Stockholms universitet

Backtesting and Forecasting the DCC-GARCH model

Tony Fayez

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2025:25
Matematisk statistik
September 2025

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2025:25**
http://www.math.su.se

Backtesting and Forecasting the DCC-GARCH model

Tony Fayez*

September 2025

Abstract

Modelling multiple assets requires multivariate approaches to capture co-movements in returns. While GARCH models are well established for volatility forecasting, their extension to multivariate settings allows analysis of portfolio risk. This thesis evaluates the DCC-GARCH(1,1) model, applied to daily stock returns from 2014-2024. Two specifications are considered, with Gaussian and multivariate Student's- t innovations. Their performance is assessed through Value-at-Risk backtesting and statistical tests. Results show that both models capture volatility dynamics well, with only minor differences between the Gaussian and Student's-t distributions.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: tofa0849@student.su.se. Supervisor: Kristoffer Lindensjö, Jan-Olov Persson.

Acknowledgements

I would like to extend my sincere gratitude to my supervisors, Kristofer Lindensjo and Jan Olov Persson, for their invaluable guidance and support throughout the course of this thesis.

I also wish to acknowledge the use of ChatGPT as a complementary tool, which provided assistance in R programming, optimizing computationally intensive scripts, and supporting the structuring of text as well as the formatting of tables and figures.

Contents

1 Introduction

Modeling and forecasting volatility is a central theme in modern financial econometrics. Asset returns often display volatility clustering, heavy tails and time-varying correlations features that cannot be captured by classical models assuming constant variance.

The introduction of the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle in 1982 (1) and its extension to the Generalized ARCH (GARCH) model by Bollerslev in 1986 (3), provided a framework for modeling conditional variances in financial time series. These models have since become a cornerstone in risk management portfolio allocation and derivative pricing.

While univariate GARCH models are useful for analyzing the volatility of a single asset, many practical applications require understanding the joint dynamics of multiple assets. For example, the construction of an optimal portfolio or the evaluation of systemic risk necessitates the modeling of time-varying covariances and correlations. This has motivated the development of multivariate GARCH (MGARCH) models, which extend the univariate framework to higher dimensions. Among these, the Dynamic Conditional Correlation (DCC-GARCH) model proposed by Engle and Sheppard (13) has gained attention due to its flexibility compared to earlier specifications.

This thesis evaluates a DCC GARCH (1,1) model for forecasting conditional covariances and correlations, testing for cross sector correlation within the OMX30 using Atlas Copco B and Svenska Handelsbanken A and see if there is significant correlation. We compare Gaussian and multivariate Student's t innovations on daily returns from 2014 to 2024, sourced from **Yahoo Finance**. All analysis is in **R** with `tidyquant` for data and `rugarch` and `rmgarch` for estimation. Model adequacy is assessed through Value at Risk (VaR) backtests using the tests of Kupiec and Christoffersen.

The thesis is structured as follows. Chapter 2 reviews univariate and multivariate GARCH theory. Chapter 3 describes the forecasting procedure, and Chapter 4 presents evaluation methods. Chapter 5 reports the empirical results. Chapter 6 assesses goodness of fit, and Chapter 7 concludes with a discussion.

2 Theoretical Framework

2.1 General Definitions

Before proceeding with the theoretical development, it is essential to establish a set of general definitions and notational conventions that will be employed throughout this section. These definitions serve to clarify the mathematical constructs and assumptions underlying the models and methodologies discussed in the subsequent analysis.

To make the generall defintions of all modells more clear the most vital notations are going to be described below:

- r_t: asset return at time t,
- $y_t = r_t - \mu_t$: mean-corrected return at time t,
- μ_t: conditional mean of r_t,
- ε_t: i.i.d. innovation with $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^2] = 1$,
- h_t: conditional variance of r_t (univariate case),
- H_t: conditional covariance matrix of r_t (multivariate case),
- α_i, β_j: model parameters,
- q, p: orders of ARCH/GARCH lags.

Note that the modelling of μ_t will not be of focus in this thesis and be set as a constant in this thesis namely $\mu_t = 0$.

2.1.1 ARCH model

Before the Autoregressive Conditional Heteroscedasticity (ARCH) process was introduced by Robert F. Engle in 1982 (1), traditional econometric models assumed constant one-period variance. With the introduction of the ARCH model, the conditional variance could be modelled as a function of past errors, leaving the unconditional variance constant. The ARCH processes are zero-mean, serially uncorrelated processes with non-constant variance conditional on the past. Hence, information about the recent past can, with an ARCH model, describe future variance. An ARCH(1) process, y_t, is defined in Tsay (2005, p. 103) (12) as,

$$y_t = \sqrt{h_t}\varepsilon_t \tag{1}$$

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

where y_t is a zero-mean process at time t, ε_t is the error term at time t, h_t is the conditional variance at time t, α_t is the model parameter for lag t, and the parameters satisfy

$$\alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \alpha_1 > 0$$

It is common to assume ε_t follows either a standard normal, or a standardized Student-t distribution. The ARCH model can be generalized to ARCH(p), where p is the number of past terms in the process that should be included in the model where, that is

$$y_t = \sqrt{h_t}\,\varepsilon_t \tag{3}$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2. \tag{4}$$

similarly as above

$$\alpha_0 > 0, \quad \alpha_i \geq 0, \quad \sum_{i=1}^{p} \alpha_i < 1 \quad (i = 1, \dots, p).$$

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

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

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_q y_{t-p}^2 \tag{6}$$

2.1.2 Generalized Autoregressive Conditional Heteroskedasticity

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process is an econometric model introduced by Engle (1) and later extended by Bollerslev (3). It provides a framework for modeling and forecasting time-varying volatility in financial return series.
Consider a series of daily asset returns, $r_1, \dots, r_n$, derived from the logarithmic differences of an observed price or index series $\{S_t\}_{t=0}^{n}$. These returns are calculated as:

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right). \tag{7}$$

The GARCH model assumes that the return at time t, r_t, is conditionally normally distributed given the information up to time $t-1$, with conditional mean μ_t and conditional variance $h_t = \sigma_t^2$, i.e.,

$$r_t \mid \mathcal{F}_{t-1} \sim N(\mu_t, h_t), \tag{8}$$

where $\mathcal{F}_{t-1}$ denotes the information set available at time $t-1$.

The hallmark of GARCH models is that the conditional variance h_t is time-varying and depends on past squared returns and past conditional variances. The GARCH(1,1) model, which is the most commonly used specification in empirical finance, is given by:

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1}, \tag{9}$$

where $\alpha_0 > 0$ is a constant term, $\alpha_1 > 0$ captures the short-term reaction of volatility to shocks in returns, and $\beta > 0$ measures the persistence of volatility from one period to the next.

For the model to be stationary, i.e., for the variance process to revert to a long-run average level, the parameters must satisfy:

$$\alpha_1 + \beta < 1. \tag{10}$$

Under this condition, the unconditional (long-run) variance is given by:

$$h = \frac{\alpha_0}{1 - \alpha_1 - \beta}. \tag{11}$$

The GARCH model also provides a framework for volatility forecasting. The one-step ahead forecast of the conditional variance is:

$$\mathbb{E}_{t-1}(r_t^2) = h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1}. \tag{12}$$

For forecasting multiple steps ahead, the model relies on its recursive structure. The two-step ahead forecast becomes:

$$\mathbb{E}_{t-1}(h_{t+1}) = \alpha_0 + (\alpha_1 + \beta) h_t, \tag{13}$$

and more generally, the k-step ahead forecast is:

$$\mathbb{E}_t(h_{t+k}) = \alpha_0 + \gamma \mathbb{E}_t(h_{t+k-1}), \tag{14}$$

where $\gamma = \alpha_1 + \beta$. This recursive forecasting process shows that as $k \to \infty$, the forecasted variance converges to the unconditional variance h, provided that $\gamma < 1$:

$$\lim_{k \to \infty} \mathbb{E}_t(h_{t+k}) = \frac{\alpha_0}{1 - \alpha_1 - \beta}. \tag{15}$$

The GARCH(1,1) model effectively captures volatility clustering, a common empirical feature of financial time series whereby large returns are often followed by large returns. However, it assumes symmetric responses to positive and negative shocks, which may limit its ability to capture certain asymmetries observed in real-world financial data.

2.1.3 Multivariate GARCH

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

- r_t is the log return at time t of n assets ($n \times 1$ vector),
- μ_t is the conditional mean of r_t at time t ($n \times 1$ vector),
- $y_t = r_t - \mu_t$ is the mean-corrected return at time t ($n \times 1$ vector),
- H_t is the conditional covariance matrix of y_t (equivalently, of r_t given Ψ_{t-1}) ($n \times n$ matrix),
- ε_t is an i.i.d. innovation vector with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t \varepsilon_t'] = I_n$.

Consider a stochastic return vector $r_t \in \mathbb{R}^{n\times 1}$. Let Ψ_{t-1} denote the information set generated by the observed series of r_t up to time $t-1$, and assume that r_t is conditionally heteroscedastic (5). The definition of the mean-corrected returns by

$$y_t = r_t - \mu_t, \tag{16}$$

and model them as

$$\varepsilon_t \mid \Psi_{t-1} \sim \text{i.i.d. } (0, I_n), \qquad y_t = H_t^{1/2}\,\varepsilon_t. \tag{17}$$

Here, $H_t^{1/2}$ is obtained from H_t via Cholesky decomposition. There are many ways of computing H_t. Different methods of computing the conditional covariance matrix will be discussed in the next section. The vector μ_t can be modelled as a function of time or as a constant. In this thesis, the modelling of μ_t will not be in focus and will be treated as a constant vector.

2.1.4 Introduction to the multivariate GARCH model

Silvennoinen and Terasvirta (5) categorize multivariate GARCH models into four main classes. This section reviews these classes, each representing a distinct approach to modeling the conditional covariance matrix H_t in Equation (17).

A central challenge in MGARCH modeling is to balance parsimony with flexibility, since the number of parameters increases rapidly with the number of assets n. In addition, the positive definiteness of the covariance matrices must be guaranteed. This is often ensured through eigenvalue decomposition, which can become computationally demanding in high-dimensional systems. Consequently, MGARCH specifications that minimize matrix inversions are particularly desirable.

The modeling of the conditional covariance matrix has been a central theme in the development of MGARCH models. One of the earliest approaches is the VEC model, introduced by Engle, Wooldridge, and Bollerslev (6) in 1988, as a direct generalization of the univariate GARCH framework. In this specification, both variances and covariances are expressed as functions of their own past values and lagged conditional covariances:

$$\mathrm{VECH}(H_t) = c + \sum_{j=1}^{q} A_j \,\mathrm{VECH}(y_{t-j}y'_{t-j}) + \sum_{j=1}^{p} B_j \,\mathrm{VECH}(H_{t-j}) \tag{18}$$

In this specification, VECH$(\cdot)$ denotes the half vectorization operator applied to symmetric matrices. The term c is an $n(n+1)/2 \times 1$ vector, while A_j and B_j are parameter matrices. Although the VEC model is very general, it is computationally demanding and requires additional restrictions in order to guarantee that H_t remains positive definite. To address these difficulties, Engle and Kroner introduced the BEKK model in 1995 (7), which secures the positive definiteness of H_t by construction:

$$H_t = CC' + \sum_{j=1}^{q}\sum_{k=1}^{K} A'_{kj} y_{t-j} y'_{t-j} A'_{kj} + \sum_{j=1}^{p}\sum_{k=1}^{K} B'_{kj} H_{t-j} B'_{kj} \tag{19}$$

In this formulation, C, A_{kj}, and B_{kj} represent parameter matrices.

