Financial Modelling with Copulas

Model Selection for Risk Management

Rasmus Nørrevang
Sebastian Søndergaard Kure

Master's Thesis, Group 4.106 B1, Mathematics-Economics

Dept. of Mathematical Sciences
Skjernvej 4A
DK-9220 Aalborg Ø
http://math.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title
Financial Modelling with Copulas

Theme
Model Selection for Risk Management

Project Period
Spring Semester 2023

Project Group
Group 4.106 B1

Participant(s)
Rasmus Nørrevang
Sebastian Søndergaard Kure

Supervisor(s)
J. Eduardo Vera-Valdés
Toke C. Zinn

Page Numbers
66

Date of Completion
June 1, 2023

Abstract

Copula model selection for risk management is the main topic of this master's thesis. In particular, we seek to investigate how copulas can be used for risk forecasting, and how one can evaluate the accuracy of the obtained forecasts and identify the most suitable model(s).

To answer these questions, we dive into the theory of copulas, including the probability integral transforms, the crucial Sklar's theorem, and examples of copulas. It also includes the Bayesian estimation method, a goodness-of-fit test, and the marginal distribution modelling choice.

Afterward, we present the risk measures of interest, the procedure on how to forecast them, and the tests for evaluating their accuracy. Also, we provide an introduction to the model confidence set in order to determine the most suitable model(s).

In the applications part, we consider an equally weighted portfolio composed of the stocks of NextEra Energy and British Petroleum. Using these, we do exploratory data analysis and consider marginal and copula model diagnostics. Lastly, we do risk forecasting of the portfolio, evaluate the forecasts by the introduced tests, and identify the superior model(s) by the model confidence set.

Preface

This master's thesis is written during the spring semester of 2023 as part of the master's programme in Mathematics-Economics at the Department of Mathematical Sciences at Aalborg University.

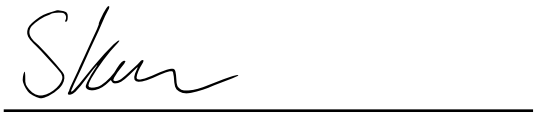
The thesis is composed of two parts including numbered chapters with corresponding sections and subsections. Citations and external references are made using the Vancouver system such that [1] refers to the first entry in the bibliography. All programming is performed using `Julia` and `R`, and the code can be found at the following GitHub repository:

https://github.com/skure18/MastersThesisCode

Finally, the group would like to extend our appreciation to our supervisors J. Eduardo Vera-Valdés and Toke C. Zinn.

Signatures

Rasmus Nørrevang



Sebastian Søndergaard Kure

Contents

Preface iii

Introduction vii

I Theory 1

1 Copulas 3
- 1.1 General Framework . . . 3
- 1.2 Main Definitions and Results . . . 3
- 1.3 Examples of Copulas . . . 6
  - 1.3.1 Special Copulas . . . 6
  - 1.3.2 Parametric Copulas . . . 7
  - 1.3.3 Archimedean Copulas . . . 8
- 1.4 Tail Dependence . . . 10
- 1.5 Copula Estimation . . . 12
- 1.6 Goodness-of-Fit Test . . . 13
- 1.7 Marginal Distribution Modelling . . . 14
  - 1.7.1 Distribution Specification . . . 16

2 Risk Forecasting 17
- 2.1 Risk Measures . . . 17
- 2.2 Risk Forecasting Procedure . . . 17
- 2.3 Risk Forecasting Evaluation . . . 19
  - 2.3.1 Accuracy of VaR and ES Forecasts . . . 19
  - 2.3.2 Model Selection . . . 21

II Application 25

3 Data Presentation 27

4 In-Sample Model Diagnostics 33
- 4.1 Introduction to the Main Packages Used in `Julia` . . . 33
- 4.2 Marginal Modelling . . . 34
- 4.3 Copula Modelling . . . 37

5 Out-of-Sample Risk Forecasting 43

6 Conclusion 49

7 Future Considerations 51

Appendices 53

A The Fréchet-Hoeffding Bounds 55

B Miscellaneous Tables and Figures 57

Bibliography 66

Introduction

In recent times, the COVID-19 pandemic and the Russian invasion of Ukraine have resulted in clear increases in volatility and overall insecurities in the energy markets [12]. As a consequence, effective risk management has become of paramount importance for investors and financial institutions. To perform effective risk management, accurate modelling of dependencies between financial variables plays a critical role. Historically, the linear correlation approach has been used to determine dependence structures. However, studies have demonstrated that copula models, which provide a flexible and powerful framework for capturing the dependence structure between random variables, are a considerably more useful tool in financial risk management [13]. In particular, the correlation approach fails to capture dependence in extreme events, whereas many copula models do not suffer from this limitation. Hence, copulas are ideal candidates for modelling the joint behavior of financial assets, as we have seen the tendency of simultaneous extreme events during the COVID-19 pandemic and the invasion of Ukraine. However, as of writing this thesis, there is no general consensus on choosing the optimal copula model(s) in applications regarding risk management.

