	5	25
Full period	20	40	147	13	24	136
Expected number	15	30	150	15	30	150
			DCC			
Experiment no	0.5%	1%	5%	99.5%	99%	95%
1	1	4	22	1	4	20
2	0	4	11	4	5	22
3	2	2	23	1	2	20
4	4	10	28	5	7	34
5	4	7	34	1	3	26
6	2	7	32	0	1	18
Expected number	2.5	5	25	2.5	5	25
Full period	13	34	150	12	22	140
Expected number	15	30	150	15	30	150

Skew Student's t-distribution

The number of violations of VaR of DCC-GARCH and CCC-GARCH for multivariate skew Student's t-distributed errors are given in Table 9.

Table 9: Number of violations of VaR of DCC-GARCH and CCC-GARCH for multivariate skew Student's t-distributed errors.

Multivariate skew Student's t

Experiment no	CCC 0.5%	1%	5%	99.5%	99%	95%
1	5	10	27	3	4	21
2	0	2	8	4	4	16
3	2	2	17	1	1	13
4	6	10	30	4	10	38
5	4	6	27	1	1	26
6	3	9	35	0	4	22
Expected number	2.5	5	25	2.5	5	25
Full period	20	39	144	13	24	136
Expected number	15	30	150	15	30	150

Experiment no	DCC 0.5%	1%	5%	99.5%	99%	95%
1	1	3	20	3	4	21
2	0	3	10	5	8	25
3	2	2	19	1	4	22
4	4	7	24	5	10	39
5	2	4	32	1	3	32
6	1	2	31	1	6	24
Expected number	2.5	5	25	2.5	5	25
Full period	10	21	136	16	35	163
Expected number	15	30	150	15	30	150

8.5.2 Kupiec test

The p-values for the Kupiec test described in Section 7.2 is calculated, for each experiment and quantile for each of the error distributions; multivariate Gaussian, Student's t and skew Student's t.

Gaussian distribution

The results for the multivariate Gaussian distributed errors are given in Table 10 for both CCC-GARCH and DCC-GARCH.

Table 10: p-values for Kupiec test for DCC-GARCH and CCC-GARCH with multivariate Gaussian distributed errors.

Multivariate Gaussian

			CCC			
Experiment no	0.5%	1%	5%	99.5%	99%	95%
1	0.02	0.02	1.00	0.38	1.00	0.40
2	0.74	1.00	0.00	0.38	0.40	0.03
3	0.74	0.13	0.05	0.28	0.13	0.01
4	0.00	0.01	0.55	0.00	0.01	0.02
5	0.06	0.40	0.69	0.28	0.13	0.84
6	0.06	0.00	0.08	0.76	0.66	0.40
Full period	0.00	0.00	0.31	0.09	0.47	0.10

			DCC			
Experiment no	0.5%	1%	5%	99.5%	99%	95%
1	0.38	0.21	0.29	0.38	0.64	0.05
2	0.38	0.40	0.00	0.06	0.21	0.13
3	0.74	0.21	0.20	0.74	0.64	0.08
4	0.00	0.01	0.84	0.02	0.02	0.12
5	0.06	0.11	0.12	0.74	0.33	0.84
6	0.06	0.05	0.23	0.28	0.66	0.08
Full period	0.00	0.00	0.35	0.09	0.29	0.03

If we use a 5% level for Kupiec test, the null hypothesis is rejected 12 times for the CCC-GARCH model and 7 times for the DCC-GARCH model as shown in Table 10.

Further, the null hypothesis is rejected 2 times for the full period for the CCC-GARCH model and 3 times for the DCC-GARCH model.

Student's t-distribution

The p-values for the Kupiec test for the Student's t-distribution for both CCC-GARCH and DCC-GARCH are given in Table 11.

Table 11: p-values for Kupiec test for DCC-GARCH and CCC-GARCH with multivariate Student's t-distributed errors.

Multivariate Student's t						
			CCC			
Experiment no	0.5%	1%	5%	99.5%	99%	95%
1	0.16	0.11	0.69	0.76	0.64	0.40
2	0.03	0.13	0.00	0.38	0.64	0.05
3	0.74	0.13	0.08	0.28	0.03	0.01
4	0.06	0.05	0.32	0.38	0.05	0.01
5	0.38	0.66	0.42	0.28	0.03	0.84
6	0.76	0.02	0.03	0.03	0.64	0.53
Full period	0.22	0.08	0.80	0.60	0.25	0.23
			DCC			
Experiment no	0.5%	1%	5%	99.5%	99%	95%
1	0.28	0.64	0.53	0.28	0.64	0.29
2	0.03	0.33	0.00	0.38	1.00	0.53
3	0.74	0.13	0.68	0.28	0.13	0.29
4	0.38	0.05	0.55	0.16	0.40	0.08
5	0.38	0.40	0.08	0.28	0.33	0.84
6	0.74	0.40	0.17	0.03	0.03	0.13
Full period	0.60	0.47	1.00	0.42	0.12	0.40

If we use a 5% level for Kupiec test, the null hypothesis is rejected 12 times for the CCC-GARCH model and 5 times for the DCC-GARCH model as shown in Table 11.