An alternative approach is offered by factor models, which are motivated by asset pricing theory. The earliest factor based GARCH model was proposed by Engle, Ng and Rothschild in 1990 (8). Within this framework, the underlying factors are assumed to be conditionally heteroskedastic and can themselves be modeled using GARCH processes:

$$H_t = \Omega + \sum_{k=1}^{K} w_k w'_k f'_{k,t} \tag{20}$$

In this context, Ω denotes a positive semi-definite $n \times n$ matrix, w_k are the portfolio weight vectors, and $f_{k,t}$ represent the factor-specific variances. Among the factor-based GARCH models, one of the most widely applied is the GO-GARCH model, introduced by Van der Weide in 2002 (10), which is characterized by the assumption of uncorrelated factors:

$$H_t = W H_t^z W = \sum_{k=1}^{N} w_k w'_k h^z_{k,t} \tag{21}$$

In this formulation, the vectors w_k correspond to the columns of the matrix W, while $h^z_{k,t}$ represent the diagonal elements of H^z_t. The key distinction between Equations (20) and (21) lies in the latter's assumption of uncorrelated factors.

Non-parametric and semi-parametric approaches offer alternatives to parametric models by not imposing strict structure on the conditional covariance matrix. A method was developed which was an approach that begins with estimating a parametric MGARCH model, extracting standardized residuals $\hat{\eta}_t$, and then using a kernel-weighted estimator to capture any remaining structure:

$$H_t = \hat{H}_t^{1/2} \frac{\sum_{r=1}^{T} \hat{\eta}_r \hat{\eta}_r' K_h(s_t - s_r)}{\sum_{r=1}^{T} K_h(s_t - s_r)} \hat{H}_t^{1/2} \tag{22}$$

In this context, $\hat{H}_t$ denotes the conditional covariance obtained from the initial MGARCH estimation, $s_t \in \Psi_{t-1}$ represents an observed variable within the available information set, and $K_h(\cdot) = K(\cdot/h)/h$ refers to a kernel function with bandwidth h.

Subsequently, models that focus on conditional variances and correlations decompose the conditional covariance matrix into its two components: conditional standard deviations and conditional correlations. Within this class of models, the Constant Conditional Correlation (CCC) model and the Dynamic Conditional Correlation (DCC) model are the most prominent. The DCC specification, which is the subject of the next section, extends the CCC framework by allowing correlations to vary over time while preserving computational feasibility.

2.1.5 CCC-GARCH model

The Constant Conditional Correlation (CCC) model, introduced by Bollerslev in 1990 (11), is a multivariate time series framework that allows conditional variances and covariances to vary over time, while assuming correlations remain constant. Under this assumption, the conditional covariance matrix H_t is expressed as:

$$H_t = D_t R D_t \tag{23}$$

Here, R is the constant correlation matrix with elements $R = \rho_{ij}$, where $\rho_{ii} = 1$ for $i = 1, \dots, n$, and n is the number of assets. D_t is a diagonal matrix containing the conditional standard deviations $\sqrt{h_{it}}$ for $i = 1, \dots, n$ of the fitted univariate GARCH models for each asset, i.e.,

$$D_t = \operatorname{diag}\left(\sqrt{h_{1t}}, \dots, \sqrt{h_{it}}\right)$$

The off-diagonal elements of H_t are given by:

$$[H_t]_{ij} = \sqrt{h_{it}}\sqrt{h_{jt}}\rho_{ij}, \quad i \neq j \tag{24}$$

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

$$h_t = \omega + \sum_{j=1}^{q} A_j y^*_{t-j} + \sum_{j=1}^{p} B_j h_{t-j} \tag{25}$$

In this formulation, $y^*_{t-j} = y_{t-j} \circ y_{t-j}$, where $\circ$ represents the Hadamard (element-wise) product. The vector ω has dimension $n \times 1$, while A_j and B_j are diagonal matrices of size $n \times n$.

For the conditional covariance matrix H_t to remain positive definite, it is necessary that the correlation matrix R is positive definite and that the diagonal entries of A_j and B_j are strictly positive. This condition, however, is only essential in the special case where $p = q = 1$, as noted by Nelson and Cao (2).

Although the CCC-GARCH model offers a parsimonious parameterization, its assumption of constant correlations is often considered too restrictive in practice. Empirical evidence, such as the study by Chevallier (9), highlights this limitation and provides motivation for adopting more flexible specifications, most notably the DCC-GARCH model.

2.1.6 DCC-GARCH model

The Dynamic Conditional Correlation model (DCC-GARCH) was introduced by Engle and Sheppard in 2001 as an extension of the CCC-GARCH model (13). The DCC model uses Equation (23), but instead of modelling R as a constant matrix, it is modelled dynamically, with R_t depending on time t, such that:

$$H_t = D_t R_t D_t \tag{26}$$

The off-diagonal elements in H_t then follow the structure below and the Equation (24) becomes:

$$[H_t]_{ij} = \sqrt{h_{it}}\sqrt{h_{jt}}\rho_{ij,t} \tag{27}$$

The matrix D_t in Equation (26) is defined as:

$$D_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 & \cdots & 0 \\ 0 & \sqrt{h_{2t}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{h_{nt}} \end{bmatrix} \tag{28}$$

Each h_{it} is defined as:

$$h_{it} = \omega_{i0} + \sum_{j=1}^{q_i} \alpha_{ij} y_{i,t-j}^2 + \sum_{j=1}^{p_i} \beta_{ij} h_{i,t-j}$$

For models specified as in Equation (26), the positive definiteness of H_t is ensured only if both the conditional variances h_{it} (for $i = 1, \ldots, n$) and the correlation matrix R_t remain positive definite at all time points. Compared to the CCC model, this requirement increases the computational burden of the DCC model, since R_t must be continuously updated and inverted at each step.

The dynamic conditional correlation structure introduced by Engle is given by:

$$Q_t = (1 - \alpha - \beta)S + \alpha\, \varepsilon_{t-1}\varepsilon_{t-1}' + \beta Q_{t-1} \tag{29}$$

$$R_t = (I \circ Q_t^{-1/2}) Q_t (I \circ Q_t^{-1/2}) \tag{30}$$

In Equation (29), Q_t is the evolving covariance matrix, and S is the unconditional correlation matrix of the standardized residuals ε_t, defined as:

$$S = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t'$$

This follows from the assumption that $S = \text{Cov}[\varepsilon_t \varepsilon_t'] = \mathbb{E}[\varepsilon_t \varepsilon_t']$.

Parameters α and β are the DCC-GARCH parameters satisfying: $\alpha > 0$, $\beta > 0$, and $\alpha + \beta < 1$. This process ensures positive definiteness of Q_t, though R_t must be normalized to yield valid correlation matrices. Specifically, the normalization in Equation (30) rescales Q_t using a diagonal matrix:

$$I \circ Q_t^{-1/2} = \begin{bmatrix} \sqrt{q_{1t}} & 0 & \cdots & 0 \\ 0 & \sqrt{q_{2t}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{q_{nt}} \end{bmatrix} \tag{31}$$

This rescaling ensures that $\rho_{ij,t} \leq 1$ and $\rho_{ii,t} = 1$, maintaining the validity of the correlation matrix R_t.

Finally, under the assumption of conditional normality, the mean-corrected return vector y_t can be expressed as:

$$y_t \mid \Psi_{t-1} \sim N(0, D_t R_t D_t) = N(0, H_t) \tag{32}$$

This assumption facilitates likelihood-based estimation. Even when modeling y_t with Student's t-distributed errors, the normality assumption is often retained in the context of the two-stage Quasi Maximum Likelihood Estimation (QMLE), which will be discussed in the following section.

3 Forecasting Procedure and Evaluation

Engle and Sheppard (13) proposed a two-step estimation procedure for the parameters of the DCC model, based on a decomposition of the quasi-likelihood function. The idea is to separate the estimation into two components: a volatility part and a correlation part. In the first step, the volatility parameters are estimated. These results are then used as inputs in the second step, where the correlation parameters are obtained.

An important feature of this approach is that the first-stage volatility estimation is unaffected by the choice of innovation distribution.As demonstrated in earlier studies (see e.g. (18)), altering the univariate distribution in the volatility stage does not influence the correlation parameters estimated in the second stage.

In the following, the two-step quasi-likelihood estimation procedure will be presented for both Gaussian innovations and multivariate Student's t innovations.

3.1 Gaussian distributed innovations

First, we define the multivariate Gaussian distribution according to (17):

$$f(\varepsilon_t) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\varepsilon_t'\varepsilon_t\right) \tag{33}$$

Consider a mean-corrected return vector y_t of dimension n, with mean μ and conditional covariance matrix H_t as defined in Equations (17) and (16). Under these assumptions, y_t follows a multivariate normal distribution, i.e. $y_t \sim MN(\mu, H_t)$. By the properties of linear transformation and scaling, the error term $\varepsilon_t = y_t - \mu_t$ also satisfies $\varepsilon_t \sim MN(0, H_t)$. Here $t = 1, \ldots, T$ is the time period used to estimate the model.

The corresponding likelihood function for the errors at time t is given by:

$$L(\theta) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{n/2}|H_t|^{1/2}} \exp\left(-\frac{1}{2}\varepsilon_t' H_t^{-1}\varepsilon_t\right) \tag{34}$$

Note that θ denotes the parameters of the model. $|H_t|$ is the determinant of H_t. Given Equation (34), the log-likelihood estimator can be written as:

$$
\begin{aligned}
ln(L(\theta)) &= -\frac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi)+\log|H_t|+y_t'H_t^{-1}y_t\right)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi)+\log|R_tD_tR_t|+y_t'D_t^{-1}R_t^{-1}D_t^{-1}y_t\right)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi)+2\log|D_t|+\log|R_t|+\varepsilon_t'R_t^{-1}\varepsilon_t\right)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi)+2\log|D_t|+y_t'D_t^{-1}D_t^{-1}y_t'-\varepsilon_t'\varepsilon_t log|R_t|+\varepsilon_t'\varepsilon_t R_t\right)
\end{aligned}
\tag{35}
$$

$$\text{where } \varepsilon_t = D_t^{-1}y_t \quad \text{and} \quad H_t = D_tR_tD_t.$$

If we denote the parameters in D_t by θ and the parameters in R_t by ϕ, the log-likelihood function can be decomposed into a volatility part and a correlation part:

$$L(\theta,\phi) = L_v(\theta) + L_c(\theta,\phi) \tag{36}$$

Where:

$$ln(L_v(\theta)) = -\frac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi)+2\log|D_t|^2+y_t'D_t^{-2}y_t'\right) \tag{37}$$

$$ln(L_c(\theta,\phi)) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|R_t|+\varepsilon_t'R_t^{-1}\varepsilon_t-\varepsilon_t'\varepsilon_t\right) \tag{38}$$

Furthermore, the volatility part $L_v(\theta)$ can be expressed as the sum of univariate GARCH log-likelihoods:

$$ln(L_v(\theta)) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\left(\log(2\pi)+\log\left((h_{i,t})+\frac{y_{i,t}^2}{h_{i,t}}\right)\right) \tag{39}$$

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

$$\text{Step 1: } \hat{\theta} = \max L_v(\theta), \quad \text{Step 2: } \hat{\phi} = \max L_c(\hat{\theta},\phi) \tag{40}$$

This two-stage estimation procedure is used to estimate Gaussian distribution innovations within the Quasi Maximum Likelihood framework.

3.2 Multivariate Student's t distributed

As noted previously, the volatility component $L_v(\theta)$ is identical for both Gaussian and multivariate Student's t errors. The distinction arises in the second stage of the likelihood estimation, where the Student's t distribution requires a different specification. Therefore, the correlation component $L_c(\theta, \phi)$ must be adapted for this case.