In this thesis, we propose an approach to evaluate the accuracy of copulas' risk forecasts and a method for identifying the most suitable copula model(s). To this end, we investigate the joint behavior of two stocks within the energy and oil/gas markets, namely NextEra Energy and British Petroleum, respectively. We elect to investigate these particular stocks as the aforementioned recent crises have shown large increases in the volatility of electricity, oil, and gas prices, which have been reflected in the prices of these stocks. In particular, we consider an equally weighted portfolio composed of the two stocks. We then seek to forecast value-at-risk and expected shortfall for this portfolio, and evaluate the utilized copula models' forecasting accuracy. After evaluating the accuracy of the copula models' risk forecasts, the copula model(s) with the most precise risk forecasts are determined. To this end, the so-called model confidence set is employed, which systematically eliminates the statistically inferior models, such that the remaining copula model(s) have no significant difference in risk forecasting performance.

This master's thesis aims to contribute to the field of risk management by systematically evaluating and selecting the most suitable copula model(s). The research findings will offer practical implications for risk managers seeking accurate frameworks to measure risk in today's complex financial landscape.

Problem Statement

Based on the introduction above, the problem statement for the thesis is as follows:

How can copulas be employed for risk forecasting in stock portfolios, and how can one evaluate the accuracy of these forecasts and identify the most suitable copula model(s) for precise risk assessment?

Part I
Theory

1 Copulas

Unless explicitly stated otherwise this chapter is based on [2], [10], [16], [19] and [20].

Copulas are a powerful tool in statistics for modeling the dependence structure between variables. Unlike traditional methods, copulas enable us to model the dependence structure separately from the marginal distributions of the individual variables. In finance, copulas are particularly useful for analyzing the dependencies between financial assets, such as stocks or bonds, which is crucial for risk management. This chapter aims to provide an overview of the necessary copula theory utilized in this thesis.

1.1 General Framework

Throughout this thesis, an arbitrary probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is fixed such that any random variable X and stochastic process $Y = (Y_t)_{t\in\mathbb{N}_0}$ is defined on this probability space with state space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. A distribution function F of a random variable X is described as a monotonically increasing càdlàg function mapping the set of real numbers to the interval $[0, 1]$. Furthermore, lowercase letters will be used to denote some realization of a random variable X or stochastic process Y, that is, a realization of X will be denoted by x and a realization of Y will be denoted by $y = (y_t)_{t=0}^{T}$, where $T \in \mathbb{N}_0$.

In econometrics, a stochastic process Y is often assumed to be decomposed in the following manner:

$$Y_t = \mu_t + \varepsilon_t,$$

where $\mu = (\mu_t)_{t\in\mathbb{N}_0}$ is considered a mean process and $\varepsilon = (\varepsilon_t)_{t\in\mathbb{N}_0}$ is an error process. This decomposition will be utilized in the final section of this chapter. Note that the term time series will be used synonymously for a stochastic process.

1.2 Main Definitions and Results

The fundamental definitions and results related to copulas will be introduced in this section. To begin with, we will offer a formal definition of a copula. While a wide variety of definitions can be found in the literature, we will be using the following.

Definition 1.1 (Copula)**.**
A d-dimensional copula is a distribution function $C : [0, 1]^d \to [0, 1]$ with standard uniform marginals that satisfies:

(i) $C(u_1, \ldots, u_d) = 0$ whenever $u_i = 0$ for at least one $i \in \{1, \ldots, d\}$.

(ii) $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}$.

(iii) C is d-increasing, i.e., for all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0,1]^d$ with $a_i \leqslant b_i$:

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geqslant 0,$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \ldots, d\}$.

The third property ensures that if the random vector $U = (U_1, \ldots, U_d)$ has distribution function C, then $\mathbb{P}(a_1 \leqslant U_1 \leqslant b_1, \ldots, a_d \leqslant U_d \leqslant b_d)$ is non-negative. For notational purposes, we define $\mathscr{C}_d$ as the family of d-dimensional copulas. Also, as in the above definition, u will denote a realization of $U \sim \text{Unif}(0,1)$ for the remainder of this thesis. We now recall the definition of the quantile function.