For the full period the null hypothesis is never rejected, either for the CCC-GARCH model or the DCC-GARCH model.

Skew Student's t-distribution

The p-values for the Kupiec test for the skew Student's t-distribution for both CCC-GARCH and DCC-GARCH are given in Table 12.

Table 12: p-values for Kupiec test for DCC-GARCH and CCC-GARCH with multivariate skew Student's t-distributed errors.

Multivariate skew Student's t

Experiment no	0.5%	1%	5%	99.5%	99%	95%
CCC						
1	0.16	0.05	0.69	0.76	0.64	0.40
2	0.03	0.13	0.00	0.38	0.64	0.05
3	0.74	0.13	0.08	0.28	0.03	0.01
4	0.06	0.05	0.32	0.38	0.05	0.01
5	0.38	0.66	0.69	0.28	0.03	0.84
6	0.76	0.11	0.05	0.03	0.64	0.53
Full period	0.22	0.11	0.61	0.60	0.25	0.23
DCC						
1	0.28	0.33	0.29	0.76	0.64	0.40
2	0.03	0.33	0.00	0.16	0.21	1.00
3	0.74	0.13	0.20	0.28	0.64	0.53
4	0.38	0.40	0.84	0.16	0.05	0.01
5	0.74	0.64	0.17	0.28	0.33	0.17
6	0.28	0.13	0.23	0.28	0.66	0.84
Full period	0.17	0.08	0.23	0.80	0.37	0.28

If we use a 5% level for Kupiec test, the null hypothesis is rejected 11 times for the CCC-GARCH model and 4 times for the DCC-GARCH model as shown in Table 12.

Further, the null hypothesis for the full period is never rejected, either for the CCC-GARCH model or the DCC-GARCH model.

For the CCC-GARCH model, the number of failures are 12, 12 and 11 for the Gaussian, Student's t- and skew Student's t-distribution respectively. For the DCC-GARCH model the number of failures are 7, 5 and 4 for the Gaussian, Student's t- and skew Student's t-distribution, respectively. From this result the DCC-GARCH model performs better than the CCC-GARCH model for all the distributions. Morever, the test might give an indication that the DCC-GARCH model with skew Student's t-distributed errors is best fit to the data.

8.5.3 Christoffersen's Markov test

For Christoffersen's Markov test only the p-values for the full period is computed for simplicity. These are shown in Table 13 for all the distributions. At 5% significance level, the model with Gaussian distributed errors is rejected 3 times, and none for the Student's t- and skew Student's t-distributed errors. Hence Christoffersen's Markov test indicates that Student's t and skew Student's t fit the data better than the Gaussian distribution. The DCC-GARCH with skew Student's t-distributed errors has greater p-values than the model with Student's t-distributed errors 2 out of 6 times, and Student's t has greater p-values than the DCC-GARCH with skew Student's t-distributed errors 4 out of 6 times. From this test the model with Student's t-distributed errors seems to perform best, but there is no clear indication.

Table 13: p-values for Christoffersen's Markov test of DCC-GARCH.

	DCC					
Distribution	0.5%	1%	5%	99.5%	99%	95%
Multivariate Gaussian	0.00	0.00	0.25	0.07	0.15	0.03
Multivariate Student's t	0.53	0.25	0.71	0.39	0.10	0.35
Multivariate skew Student's t	0.16	0.07	0.15	0.62	0.20	0.18

8.6 Conclusion

In all tests for marginal goodness of fit the DCC-GARCH with skew Student's t-distributed errors outperformed the DCC-GARCH with Gaussian and Student's t-distributed errors. For the Ljung-Box test, Turning point test and the Difference-sign test, there was no significant difference between the models with Gaussian, Student's t- and the skew Student's t-distributed errors. The acf-plot and Q-Q-plots, however indicate that the model with skew Student's t-distributed errors give a better fit to the data than the model with Gaussian and Student's t-distributed errors.

The marginal distribution of the European, American and Japanese series seemed to be a bit different. For the Japanese data all the three models gave an appropriate fit, but the model with skew Student's t-distributed errors gave best fit. All tests accepted the fitted models for the Japanese data. For the European data the test acf-plot, the Turning point test and the Difference sign test indicated appropriate fit for all models, while the Q-Q-plot and Ljung-Box test did not give clear indications. For the American data none of the models gave a very good fit.

Concerning multivariate fit, both the Baringhaus-Franz multivariate test and the Kupiec test favourized the model with skew Student's t-distributed errors. This model was the only one that got a p-value different from 0 for the Baringhaus-Franz multivariate test. The model with skew Student's t-distributed errors was accepted with a 1% significance level. The results of the Christoffersen's test did not give a clear indication what model best fitted the data. However, the models with Student's t- and skew Student's t- distributed errors performed much better than the model with a Gaussian error distribution.

Comparing the DCC-GARCH model with the CCC-GARCH model using the Kupiec test showed that the DCC-GARCH model gave a better fit to the data.

9 Concluding Remarks and Future Work

In this thesis we have studied the DCC-GARCH model with Gaussian, Student's t and skew Student's t-distributed errors. For a basic understanding of the GARCH model, the univariate GARCH and multivariate GARCH models in general were discussed before the DCC-GARCH model was considered.