First, we define the multivariate Student's t distribution:

$$f(\varepsilon_t \mid \nu) = \prod_{t=1}^{T} \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)[\pi(\nu-2)]^{n/2}} \left[1 + \frac{\varepsilon_t' \varepsilon_t}{\nu-2}\right]^{-\frac{\nu+n}{2}} \tag{41}$$

Here, $\Gamma(\cdot)$ denotes the Gamma function. Unlike the Gaussian distribution, the multivariate Student's t distribution introduces an additional parameter, ν, which governs the heaviness of the tails. This parameter appears because the Student's t distribution can be viewed as a mixture of a multivariate Gamma and a multivariate Normal distribution. As a result, it preserves symmetric tail dependence while allowing for heavier tails than the Gaussian case.

Formally, if the mean-corrected return vector y_t has mean μ_t, conditional covariance matrix H_t, and shape parameter ν, then it follows a multivariate Student's t distribution:

$$y_t \sim MT(\mu_t, \Omega_t, \nu)$$

where Ω_t is a scale matrix such that $H_t = \frac{\nu}{\nu-2}\Omega_t$. The likelihood function of $y_t = H_t^{1/2}\varepsilon_t$ at time t of the errors is then:

$$L(\theta) = \prod_{t=1}^{T} \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)[\pi(\nu-2)]^{n/2}|\Omega_t|^{1/2}} \left[1 + \frac{\varepsilon_t' \Omega_t^{-1} \varepsilon_t}{\nu-2}\right]^{-\frac{\nu+n}{2}} \tag{42}$$

Given Equation (42), the corresponding log-likelihood estimator is:

$$\begin{aligned} ln(L(\theta)) = \sum_{t=1}^{T} \Bigg(\log\Gamma\left(\frac{\nu+n}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2}\log(\pi(\nu-2)) \\ - \frac{1}{2}\log|D_t R_t D_t| - \frac{\nu+n}{2}\log\left(1 + \frac{y_t' D_t^{-1} R_t^{-1} D_t^{-1} y_t}{\nu-2}\right)\Bigg) \end{aligned} \tag{43}$$

$$\text{where} \quad H_t = D_t R_t D_t.$$

As with the Gaussian case, we adopt a two-step estimation procedure. Step 1 is identical to the Gaussian likelihood estimation, and step 2 includes the additional shape parameter ν:

$$\text{Step 1: } \hat{\theta} = \max L_v(\theta), \quad \text{Step 2: } \hat{\phi}, \hat{\nu} = \max L_c(\hat{\theta}, \phi, \nu) \tag{44}$$

Since the variance parameters in $L_v(\theta)$ are already estimated in Step 1, we can treat D_t as constant and exclude it from the second-stage likelihood. As such, we simplify $L_c(\hat{\theta}, \phi, \nu)$ as follows:

$$\begin{aligned} ln(L_c(\hat{\theta}, \phi, \nu)) = \sum_{t=1}^{T} \Bigg(& \log \Gamma\left(\frac{\nu + n}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2}\log(\pi(\nu - 2)) \\ & - \frac{1}{2}\log|R_t| - \frac{\nu + n}{2}\log\left(1 + \frac{\varepsilon_t' R_t^{-1} \varepsilon_t}{\nu - 2}\right)\Bigg) \end{aligned} \tag{45}$$

This formulation represents the second stage likelihood estimation assuming multivariate Student's t distribution innovations.

3.3 Forecasting GARCH methods

Forecasting variances is a fundamental requirement of GARCH models. Most GARCH models provide a convenient framework for generating k-step ahead forecasts. For a univariate GARCH(1,1) process which is given by (18) we get the following:

Recall $H_t = D_t R_t D_t$ with $D_t = \text{diag}\big(\sqrt{h_{1,t}}, \ldots, \sqrt{h_{n,t}}\big)$ and R_t the conditional correlation matrix.

Step one, univariate variance forecasts. The diagonal entries of D_{t+k} come from the univariate variance forecasts, which can be computed asset by asset. For the general GARCH(p,q) model the forecast procedure becomes,

$$\mathbb{E}_t[h_{i,t+k}|F_t] \;=\; \omega_0 + \sum_{j=1}^{max(p,q)} \big(\alpha_j + \beta_j\big)\, \mathbb{E}_t[h_{i,t+k-j}|F_t],$$

where $\mathbb{E}_t[h_{i,t+k}|F_t] = y^2_{i,t+k}$ for $k < 0, i = 1, \ldots, n$

For the widely used GARCH$(1,1)$ the k-step ahead forecasting procedure becomes,

$$\mathbb{E}_t[h_{i,t+k}] \;=\; \sum_{i=0}^{k-2} \omega_0 \left(\alpha_1 + \beta_1\right)^i \;+\; \left(\alpha_1 + \beta_1\right)^{k-1} \mathbb{E}_t[h_{i,t+1}|F_t].$$

where

$$\mathbb{E}_t[h_{i,t+k}|F_t] = \omega_0 + \alpha_1 y^2_{i,t+k} + \beta_1 h_{i,t}$$

Note that in theory the memory will decline with a exponential rate $(\alpha_1 + \beta_1)$. The forecast of the conditional variance becomes the following:

$$\mathbb{E}_t[D_{t+k}|F_t] = \left(\text{diag}\sqrt{E_t[h_{1,t+k}|F_t]}, \ldots, \sqrt{E_t[h_{n,t+k}|F_t]}\right)$$

Step two, conditional correlation forecasts. The elements in the conditional correlation matrix, 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. $\hat{\rho}_{ij} = \frac{\hat{q}_{ij}}{\sqrt{\hat{q}_{ii}\hat{q}_{jj}}}$, where $\hat{q}_{ij}$, $\hat{q}_{ii}$ and $\hat{q}_{jj}$ are the forecast elements in Q_{t+k}. Thus is the expectation of Q_{t+k}:

$$\begin{cases} (1-\alpha-\beta)S + \alpha\, \varepsilon_t \varepsilon_t' + \beta\, Q_t, & \text{for } k = 1, \\ (1-\alpha-\beta)S + a\, E[\varepsilon_{t+k-1}\varepsilon_{t+k-1}' \mid F_t] + \beta\, E[Q_{t+k-1} \mid F_t], & \text{for } k > 1. \end{cases} \tag{46}$$

where

$$E[\varepsilon_{t+k-1}\varepsilon'_{t+k-1}|F_t] = E[R_{t+k-1}|F_t] = E[(I \circ Q^{-1/2}_{t+k-1})Q_{t+k-1}(I \circ Q^{-1/2}_{t+k-1})|F_t]$$

Due to this dependency, the k-step ahead forecast cannot be computed directly. Engle and Sheppard (13) proposed two different approaches for forecasting the DCC parameters. Since $E[(I \circ Q^{-1/2}_{t+k-1})Q_{t+k-1}(I \circ Q^{-1/2}_{t+k-1})|F_t]$ is unknown, we cannot directly compute the k-step ahead forecast in 46. However, there exists two methods that approximates this forecast,

1. Method 1; Assumes that $E[\varepsilon_{t+i}\varepsilon'_{t+i} \mid F_t] \approx E[Q_{t+i} \mid F_t]$ for $i = 1, \ldots, k$.
2. Method 2; Assumes that $\bar{R} \approx S$ and $E[R_{t+i} \mid F_t] \approx E[Q_{t+i} \mid F_t]$ for $i = 1, \ldots, k$.

Method 1.

$$E\big[\varepsilon_{t+i}\varepsilon'_{t+i} \mid F_t\big] \approx E[Q_{t+i} \mid F_t] \quad \text{for } i = 1, \ldots, k.$$

For $k > 1$,

$$\begin{aligned}
E[Q_{t+k} \mid F_t] &= (1-\alpha-\beta)S \;+\; \alpha\, E\big[\varepsilon_{t+k-1}\varepsilon'_{t+k-1} \mid F_t\big] \;+\; \beta\, E[Q_{t+k-1} \mid F_t] \\
&\approx (1-\alpha-\beta)S \;+\; (\alpha+\beta)\, E[Q_{t+k-1} \mid F_t] \\
&\approx (1-\alpha-\beta)S \;+\; (\alpha+\beta)\Big[(1-\alpha-\beta)S \;+\; (\alpha+\beta)\, E[Q_{t+k-2} \mid F_t]\Big] \\
&= (1-\alpha-\beta)S\sum_{i=0}^{k-2}(\alpha+\beta)^i \;+\; (\alpha+\beta)^{k-1}\, E[Q_{t+1} \mid F_t] \\
&= \Big(1-(\alpha+\beta)^{k-1}\Big)S \;+\; (\alpha+\beta)^{k-1}\, E[Q_{t+1} \mid F_t] \;=\; \widehat{Q}_{t+k}.
\end{aligned}$$

From 46 we get that,

$$E[Q_{t+1} \mid F_t] = (1-\alpha-\beta)S + \alpha\, \varepsilon_t\varepsilon'_t + \beta\, Q_t.$$

Then

$$\widehat{R}_{t+k} \;=\; E[R_{t+k} \mid F_t] \;\approx\; (I \circ \widehat{Q}^{-1/2}_{t+k})\, \widehat{Q}_{t+k}\, (I \circ \widehat{Q}^{-1/2}_{t+k}),$$

Method 2, Assume that $\bar{R} \approx S$ and $E[R_{t+i} \mid F_t] \approx E[Q_{t+i} \mid F_t]$ for $i = 1, \ldots, k$. For $k > 1$,

$$\begin{aligned}
E[R_{t+k} \mid F_t] &\approx E[Q_{t+k} \mid F_t] \\
&= (1-\alpha-\beta)\, S \;+\; \alpha\, E[R_{t+k-1} \mid F_t] \;+\; \beta\, E[Q_{t+k-1} \mid F_t] \\
&\approx (1-\alpha-\beta)\, \bar{R} \;+\; (\alpha+\beta)\, E[R_{t+k-1} \mid F_t] \\
&\approx (1-\alpha-\beta)\, \bar{R} \;+\; (\alpha+\beta)\Big[(1-\alpha-\beta)\, \bar{R} \;+\; (\alpha+\beta)\, E[R_{t+k-2} \mid F_t]\Big] \\
&= (1-\alpha-\beta)\, \bar{R}\sum_{i=0}^{k-2}(\alpha+\beta)^i \;+\; (\alpha+\beta)^{k-1}E[R_{t+1} \mid F_t] \\
&= \big(1-(\alpha+\beta)^{k-1}\big)\bar{R} \;+\; (\alpha+\beta)^{k-1}E[R_{t+1} \mid F_t].
\end{aligned}$$

and finally, $E[R_{t+1} \mid F_t] \;\approx\; (I \circ \widehat{Q}_{t+k}^{-1/2})\, \widehat{Q}_{t+k}\, (I \circ \widehat{Q}_{t+k}^{-1/2}), \widehat{Q}_{t+1} = (1 - \alpha - \beta)\, S + \alpha\, \varepsilon_t \varepsilon_t' + \beta\, Q_t,$

With the notation $\widehat{H}_{t+k} = E[H_{t+k} \mid F_t]$, $\widehat{R}_{t+k} = E[R_{t+k} \mid F_t]$, and $\widehat{D}_{t+k} = E[D_{t+k} \mid F_t]$, the k step ahead covariance forecast is

$$\widehat{H}_{t+k} \;=\; \widehat{D}_{t+k}\, \widehat{R}_{t+k}\, \widehat{D}_{t+k}.$$

4 Methodology

4.1 Evaluation of estimations

To assess the model's goodness of fit, it is necessary to evaluate the marginal and multivariate distributions separately. This section outlines the methods employed for these evaluations.

4.1.1 Goodness of fit for univariate distributions

A visual inspection of the residuals provides an intuitive way to assess their behaviour. If the residuals are independently and identically distributed (IID), the plot should display no visible patterns and appear random.

4.1.2 The Auto Correlation Function

The autocorrelation function (ACF) of a stochastic process X is defined as (14):

$$R_X(t) = \mathrm{Cov}[X(s), X(s+t)].$$

For a sample of size n, the process ε_t can be considered independently and identically distributed (IID) if approximately 5% of the autocorrelation lags fall outside the 95% confidence bounds.