Definition 1.2 (Quantile Function)**.**
Let F be a distribution function. The function $F^{-1} : [0,1] \to \mathbb{R}$ defined by:

$$F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : \alpha \leqslant F(x)\},$$

is called the quantile function.

As the quantile function only coincides with the standard inverse when F is continuous and strictly increasing[1], we should point out that a small abuse of notation has been used here. Since copulas take uniform realizations as inputs, the subsequent transformations are highly beneficial.

Proposition 1.3 (Probability Integral Transform)**.**
Let X be a random variable. If its distribution function F_X is continuous and strictly increasing, then:

$$F_X(X) \sim \text{Unif}(0,1).$$

Proof. By defining $Z = F_X(X)$, we see that:

$$F_Z(z) = \mathbb{P}(Z \leqslant z) = \mathbb{P}(F_X(X) \leqslant z) = \mathbb{P}(X \leqslant F_X^{-1}(y)) = F_X(F_X^{-1}(z)) = z,$$

where the third step follows from the fact that F_X is continuous and strictly increasing. ■

As a result, when transforming a continuous random variable by its distribution function, a standard uniform random variable will be obtained.

Proposition 1.4 (Inverse Probability Integral Transform)**.**
Let X be a random variable with distribution function F_X. If $U \sim \text{Unif}(0,1)$, then:

$$F_X^{-1}(U) \sim F_X.$$

Proof. The proof is shown in the case of F_X being continuous and strictly increasing. By defining $Z = F_X^{-1}(U)$, we see that:

$$F_Z(z) = \mathbb{P}(Z \leqslant z) = \mathbb{P}(F_X^{-1}(U) \leqslant z) = \mathbb{P}(U \leqslant F_X(z)) = F_X(z),$$

which concludes the proof. ■

[1] This assumption is reasonable in many financial applications since continuous distributions are typically considered.

Hence, $F_X^{-1}(U)$ has F_X as its distribution function if U is a uniform random variable on $[0,1]$. When generating a uniform sample, this result is advantageous by allowing the transformation of the samples to have distribution F_X.

Now, suppose that $X_1, \dots, X_d$ have a joint distribution function F as well as continuous and strictly increasing marginals $F_1, \dots, F_d$. By the probability integral transform, it follows that each of the $F_1(X_1), \dots, F_d(X_d)$ is distributed uniformly on $[0,1]$. Therefore, the joint distribution function of $F_1(X_1), \dots, F_d(X_d)$ is a copula which we denote C. This copula contains all information about the dependencies among the components of X but has no information about the marginal distribution functions of X. An expression for C can be found in the following way:

$$\begin{aligned} C(u_1, \dots, u_d) &= \mathbb{P}(F_1(X_1) \leqslant u_1, \dots, F_d(X_d) \leqslant u_d) \\ &= \mathbb{P}(X_1 \leqslant F_1^{-1}(u_1), \dots, X_d \leqslant F_d^{-1}(u_d)) \\ &= F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)). \end{aligned} \tag{1.1}$$

By letting $u_i = F_i(x_i)$ for $i = 1, \dots, d$, (1.1) reads instead:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

This is part of the well-known Sklar's theorem which we now formally state.

Theorem 1.5 (Sklar's Theorem)**.**
Let $X_1, \dots, X_d$ be random variables with joint distribution function F and let $F_1, \dots, F_d$ be its marginals. Then there exists a d-dimensional copula $C \in \mathscr{C}_d$ such that:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \tag{1.2}$$

for all $(x_1, \dots, x_d) \in \mathbb{R}^d$. Furthermore, if the marginals $F_1, \dots, F_d$ are continuous, then C is unique; otherwise C is uniquely determined only on $\operatorname{range}(F_1) \times \cdots \times \operatorname{range}(F_d)$. Conversely, if C is a copula and $F_1, \dots, F_d$ are distribution functions, then the function F defined in (1.2) is a joint distribution function with marginals $F_1, \dots, F_d$.

Consequently, Sklar's theorem states that any joint distribution function can be expressed in terms of its marginals and a copula function. This aspect makes copulas an extremely flexible tool for statistical analysis since it is not necessary to construct a full joint model for all variables at once. Instead, one might model each variable individually before modeling their dependence using a copula. For a proof, the interested reader is referred to [7].