The Maximum likelihood method is used to estimate the parameters. The estimation of the correctly specified likelihood is difficult, and hence the DCC-model was designed to allow for two stage estimation. Usually Gaussian distributed errors are assumed in the first stage independent of the choice of the error distribution in the correctly specified likelihood. In the second stage, the parameters a and b of the dynamic correlation matrix, and the parameters of the error distribution, are estimated using the correctly specified likelihood.

After the parameters of the DCC-model have been estimated, the forecast of the conditional covariance matrix is obtained by forecasting the conditional variances and the conditional correlation matrix separately. The forecasts of the conditional variances is done by assuming Gaussian distributed errors. The forecast of the conditional correlation matrix can not be directly calculated. In this thesis, two different methods of approximating this matrix have been discussed.

An important issue is how to evaluate goodness of fit for the DCC-GARCH model. This might be done by checking both the marginal and multivariate goodness of fit. One specific approach considered is the backtesting of Value-at-Risk, this is used to measure risk of loss of a portfolio of financial asset series.

After precenting the theory, DCC-GARCH models were fit to a portfolio consisting of European, American and Japanese stocks assuming three different error distributions; multivariate Gaussian, Student's t and skew Student's t. The European, American and Japanese series seemed to have a bit different marginal distributions. The DCC-GARCH model with skew Student's t-distributed errors performed best. But even the DCC-GARCH with skew Student's t-distributed errors did explain all of the asymmetry in the asset series. Hence even better models may be considered. Comparing the DCC-GARCH model with the CCC-GARCH model using the Kupiec test showed that the first model gave a better fit to the data.

There are several possible directions for future work. It might be better to use other marginal models such as the EGARCH, QGARCH and GJR GARCH, that capture the asymmetry in the conditional variances. If the univariate GARCH models are more correct, the DCC-GARCH model might yield better results. Other error distributions, such as a Normal Inverse Gaussian (NIG) might also give a better fit. When we fitted the Gaussian, Student's t- and skew Student's t-distibutions to the data, we assumed all the distributions to be the same for the three series. This might be a too restrictive criteria. A model where the marginal distributions is allowed to be different for each of the asset series might give a better fit. One then might use a Copula to link the marginals together.

References

[1] C. ALEXANDER (2001). *Market Models*. John Wiley Sons Ltd, 1 edition.

[2] C. ALMEIDA and J. VICENTE (2009). Are interest rate options important for the assessment of interest rate risk? *No 179, Working Papers Series from Central Bank of Brazil, Research Department*.

[3] T.W. ANDERSON (1984). *An introduction to multivariate statistical analysis*. Wiley series, 2 edition.

[4] A. AZZALINI and A. CAPITANIO (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *J. Roy. Statist. Soc., B 65*, pages 367–389.

[5] L. BARINGHAUS and C. FRANZ (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis, Volume 88 , Issue 1 (January 2004)*, pages 190 – 206.

[6] M. BERG JENSEN and A. LUNDE (2001). The nig-s and arch model: A fat-tailed stochastic, and autoregressive conditional heteroscedastic volatility model. *Working paper series No. 83, University of Aarhus*.

[7] C. BLANCO and M. OKS (2004). Backtesting var models: Quantitative and qualitative tests. *The Risk Desk. Vol. IV, No. 1*.

[8] T. BOLLERSLEV (1986). Generalized autoregressiv conditional heteroskedasticity. *Journal of econometrics, Vol. 31*, pages 307–327.

[9] T. BOLLERSLEV (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The Review of Economics and Statistics, Vol. 72*, pages 498–505.

[10] JONATHAN CRYER and KUNG-SIK CHAN (2008). *Time Series Analysis*. Springer, 2 edition.

[11] R.F. ENGLE and K. SHEPPARD (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. *NBER Working Papers, No. 8554*.

[12] R.F ENGLE (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of uk inflation. *Econometrica 50*.

[13] R.F. ENGLE (2000). Dynamical conditional correlation - a simple class of multivariate garch models. *Working Paper Series with number 2000-09*.

[14] E. JONDEAU (2005). Conditional asset allocation under non-normality: how costly is the mean-variance criterion? *FAME Research Paper Series with number rp132*.

[15] K. AAS, X.K. DIMAKOS and I. HOBÆK HAFF (2005). Risk estimation using the multivariate normal inverse gaussian distribution. *Journal of Risk, Vol. 8, No. 2*, pages 39–60.

[16] T. PETERS (2004). Forecasting the covariance matrix with the dcc garch model. *Stockholm University, Examensarbete 2008:4*.

[17] A. SILVENNOINEN and T. TERÄSVIRTA (2007). Multivariate garch models. *Working Paper Series in Economics and Finance, No. 669.*

[18] RUEY S. TSAY (2002). *Analysis of Financial Time Series.* Wiley-Interscience, 1 edition.

[19] J.H. VENTER and P.J.. JONGH (2002). Risk estimation using the normal inverse gaussian distribution. *Journal of Risk, Vol. 4*, pages 1–24.