4.1.3 Ljung-Box test

A central task in time series analysis is to examine the presence of autocorrelation, i.e., whether correlations exist between observations at different lags.

One widely used method for this purpose is the **Ljung-Box test**. To introduce the test, we first define the sample autocorrelation at lag l for a return series $\{y_t\}_{t=1}^{T}$:

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{T}(y_t - \bar{y})(y_{t-l} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}, \quad 0 \leq l < T,$$

where $\bar{y} = \frac{1}{T}\sum_{i=1}^{T} y_i$ (Tsay, 2005, p. 26).

Using these sample autocorrelations, the Ljung-Box statistic for m lags in a sample of size T is defined as:

$$Q(m) = T(T+2)\sum_{l=1}^{m}\frac{\hat{\rho}_l^2}{T-l},$$

see Tsay (2005, p. 27). The $Q(m)$ statistic tests the joint null hypothesis that all autocorrelations up to lag m are zero. A common choice for m is $m \approx \ln(T)$.

The null hypothesis is that:

$$\hat{\rho}_1 = \hat{\rho}_2 = \cdots = \hat{\rho}_m = 0.$$

Under the null hypothesis, the statistic $Q(m)$ follows asymptotically a chi-squared distribution with m degrees of freedom. Consequently, if the corresponding p-value is smaller than the chosen significance level (commonly 5%), the null hypothesis is rejected. This outcome indicates the presence of autocorrelation in the series.

4.2 Goodness of fit for multivariate distributions

While the univariate distributions may be adequately specified, this does not necessarily guarantee that the corresponding multivariate distributions are well defined. To address this, two backtesting procedures for Value-at-Risk (VaR) are considered in this section. These tests provide a framework for assessing whether the tails of the portfolio return distribution conform to the assumed model.

4.2.1 Value at Risk

We begin by defining Value-at-Risk (VaR). VaR is a risk measure that quantifies the maximum expected loss of a portfolio over a specified time horizon, given a certain probability level. Denote this probability level by α, and let l represent the corresponding loss threshold. In this context, l is interpreted as the VaR at confidence level α.

Extending VaR to a multivariate setting requires knowledge of the underlying portfolio return distribution. As discussed in Section 2.1.3, the multivariate normal distribution exhibits linearity properties that facilitate this analysis. Recall that the error terms were defined as $\varepsilon_t \sim MN(0, H_t)$. For portfolio returns $\mathbf{R}_{p,t}$, we have

$$\mathbf{R}_{p,t} = \mathbf{w}\varepsilon_t + \mathbf{w}\mu_t,$$

where $\mathbf{w}$ denotes the portfolio weights, r_t represents the returns, and ε_t are the Gaussian error terms. Consequently, under the assumption of multivariate normality, the distribution of the portfolio returns becomes:

$$R_{p,t,MN} \sim MN(\mathbf{w}\mu_t, \mathbf{w}' H_t \mathbf{w}). \tag{42}$$

Similarly, in Section 2.1.3, the errors of the multivariate Student's t-distribution were defined as $y_t \sim MT(\mu_t, \Omega_t, \nu)$. The portfolio returns $R_{p,t}$ under the multivariate Student's-t assumption are then given by

$$R_{p,t,MT} \sim MT(w\mu_t, w' H_t w, \nu). \tag{43}$$

Here, w is assumed to be equally distributed among the stock pairs.

4.2.2 Kupiec's test

The Kupiec's test is designed to assess whether the observed number of Value-at-Risk (VaR) violations is consistent with a specified confidence level α (15). The null hypothesis, H_0, states that the frequency of violations matches the expected rate implied by α. In contrast, the alternative hypothesis, H_α, posits that the observed violation rate deviates from the expected level. Since the occurrence of violations can be modelled as a binomial process, the test is formulated as a likelihood ratio test. The corresponding Kupiec test statistic is expressed as:

$$L_{KT} = -2\log\left((1-p)^{N-x}p^{x}\right) + 2\log\left((1-x)^{N-x}\left(\frac{x}{N}\right)^{x}\right), \tag{47}$$

If H_0 is true, the test statistic asymptotically follows a chi-squared distribution with 1 degree of freedom, i.e. $L_{KT} \sim \chi^2(1)$. Large deviations from the expected number of violations indicate that the model may not be correctly specified (16).

4.2.3 Christoffersen's test

The Christoffersen's test examines whether the likelihood of observing a violation on a given day is influenced by the occurrence of a violation on the preceding day. In contrast to unconditional coverage tests, which only assess the overall frequency of violations, the Christoffersen's test explicitly evaluates the dependence structure between consecutive violations. As such, the test relies on a Markov chain framework to capture the potential clustering of exceptions.

The hypotheses are defined as:

$$\begin{aligned} H_0 &: \text{Failures are independently and identically distributed (IID)} \\ H_\alpha &: \text{Failures are not IID} \end{aligned}$$

The corresponding likelihood ratio test statistic is:

$$L_{CT} = 2\ln\left(\frac{(1-\pi_{01}^{n_{00}})\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}}{\alpha^{x}(1-\alpha)^{n-x}}\right), \tag{48}$$

where:

- n_{ij} is the number of transitions from state i to state j, where $i, j \in \{0, 1\}$, and each state represents failure (1) or non-failure (0)
- π_{01} is the conditional probability of a failure following a non-failure

- π_{11} is the conditional probability of a failure following a failure
- α is the expected failure rate
- x is the total number of failures
- n is the total number of observations

4.3 Statstical Time Series test

Model validation plays a crucial role in determining whether a GARCH-type model is appropriate for financial time series analysis. A key requirement of these models is that the underlying return series must be stationary. To assess this property, the Augmented Dickey-Fuller (ADF) test is applied to the log-return series of each stock. The test evaluates the presence of a unit root, where rejection of the null hypothesis implies stationarity. If the null is rejected, the series can be considered suitable for volatility modeling within the GARCH and DCC-GARCH frameworks.

4.3.1 Augmented Dickey-Fuller test

The whole analysis will be conducted under the assumption that the time series of daily log returns is stationary. Therefore, it is crucial to test whether this assumption is reasonable. To verify stationarity of a series $\{x_t\}$, we apply the Augmented Dickey-Fuller (ADF) test. The ADF regression is defined as:

$$\Delta x_t = \alpha_t + (\beta - 1)x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{49}$$

where $\Delta x_j = x_j - x_{j-1}$, ϕ_i are constants, α_t is a function of time (constant or trend), and ε_t is assumed to be white noise (12).

The null hypothesis of the test is:

$$H_0 : \beta = 1,$$

which implies the presence of a unit root, that is, the series is non stationary, and each shock ε_t has a permanent effect. If we set $\alpha_t = 0$ for simplicity, the ADF equation reduces to:

$$x_t = x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{50}$$

This is a random walk, where the innovations ε_t serve as nonstationary increments.

Under the alternative hypothesis:

$$H_1 : \beta < 1,$$

the model becomes:

$$x_t = \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{51}$$

Since $|\beta| < 1$, the influence of past values x_{t-1} decays geometrically and the shocks ε_t dissipate over time, resulting in a stationary process.

Moreover, the ADF test statistic is defined as:

$$t_\beta = \frac{\hat{\beta} - 1}{\text{std}(\hat{\beta})},$$

Here, $\hat{\beta}$ denotes the least squares estimate of β, and $\text{std}(\hat{\beta})$ its standard error. The test statistic t_β is compared against the critical values from the Dickey-Fuller distribution. If t_β is sufficiently negative, the null hypothesis H_0 is rejected, implying that the series is stationary. The detailed results are reported in the Appendix 8.

4.3.2 ARCH effects

Before specifying a volatility model, we first check for the presence of time-varying variance. After fitting an ARMA(p, q) model to the mean equation μ_t via maximum likelihood estimation, we compute the residuals:

$$a_t = y_t - \mu_t,$$

which, in practice, reduces to $a_t = y_t$, since we assume $\mu_t = 0$.

Even though the series $\{a_t\}$ may exhibit characteristics of white noise, its squared series $\{a_t^2\}$ often shows significant serial correlation. This suggests that while the first moment (mean) may be independent, there is dependence in the second moment (variance).

To detect this phenomenon, we apply the Ljung-Box test to the squared residuals $\{a_t^2\}$ (12). For results of the ARCH effects see Appendix, Table 9.

4.3.3 Assessing the Orders of the GARCH model

We proceed by fitting GARCH(p, q) models to each stocks daily log return series separately using the maximum likelihood estimation, under the assumption that residuals follow a Student-t distribution. In Table 8 see Appendix A, reports the AIC and BIC values for models of order (p, q) ranging from (1,1) to (3,3).

The results indicate that, for most stocks, the GARCH(1,1) model provides the lowest AIC and BIC values, suggesting it offers the best trade-off between model fit and complexity. This finding is consistent with the general results discussed in Section 2.1.2.

5 Empirical results

In order to test the DCC-model, an application to real data was performed. The chosen data set consists of the stocks Svenska Handelsbanken and Atlas Copco:

$$Y_{t,i} = (\text{SHB A, ATCO B})$$

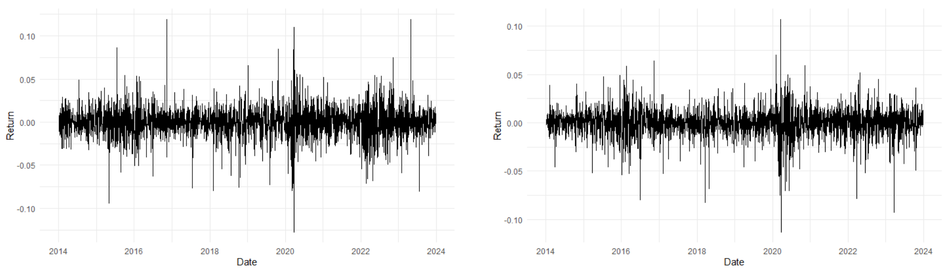
The dataset covers daily trading data from 2014-01-01 to 2024-01-01, corresponding to a time span of $t = 1, \ldots, 2513$, where $t = 1$ represents 2014-01-01 and $t = 2513$ represents 2024-01-01. Daily data were chosen instead of weekly data, as the higher frequency provides more observations and allows for a more accurate representation of volatility dynamics. The log returns are defined as:

$$r_{t,i} = \log Y_{t,i} - \log Y_{t-1,i}$$

As stated in 16, $r_t = y_t + \mu_t$ where r_t is the log returns, y_t are the mean corrected returns and μ_t is the mean vector. Here $E[y_t] = 0, Cov[y_t] = H_t$.

5.1 Data

Looking at the mean-corrected returns in Figure 1, it is obvious that each series has volatility clustering. In other words, we see that large changes tends to be followed by large changes and small changes are followed by small changes. Given the heteroscedastic properties of the series it indicates that GARCH-models are appropriate to use for this data set.



(a) Mean corrected-returns for ATCO B (b) Mean corrected-returns for SHB A

Figure 1: Mean-corrected returns for ATCO B and SHB A

5.2 Autocorrelation function

By examining the autocorrelation function of y_t^2 in Figure 3 one can see that more than 5% of the lags fall outside the confidence limits of 95%, and there is a clear pattern in which the autocorrelations are large at the beginning but decay over time. This suggests that y_t^2 is not serially uncorrelated, and consequently y_t exhibits dependence.

On the other hand, when inspecting the ACF of y_t in Figure 2, we see that fewer than 5% of the lags fall outside the confidence interval, indicating that y_t is serially uncorrelated.

Based on these observations, we conclude that a GARCH process is appropriate for modelling the data, as y_t is uncorrelated but dependent. This property holds across all stock pairs in the dataset.