In connection with Sklar's theorem, we introduce the notation of the copula density c in the following manner:

$$c(u_1, \dots, u_d) = \frac{\partial^d}{\partial u_1, \dots, \partial u_d} C(u_1, \dots, u_d). \tag{1.3}$$

Differentiating (1.2) by use of the chain rule then yields:

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) f_1(x_1) \cdot \cdots \cdot f_d(x_d), \tag{1.4}$$

where f_i denotes the marginal density of X_i for $i \in \{1, \dots, d\}$. Thus, any joint density function can be expressed by the product of its marginal densities and a copula density.

The version of Sklar's theorem in Theorem 1.5 only provides a way to model the joint distribution using a copula, without considering any conditioning variables. However, in finance, it may be important to model the conditional dependence between stock prices

given their past values or some other economic indicators. Therefore, the conditional Sklar's theorem, which is presented below, is often considered more useful than the unconditional version. To this end, we consider a conditional copula denoted as $C(\cdot \mid \mathcal{F})$ which is defined in a similar manner to the definition of a copula in Definition 1.1, but with the inclusion of some conditioning set $\mathcal{F}$.

Theorem 1.6 (Sklar's Theorem for Conditional Distributions)**.**
Let $X_1, \ldots, X_d$ be random variables, $\mathcal{F}$ be some conditioning set, and F be the joint conditional distribution of $(X_1, \ldots, X_d) \mid \mathcal{F}$ with conditional marginal distribution functions $F_1, \ldots, F_d$. Then there exists a d-dimensional conditional copula $C \in \mathscr{C}_d$ such that:

$$F(x_1, \ldots, x_d \mid \mathcal{F}) = C(F_1(x_1 \mid \mathcal{F}), \ldots, F_d(x_d \mid \mathcal{F}) \mid \mathcal{F}), \tag{1.5}$$

for all $(x_1, \ldots, x_d) \in \mathbb{R}^d$. Furthermore, if the conditional marginals $F_1, \ldots, F_d$ are continuous in $x_1, \ldots, x_d$ given $\mathcal{F}$, then C is unique. Conversely, if C is a conditional copula and $F_1, \ldots, F_d$ are conditional distribution functions, then the function F defined in (1.5) is a joint conditional distribution function with conditional marginals $F_1, \ldots, F_d$.

Hence, the fact that any joint distribution function can be defined by means of its marginals and a copula function holds true when considering conditional distribution functions as well. For a proof, the interested reader is referred to [19].

Lastly, we note that any high-dimensional copula can be constructed by multiplying bivariate copulas, that is 2-dimensional copulas, after a suitable transformation of the marginals. Hence, it is possible to consider high-dimensional copulas by solely considering bivariate copulas. As a result, any results shown in this thesis concerning bivariate copulas can be taken into account even if one is considering high-dimensional copulas. For more details, the interested reader is referred to [8].

1.3 Examples of Copulas

In the following subsections, different types of copulas will be introduced. First, the so-called special copulas will be presented, which as the name suggests deal with copulas for special cases of dependence structure. The presented copulas are the unconditional versions.

1.3.1 Special Copulas

Three copulas, notably the independence, co-monotonicity, and counter-monotonicity copulas, are of particular interest. The independence copula represents independence between variables, whereas the co- and counter-monotonicity copulas are two different extreme cases of dependence. In particular, these copulas are of interest, as many copulas approach special copulas in their limits. First, the independence copula is defined.

Definition 1.7 (Independence Copula)**.**
Let $C_0 \in \mathscr{C}_d$, then C_0 is called the independence copula if:

$$C_0(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i. \tag{1.6}$$

Thus, C_0 is the joint distribution of d mutually independent standard uniform random variables. The equality in (1.6) follows immediately from the fact that $u_i \perp\!\!\!\perp u_j$ for all $i, j \in \{1, \ldots, d\}$, $i \neq j$. Having defined a copula for independence, it is natural to consider copulas of perfect positive and negative dependence. In the former case, the following definition applies.

Definition 1.8 (Co-Monotonicity Copula)**.**
Let $C_+ \in \mathscr{C}_d$, then C_+ is called the co-monotonicity copula if:

$$C_+(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d). \tag{1.7}$$

To realize that this is the copula corresponding to perfect dependence, consider $U \sim \text{Unif}(0,1)$ and let $u_1, \ldots, u_d$ be realizations of $U, \ldots, U$. Then:

$$\mathbb{P}(U \leqslant u_1, \ldots, U \leqslant u_d) = \mathbb{P}(U \leqslant \min(u_1, \ldots, u_d)) = \min(u_1, \ldots, u_d).$$

It holds that the co-monotonicity copula corresponds to the Fréchet-Hoeffding upper bound presented in Appendix A, such that $C(u_1, \ldots, u_d) \leqslant C_+(u_1, \ldots, u_d)$ for all $C \in \mathscr{C}_d$. Equivalently, a lower bound exists for all copulas, namely the Fréchet-Hoeffding lower bound. This bound is not a copula for $d \geqslant 3$, however, for $d = 2$, this is indeed a copula. This copula coincides with the bivariate copula corresponding to perfect negative dependence, which is defined in the following.

Definition 1.9 (Counter-Monotonicity Copula)**.**
Let $C_- \in \mathscr{C}_2$, then C_- is called the counter-monotonicity copula if:

$$C_-(u_1, u_2) = \max(u_1 + u_2 - 1, 0). \tag{1.8}$$

To realize that this is the copula corresponding to perfect negative dependence, consider again $U \sim \text{Unif}(0,1)$ and two realizations u_1, u_2. Then:

$$\mathbb{P}(U \leqslant u_1, 1 - U \leqslant u_2) = \mathbb{P}(1 - u_2 \leqslant U \leqslant u_1) = \max(u_1 + u_2 - 1, 0).$$

As remarked above, $C_-(u_1, u_2)$ corresponds to the Fréchet-Hoeffding lower bound for $d = 2$ which implies that $C_-(u_1, u_2) \leqslant C(u_1, u_2)$ for all $C \in \mathscr{C}_2$.

1.3.2 Parametric Copulas

In this subsection, we define two parametric copulas, namely the Gaussian and t-copulas. The Gaussian copula assumes the dependence structure follows that of a multivariate Gaussian distribution and is defined as follows.

Definition 1.10 (Gaussian Copula)**.**
Let $C_N \in \mathscr{C}_d$ and Φ denote the distribution function of the standard univariate Gaussian distribution. Then C_N is called a Gaussian copula with correlation matrix $R \in \mathbb{R}^{d \times d}$ if:

$$C_N(u_1, \ldots, u_d; R) = \Phi_R\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big),$$

where Φ_R denotes the distribution function of the multivariate Gaussian distribution with mean zero and correlation matrix R.

In the bivariate case, it is noteworthy that a correlation matrix R can be represented by a single scalar $\rho \in (-1, 1)$. The Gaussian copula can be interpreted as transforming the uniform realizations $u_1, \ldots, u_d$ to Gaussian realizations by Proposition 1.4, and evaluating them in a multivariate Gaussian distribution. Note that a Gaussian copula reduces to the independence copula C_0 in (1.6) if $R = I_d$. If all correlations in R approach 1, then C_N approaches the co-monotonicity copula C_+ in (1.7). Furthermore, for the bivariate case, C_N approaches the counter-monotonicity copula C_- in (1.8) if the pair-wise correlation ρ approaches -1.

Similar to how we are able to derive a copula from the Gaussian distribution, we can also derive a copula from the t-distribution.

Definition 1.11 (t-Copula)**.**
Let $C_t \in \mathscr{C}_d$ and t_ν denote the distribution function of a univariate t-distribution with $\nu > 0$ degrees of freedom. Then C_t is called a t-copula with correlation matrix $R \in \mathbb{R}^{d\times d}$ if:

$$C_t(u_1, \dots, u_d; \nu, R) = t_{\nu,R}\big(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)\big), \tag{1.9}$$

where $t_{\nu,R}$ denotes the distribution function of the multivariate t-distribution with $\nu > 0$ degrees of freedom and correlation matrix R.

Note that the degrees of freedom affect both the multivariate distribution $t_{\nu,R}$ and the quantile functions t_ν^{-1}. As a consequence, both the multivariate distribution and the quantile functions converge to those of the standard Gaussian distribution for $\nu \to \infty$. From this, it follows easily that $C_t \to C_N$ for $\nu \to \infty$. Examples of Gaussian and t-copula densities are displayed in Figure 1.1. It is evident that the t-copula assigns much more probability mass to the corners when compared to the Gaussian copula. The reason for this is clarified by considering the tail dependence of the t-copula, which is discussed in Section 1.4.