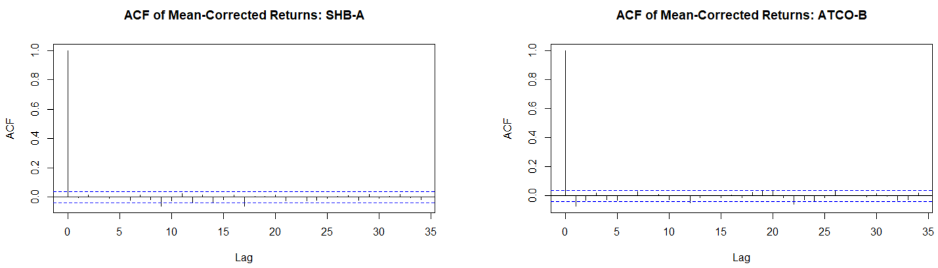


Figure 2: ACF plots of mean-corrected returns(y_t) for SHB A and ATCO B

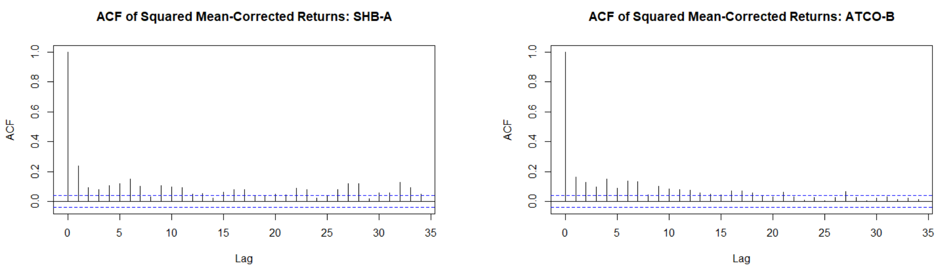


Figure 3: ACF plots of squared mean-corrected returns(y_t^2) for SHB A and ATCO B

5.3 Modelling and estimations

In this section the estimated parameters of the DCC-model will be presented for the stock pair $Y_{t,i}$ =(SHB A, ATCO B). For the underlying univariate models, they are all p = q = 1.

5.3.1 First stage parameters

The first stage estimation parameters, estimated from $\max\{L_v(\theta)\}$ which is explained in Section 3.1, see Equation 40, are the same for both the Gaussian distribution and the multivariate Student's t-distribution, since both approaches assume normally distributed univariate GARCH models. The estimated parameters from the first stage are:

Estimated GARCH(1,1)			
Stock	ω	α_1	β_1
SHB	2.257861e-07	5.074903e-02	9.002305e-01
ATCO	3.320269e-07	5.170496e-02	9.006502e-01

Table 1: Estimated GARCH(1,1) for SHB A and ATCO B

Both series show high persistence in Table 1, $\alpha_1 \approx 0.051$, $\beta_1 \approx 0.90$, and $\alpha_1 + \beta_1 \approx 0.95 < 1$, which is consistent with stationarity. The implied unconditional daily volatility is about 0.21% for SHB and 0.26% for ATCO, so ATCO is a bit more volatile. Parameter similarity indicates very similar volatility dynamics across the two stocks.

5.3.2 Second stage parameters

As explained in Section 3, the second stage estimation differs between the Gaussian distribution and the multivariate Student's t-distribution. For the Gaussian case, $\max\{L_v(\hat{\theta},\phi)\}$ was estimated, see Equation 40, while for the multivariate Student's t-distribution, the estimation was based on $\max\hat{\phi} = \max\{L_v(\hat{\theta},\phi,\nu)\}$ outlined in Section 3.2, see Equation 44. The second stage parameters are:

Estimated DCC(1,1) Parameters for Different Distributions			
Distribution	α_1	β_1	ν
Gaussian	0.052949	0.808967	NA
MV Student-t	0.036821	0.853241	4.9996

Table 2: Second-stage DCC parameter estimates for Gaussian and Multivariate Student's t distributions

Both DCC estimates in Table 2 are persistent but mean reverting, with $\alpha_1 + \beta_1 \approx 0.86$ for Gaussian and ≈ 0.89 for Student t, both below one. Relative to Gaussian, the Student t fit shifts weight from α_1 to β_1, implying smoother and slightly more persistent correlation dynamics. The estimated degrees of freedom $\nu \approx 5$ indicates a more pronounced tail.

5.4 DCC-forecasts

The k step ahead forecast of the conditional covariance matrix is $H_{t+k} = D_{t+k}\, R_{t+k}\, D_{t+k}$,where D_{t+k} is diagonal with the univariate volatility forecasts (conditional standard deviations) on the diagonal, and R_{t+k} is the conditional correlation matrix. This decomposition separates marginal volatility from covariation and allows examination of time varying dependence between assets.

Applied to Svenska Handelsbanken (SHB A) and Atlas Copco (ATCO B), this framework assesses whether two firms from distinct sectors, banking and industrials, are effectively uncorrelated. The estimates reported below indicate a meaningful correlation between their returns.

In this thesis a $k = 350$ step ahead forecast will be performed where k is one trade day.

5.4.1 D_t-forecasts

The D_{t+k} forecasts were obtained from the first stage of the quasi-maximum likelihood estimation. As discussed earlier in Section 3.1, the univariate normal forecasts serve as the basis for the second-stage estimations under both the Gaussian and the multivariate Student's t distributions. Note that in the Figure 4, the black line corresponds to D_t, the red line illustrates D_{t+k} and the cyan line denotes the unconditional variance.

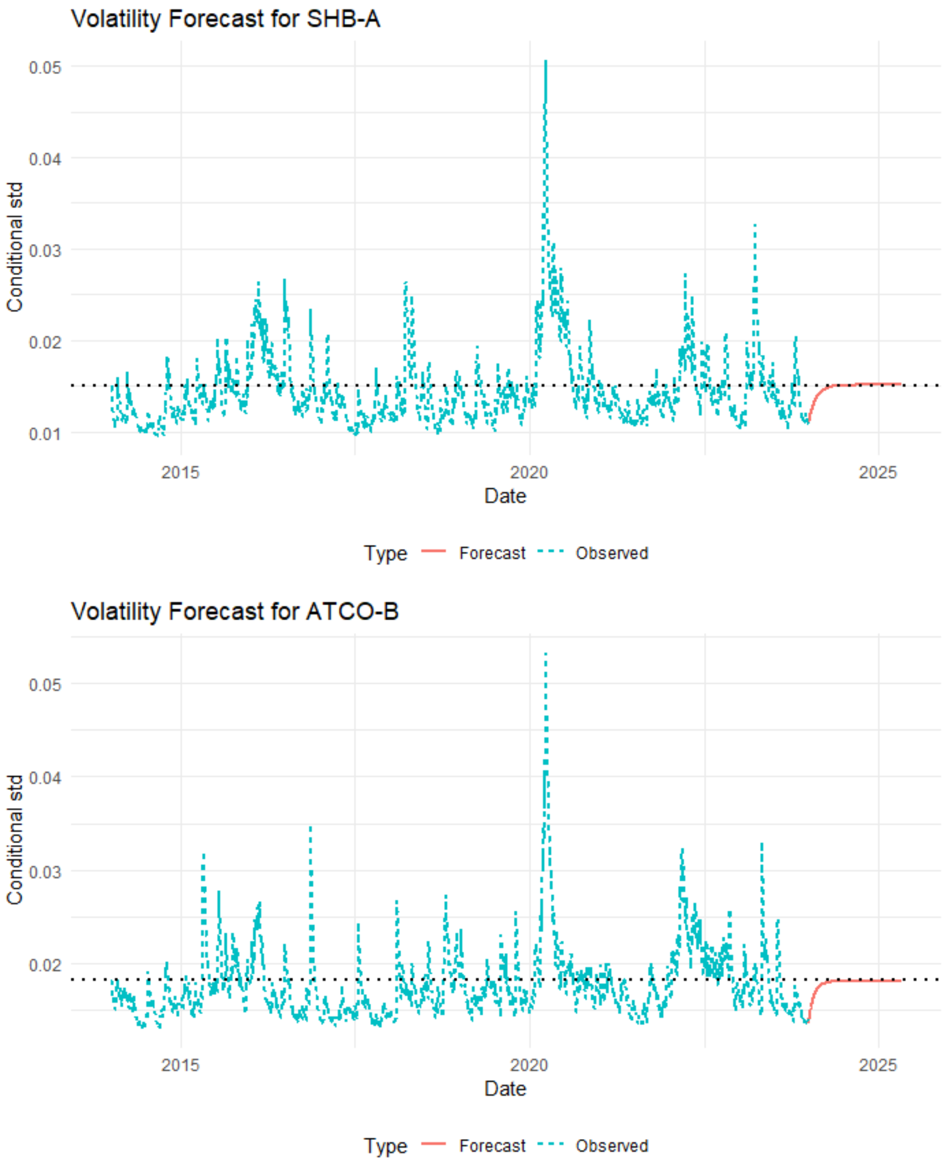


Figure 4: Plots of D_{t+k}.

Both stocks show pronounced volatility spikes (notably around 2020) but mean-revert toward their pre-shock baseline. Short-horizon forecasts from Gaussian and Student t specifications are very similar, with the Student t paths a bit smoother. ATCO B remains consistently more volatile than SHB A.

5.4.2 R_t-forecasts

The R_{t+k} forecasts were derived from the second-stage quasi-maximum likelihood estimations, as outlined in Section 3. In the Figures 5 and 6, the black line denotes R_t, the red line represents R_{t+k} and the cyan line corresponds to the unconditional correlation.

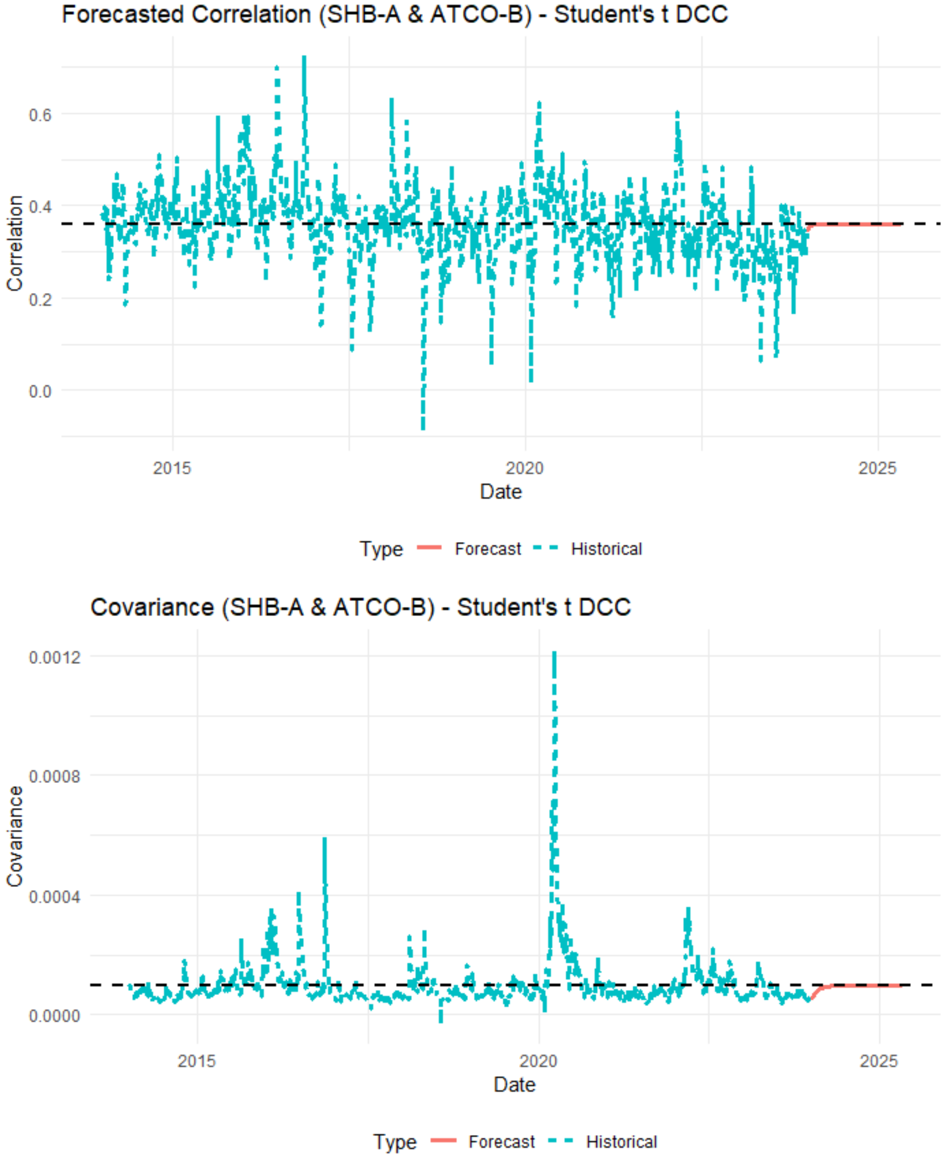


Figure 5: Plots of R_{t+k} under Multivariate Student-t distribution.

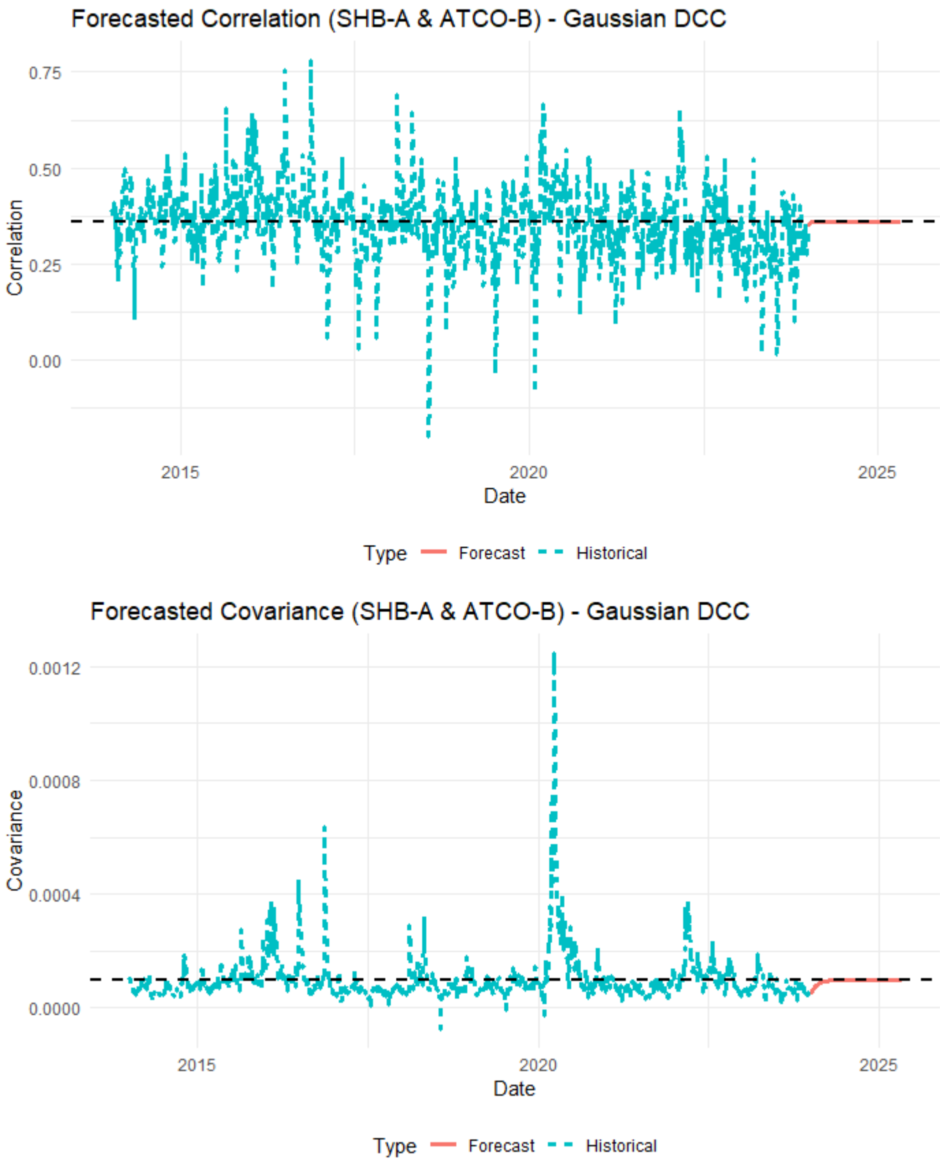


Figure 6: Plots of R_{t+k} under Multivariate Gaussian distribution.

The DCC estimates show that correlation between Svenska Handelsbanken and Atlas Copco is clearly time varying and often climbs above fifty percent, despite the firms operating in different sectors. The series tends to hover around forty percent on average, rising during market stress and drifting toward thirty percent in calmer periods. Short horizon forecasts settle near the mid thirties to about forty percent, with the Student t specification a little lower than the Gaussian one, but the overall picture is the same.

This level of covariation is plausible given shared exposure to broad market factors within OMX30, macro shocks such as interest rates, the SEK exchange rate, and global demand, links between credit conditions and industrial investment, and an overlapping investor base. Both specifications pass basic checks, so the finding is not model driven. For portfolio construction the diversification benefit from pairing these names is limited and depends on the state of the market.

5.4.3 H_t-forecasts

Since H_{t+k} combines the previously presented plots of D_{t+k} and R_{t+k}, it reflects both components jointly. In Figure 7, the black line denotes H_t, the red line represents H_{t+k} and the cyan line corresponds to the unconditional covariance.

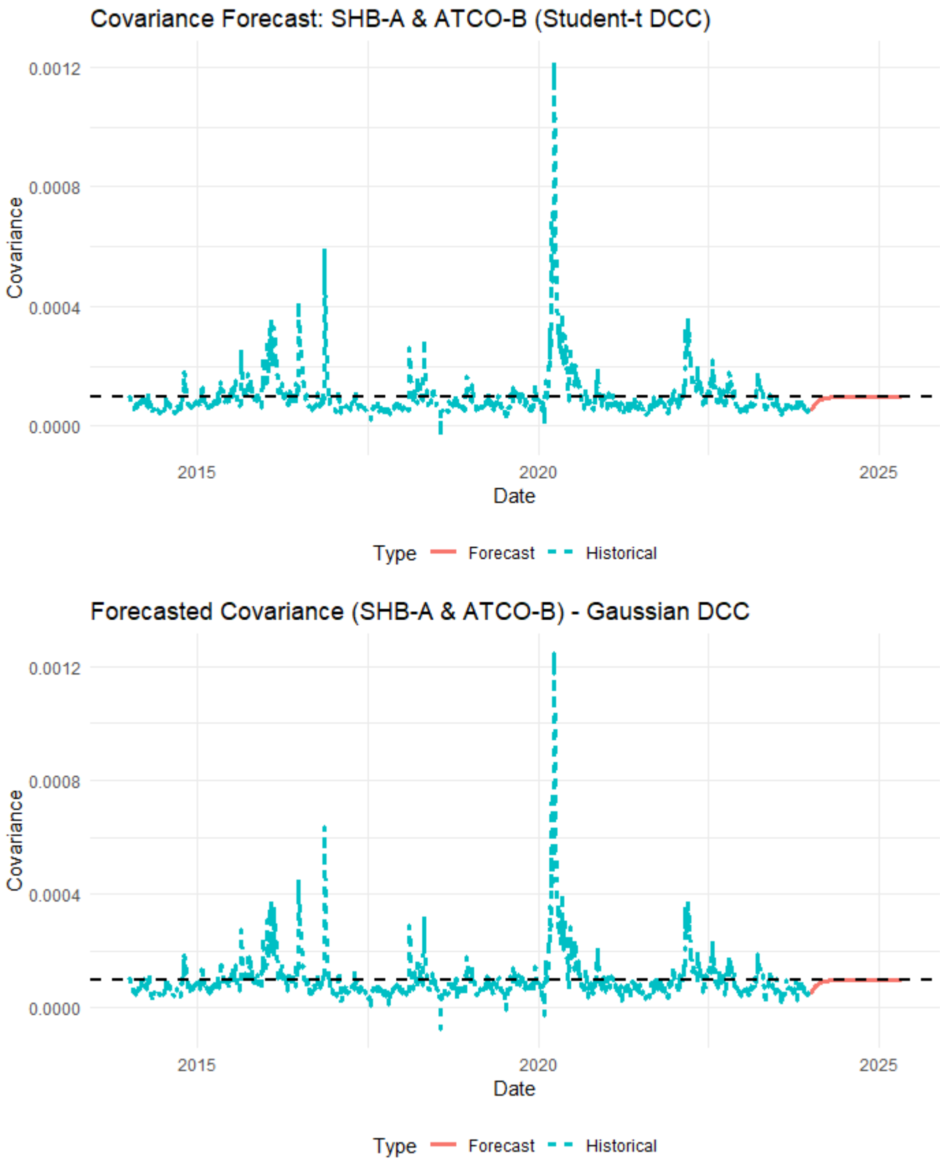


Figure 7: Plots of H_{t+k} under Multivariate Gaussian distribution and Multivariate Student-t distribution.

The covariance is positive most of the time, with sharp spikes during stress periods around 2020 and brief dips toward zero. Both models short horizon forecasts revert to a modest positive level, the Student-t path is a bit smoother and slightly lower than the Gaussian. Overall this points to a persistent but state dependent.

6 Goodnes of fit

6.1 Marginal distributions

In this section of the thesis the errors will be assessed and we will also look at the goodness of fit for each distribution with help of figures and statistical tests.

6.1.1 Visual error evaluation

The errors are calculated from the conditional covariance matrix. The errors are calculated from $\varepsilon_t = H_t^{-\frac{1}{2}} y_t$. It is not possible to draw any conclusions by looking at the error plots in Figures 8 and 9 but we can see that the tails of the errors for the gaussian distribution are slightly heavier than the multivariate student-t distribution in Figure 10.

Figure 8: Errors for SHB A under the Gaussian and Multivariate Student-t DCC-GARCH(1,1) model.

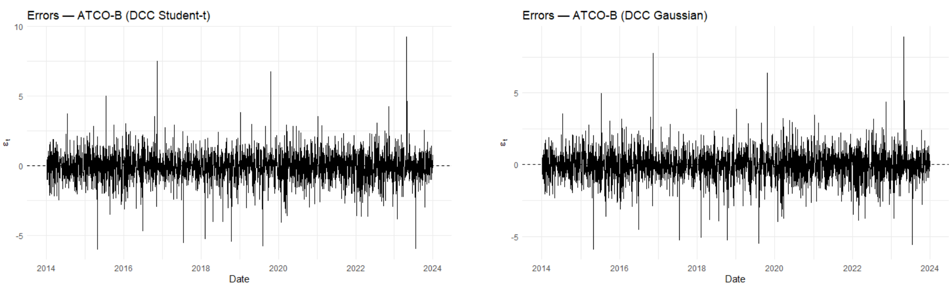
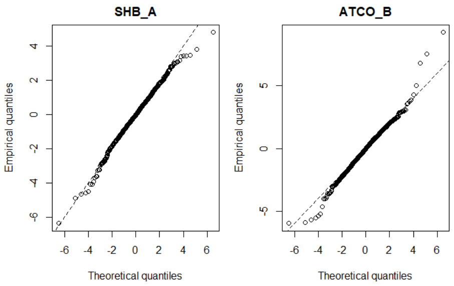
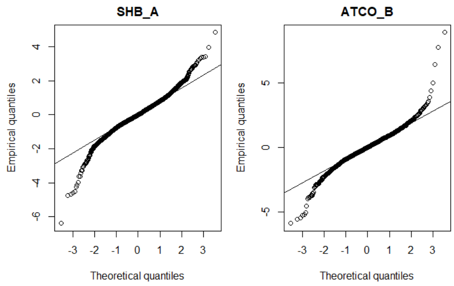


Figure 9: Errors for ATCO B under the Gaussian and Multivariate Student-t DCC-GARCH(1,1) model.



(a) Empirical quantiles of the errors from the DCC-GARCH(1,1) model with a student-t distribution.

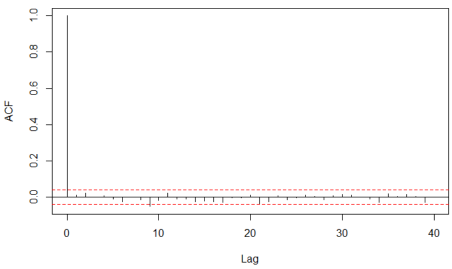


(b) Empirical quantiles of the errors from the DCC-GARCH(1,1) model with a Gaussian distribution.

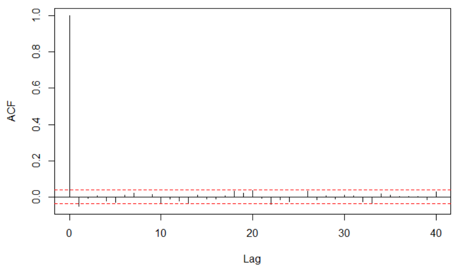
Figure 10: QQ-Plots for ε_t

6.1.2 The autocorrelation function

Displayed below in Figures 11 and 12 are the autocorrelation functions of the different error series ε_t. At the 5% significance level, approximately 5% of the lags are expected to lie outside the confidence bounds (the red lines). In practice, fewer than 5% of the lags do so, leading to the conclusion that the errors for both distributions behave randomly at the 5% level across all series. These results can also be verified in Table 3.

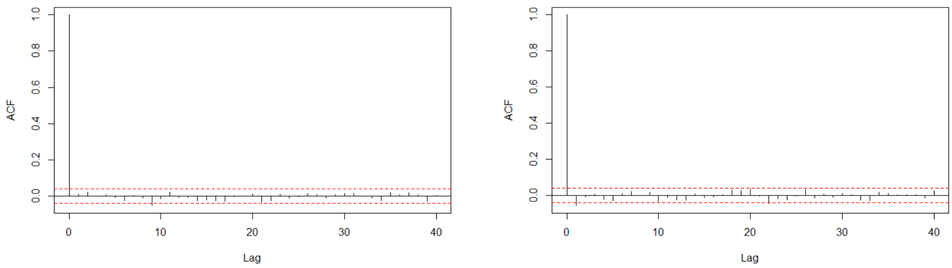


(a) SHB A



(b) ATCO B

Figure 11: ACF for multivariate Student's-t errors.



(a) SHB A

(b) ATCO B

Figure 12: ACF for Gaussian errors.

Statistic	**Series and model**	**Lags outside 95% CI**	**Percent outside**
ACF of ε_t	SHB A, DCC t	1/40	2.5%
ACF of ε_t	ATCO B, DCC t	2/40	5.0%
ACF of ε_t	SHB A, DCC Gaus.	1/40	2.5%
ACF of ε_t	ATCO B, DCC Gaus.	2/40	5.0%

Table 3: ACF diagnostic summary for residuals of the DCC-GARCH(1,1) Multivariate student-t and Gussian model.

6.1.3 Ljung-Box test

Figures 13 and 14 below shows the Ljung-Box test statistic as a function of the lag, $1, \ldots, 30$. The results indicate that, across all series and for both distributions, some lags exhibit correlation while others do not. Consequently, no definitive conclusion can be drawn from this test. The red line denotes the 5% significance level of the test statistic.

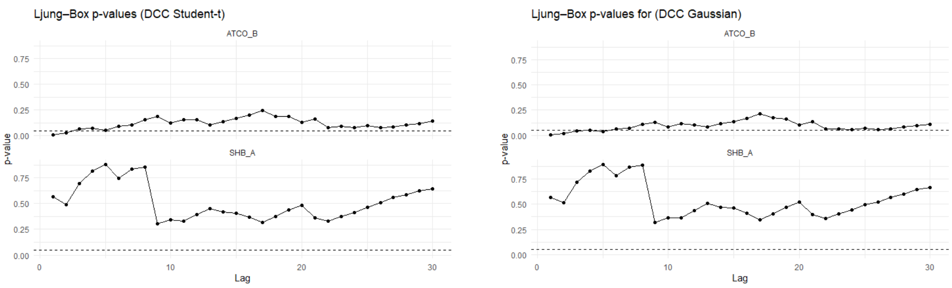


Figure 13: Ljung Box test, Gaussian and student-t errors.

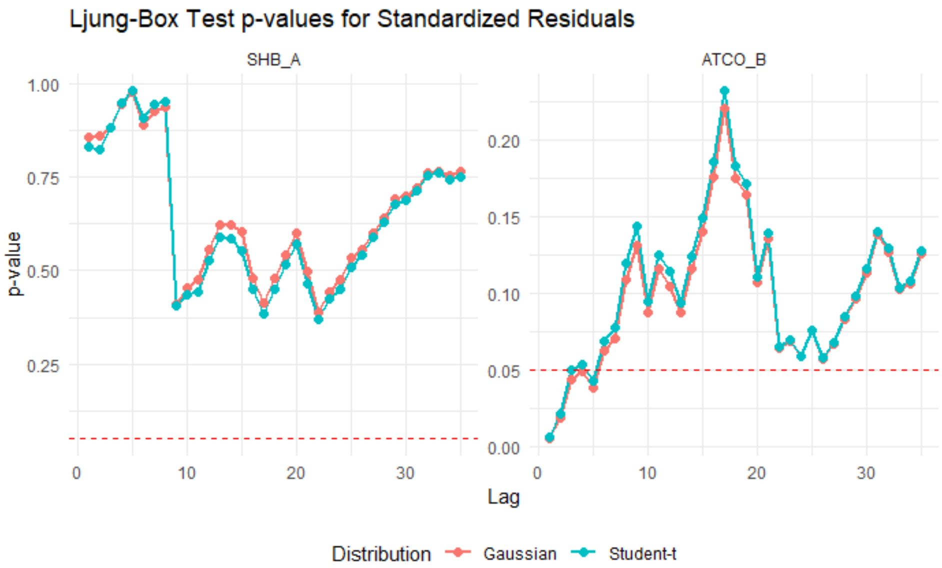


Figure 14: Ljung-box test for SHB A and ATCO B.

6.2 Multivariate goodness of fit

This section evaluates Value at Risk (VaR) backtesting with the Kupiec test and the Christoffersen test. All tests are conducted in a rolling window setting. The fitting set contains the data used to estimate the model, while the forecast period is the subsequent out of sample horizon generated from the fitted model and compared with the realized data. In Table 4 we can observe the estimated parameters when performing the VaR back-test.

Distribution	α_1	β_1	ν
Gaussian DCC	0.2154	0.5948	NA
Student's t DCC	0.1284	0.7381	4.985

Table 4: Estimated parameters α_1, β_1, and degrees of freedom under Gaussian and Student's t DCC models.

6.2.1 VaR Violations

In the section we apply two tests to the VaR violation sequences, the Kupiec-test and the conditional coverage test (or Christoffersen test). The Kupiec-test evaluates whether the total number of violations matches the expected rate, while the conditional coverage test jointly tests both the violation frequency and the independence of violations over time. The p-values for each test are presented in Table 5 and 6.

Modell	**Violation rate**	**Kupiec p-value**	**CC p-value**
DCC(gaussian), $\alpha = 10\%$	0.094	0.590	0.827
DCC(gaussian), $\alpha = 5\%$	0.050	0.960	0.997
DCC(gaussian), $\alpha = 1\%$	0.017	0.070	0.154

Table 5: Violation rate, Kupiec and Christoffersen p-values for DCC-GARCH(1,1) Gaussian.

Model	**Violation rate**	**Kupiec p-value**	**CC p-value**
DCC(Student-t), $\alpha = 10\%$	0.1088	0.4288	0.7310
DCC(Student-t), $\alpha = 5\%$	0.0597	0.2359	0.4857
DCC(Student-t), $\alpha = 1\%$	0.0040	0.0586	0.1651

Table 6: Violation rate, Kupiec and Christoffersen p-values for DCC-GARCH(1,1) Multivariate student-t.

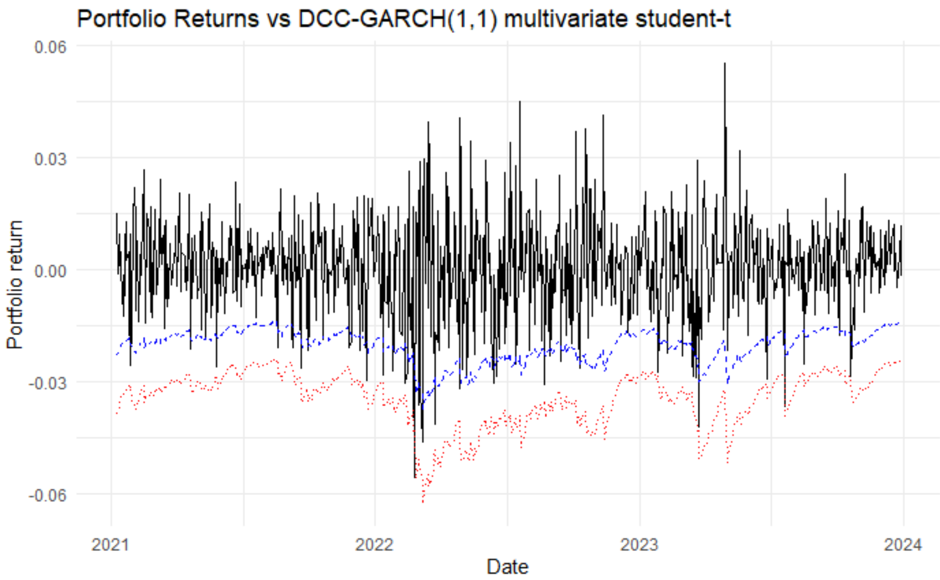


Figure 15: DCC-GARCH(1,1) multivariate Student-t VaR.

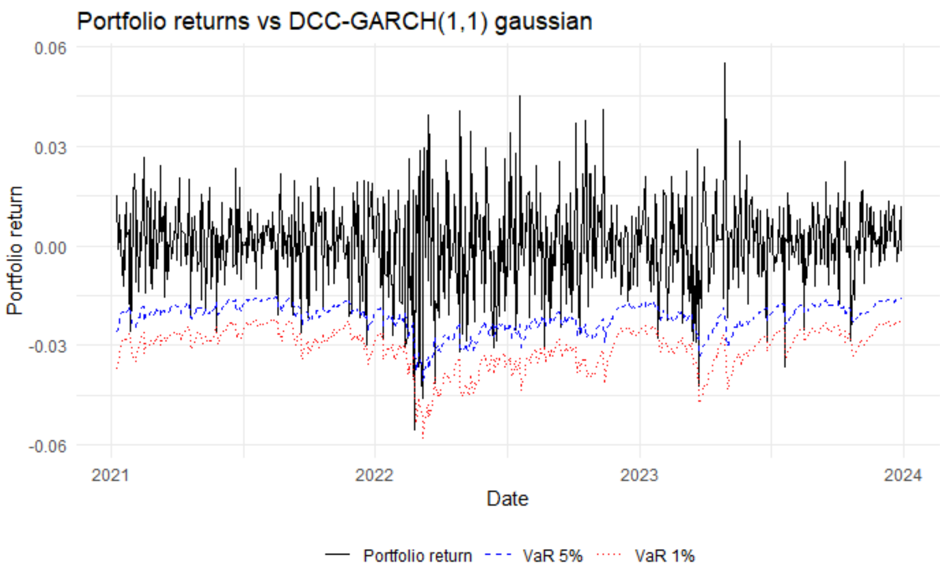


Figure 16: DCC-GARCH(1,1) Gaussian VaR.

We evaluate Value at Risk violation sequences with the Kupiec test and the Christoffersen conditional coverage test. For both DCC GARCH specifications, the 10% and 5% VaR levels show violation rates close to target, with large p values and no evidence of dependence in violations. At the 1% level, the Gaussian model produces a rate of 1.7% and the

Student-t model produces a rate of 0.4%. In both cases the Kupiec p values are near but above the 5% threshold and the Christoffersen p values do not indicate dependence.

At conventional 5% significance, both models pass coverage and independence checks across the reported levels. The Gaussian specification aligns most closely with target coverage at 10% and 5%, while the Student-t specification is more conservative in the far tail. For applications that prioritize precision around common risk thresholds, the Gaussian model is adequate, for stricter tail protection, the Student t model offers a conservative alternative. Overall, the DCC-GARCH(1,1) framework delivers VaR forecasts with acceptable coverage and no detected clustering of violations. To visulies the Value at Risk test and see the rather similar performances see Figures 15 and 16.

7 Discussion and Key findings

The objective of this thesis was to examine the application of the DCC-GARCH(1,1) model for forecasting conditional covariances and correlations, with a particular focus on comparing Gaussian and multivariate Student's t distributed innovations. The analysis combined both theoretical considerations and empirical evaluation based on daily stock returns from SHB A and ATCO B.

Several simplifying assumptions were made to narrow the scope of the study. For instance, the mean vector μ_t in the univariate GARCH(1,1) model was treated as constant rather than being modeled through an ARMA process, and only one of the forecasting approaches for R_t proposed by Engle and Sheppard was implemented. Moreover, the lag orders of the univariate and multivariate GARCH models were restricted to $(1, 1)$ but since we accessed the order of the univariate GARCH model according to the Table 8 it was not necessary to model with other lags since the AIC and BIC values suggested that the GARCH(1,1) model was overall the best fit.

The empirical results showed that both Gaussian and Student's t specifications displayed high persistence in conditional correlations, as reflected in the estimated α_1 and β_1 parameters. For the Student's t model, the degrees of freedom ν was not that large, suggesting thinner tails and behavior closer to the Gaussian distribution.

The backtesting of Value at Risk (VaR) was carried out using both the Kupiec test and the Christoffersen conditional coverage test. The Kupiec test evaluates whether the observed number of violations matches the expected level, while the Christoffersen test extends this by also checking for independence of violations over time. The results are summarized in Tables 5 and 6.

For both the Gaussian and the Student-t DCC-GARCH(1,1) models, the violation rates at the 10% and 5% levels are close to their theoretical targets, and the corresponding p-values are high, suggesting no evidence of misspecification. At the 1% level, the Gaussian model slightly overestimates risk (1.7% violation rate), while the Student-t model slightly underestimates it (0.4% violation rate). In both cases, the Kupiec p-values are borderline but remain above the 5% threshold, and the Christoffersen test does not indicate serial dependence.

Taken together, the results suggest that both models provide adequate VaR forecasts at conventional significance levels. The Gaussian specification tracks the nominal levels more closely at 10% and 5%, while the Student-t specification is more conservative in the far tail. This implies that the Gaussian model is sufficient when accuracy around common risk thresholds is most important, whereas the Student-t model may be preferred when stricter tail protection is desired. Overall, the DCC-GARCH(1,1) framework produces VaR forecasts with satisfactory coverage and no clustering of violations. The similar performance of the two models is also illustrated in Figures 15 and 16.

Overall, the comparison indicates that the Gaussian and Student's t DCC-GARCH models perform more similarly than might be expected. Although the Student's t distribution is often considered superior due to its ability to capture heavy tails in financial returns, the relatively high estimates of ν reduced this advantage in practice. Consequently, the Student's t specification can be regarded as performing marginally better in some cases, but the overall findings suggest that both distributions provide comparable performance in multivariate risk modelling.

8 Appendix A

Aktie	p-value
SHB	0.01
ATCO	0.01

Table 7: ADF-testresults for SHB and $ATCO$

(a) SHB

Model	AIC	BIC
GARCH(1,1)	-5.868558	-5.852181
GARCH(1,3)	-5.867319	-5.842754
GARCH(1,2)	-5.867174	-5.846703
GARCH(2,3)	-5.867124	-5.838465
GARCH(2,1)	-5.866985	-5.846514
GARCH(2,2)	-5.865582	-5.841017
GARCH(3,3)	-5.865529	-5.832776
GARCH(3,1)	-5.865464	-5.840899
GARCH(3,2)	-5.864066	-5.835406

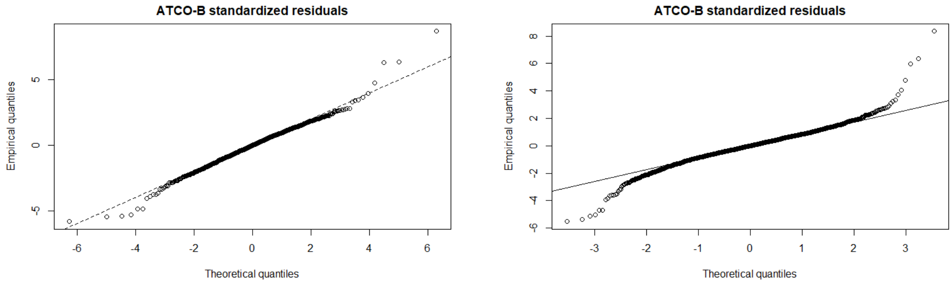
(b) $ATCO$

Model	AIC	BIC
GARCH(1,1)	-5.262768	-5.246392
GARCH(2,1)	-5.261135	-5.240664
GARCH(1,2)	-5.261130	-5.240659
GARCH(2,2)	-5.259612	-5.235047
GARCH(1,3)	-5.259535	-5.234970
GARCH(3,1)	-5.259526	-5.234961
GARCH(2,3)	-5.258268	-5.229608
GARCH(3,2)	-5.257976	-5.229316
GARCH(3,3)	-5.256673	-5.223919

Table 8: AIC and BIC values for GARCH(p, q) models for each stock

Stock	**ARCH Test p-value**
SHB	0.00000
ATCO	0.00000

Table 9: ARCH test p-values for SHB and $ATCO$.



(a) Student-t distribution (b) Gaussian distribution

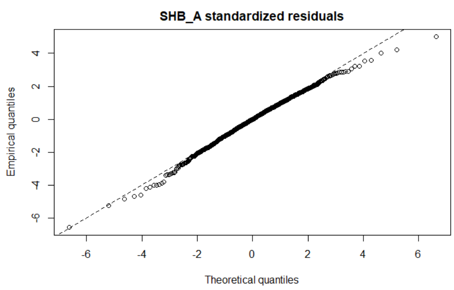
Figure 17: Residuals for ATCO B with the GARCH(1,1) modell

Test	Distribution	VaR 1%	VaR 5%	VaR 10%
1	Gaussian DCC	Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0
1	Student's t DCC	Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0
2	Gaussian DCC	Fail to Reject H_0	Fail to Reject H_0	**Reject** H_0
2	Student's t DCC	Fail to Reject H_0	Fail to Reject H_0	**Reject** H_0

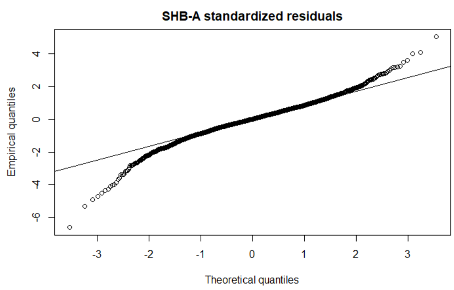
Table 10: Kupiec test decisions for VaR at 1%, 5%, and 10% levels (significance: 5%).

Test	Distribution	Unconditional (UC)			Conditional (CC)		
		1%	5%	10%	1%	5%	10%
1	Gaussian	0.075	0.134	0.372	0.180	0.216	0.092
1	Student-t	0.214	0.134	0.833	0.426	0.216	0.401
2	Gaussian	0.428	0.071	**0.000**	0.727	**0.050**	**0.000**
2	Student-t	0.428	0.071	**0.000**	0.727	**0.050**	**0.000**

Table 11: Christoffersen test p-values for both Gaussian and Student-t DCC models under Test 1 and Test 2. Bold values indicate rejection of H_0 at the 5% significance level.



(a) Student-t distribution



(b) Gaussian distribution

Figure 18: Residuals for SHB-A with the GARCH(1,1) modell

References

[1] Robert F. Engle. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." In: *Econometrica* 50.4 (July 1982), pp. 987–1007. ISSN: 0012-9682. DOI: 10.2307/1912773.

[2] Daniel B. Nelson and Charles Q. Cao. "Inequality Constraints in the Univariate GARCH Model." In: *Journal of Business and Economic Statistics* 10.2 (Apr. 1992), pp. 229–235. ISSN: 0735-0015. DOI: 10.1080/07350015.1992.10509902.

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

[4] Luc Bauwens, Christian M. Hafner. "Identification of structural multivariate GARCH models." In: *Journal of Econometrics* 219.1 (2020), pp. 21–42. ISSN: 0304-4076. DOI: `10.1016/j.jeconom.2020.05.003`. `https://www.sciencedirect.com/science/article/pii/S0304407620302098`

[5] Annastiina Silvennoinen and Timo Terasvirta. "Multivariate GARCH models." In: *Handbook of Financial Time Series*. Forthcoming. Also published as SSE/EFI Working Paper Series in Economics and Finance No. 669, Stockholm School of Economics (Jan. 2008). `https://swopec.hhs.se/hastef/papers/hastef0669.pdf`

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

[7] Robert F. Engle and Kenneth F. Kroner. "Multivariate Simultaneous Generalized ARCH." In: *Econometric Theory* 11.1 (Mar. 1995), pp. 122–150. Published by: Cambridge University Press. Stable URL: `http://www.jstor.org/stable/3532933`. Accessed: 18-08-2017.

[8] Robert F. Engle, Victor K. Ng, and Michael Rothschild. "Asset Pricing with a Factor-ARCH Covariance Structure: Empirical Estimates for Treasury Bills." In: *Journal of Econometrics* 45.1-2 (July 1990), pp. 213–237. ISSN: 0304-4076. DOI: 10.1016/0304-4076(90)90099-F.

[9] Julien Chevallier. “Time-Varying Correlations in Oil, Gas and CO_2 Prices: An Application Using BEKK, CCC and DCC-MGARCH Models.” In: *Applied Economics* 43.30 (2011), pp. 3965–3981. ISSN: 1466-4283. DOI: 10.1080/00036846.2011.589809.

[10] Roy van der Weide. “GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model.” In: *Journal of Applied Econometrics* 17.5 (Sept. 2002), pp. 549–564. ISSN: 0883-7252. DOI: 10.1002/jae.688.

[11] Tim Bollerslev. “Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model.” In: *The Review of Economics and Statistics* 72.3 (Aug. 1990), pp. 498–505. ISSN: 0034-6535. DOI: 10.2307/2109358.

[12] Ruey S. Tsay. *Analysis of Financial Time Series*. 2nd ed., Wiley-Interscience, 2005.

[13] Robert F. Engle and Kevin Sheppard. “Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH”. In: *NBER Working Paper Series* (2001).

[14] P. J. Brockwell and R. A. Davis. Time Series: Theory and Methods. Springer, 1991

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

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

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

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