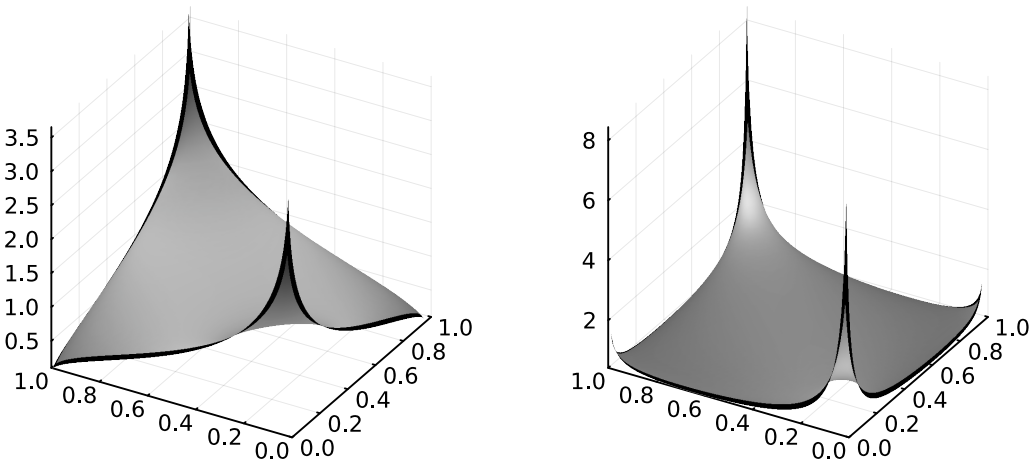


Figure 1.1: Gaussian copula density with parameter $\rho = 0.3$ (left) and t-copula density with parameters $\rho = 0.3$ and $\nu = 4$ (right).

1.3.3 Archimedean Copulas

In this subsection, we will discuss the Archimedean copula family, which is widely used in practice due to its simple and explicit formulas that typically rely on a single parameter. This simplicity is particularly advantageous when estimating the copula, as opposed to Gaussian and t-copulas which require the estimation of multivariate distributions.

Definition 1.12 (Archimedean Copula)**.**
An Archimedean copula is a copula $C \in \mathscr{C}_d$ of the form:

$$C(u_1, \dots, u_d) = \varphi\big(\varphi^{-1}(u_1) + \cdots + \varphi^{-1}(u_d)\big), \tag{1.10}$$

where the generator function $\varphi : [0, \infty] \to [0, 1]$ is a continuous, strictly decreasing, and convex function satisfying $\varphi(\infty) = 0$ and $\varphi(0) = 1$.

Note that the independence copula C_0 is a special case of an Archimedean copula for $\varphi(u) = \mathrm{e}^{-u}$. There are several different Archimedean copulas, each separated by how the φ

function is defined. As evident from (1.10), either the generator function φ or the inverse generator function φ^{-1} needs to be specified. The subsequent discussion will introduce a number of Archimedean copulas, beginning with the Frank copula.

Frank Copula

The Frank copula is an Archimedean copula with generator and inverse generator functions:

$$\varphi_F(u;\theta) = -\log\Big(\mathrm{e}^{-u}(\mathrm{e}^{-\theta}-1)+1\Big)/\theta, \qquad \varphi_F^{-1}(u;\theta) = -\log\left(\frac{\mathrm{e}^{-\theta u}-1}{\mathrm{e}^{-\theta}-1}\right),$$

where $\theta \in (0,\infty)$. Hence, the Frank copula $C_F \in \mathscr{C}_d$ is given by:

$$C_F(u_1,\ldots,u_d;\theta) = -\frac{1}{\theta}\log\left(1+\frac{\prod_{i=1}^{d}(\mathrm{e}^{-\theta u_i}-1)}{(\mathrm{e}^{-\theta}-1)^{d-1}}\right). \tag{1.11}$$

The Frank copula has some interesting convergence properties. First, as $\theta \to 0$ it holds that $C_F \to C_0$. Second, as $\theta \to \infty$ it holds that $C_F \to C_+$. Last, for $C_F \in \mathscr{C}_2$, it holds that $C_F \to C_-$ as $\theta \to -\infty$.

Gumbel Copula

The Gumbel copula is an Archimedean copula with generator and inverse generator functions:

$$\varphi_G(u;\theta) = \mathrm{e}^{-u^{\frac{1}{\theta}}}, \qquad \varphi_G^{-1}(u;\theta) = (-\log(u))^{\theta},$$

where $\theta \in [1,\infty)$. Hence, the Gumbel copula $C_G \in \mathscr{C}_d$ can be expressed as:

$$C_G(u_1,\ldots,u_d;\theta) = \exp\left(-\Big[(-\log(u_1))^{\theta}+\cdots+(-\log(u_d))^{\theta}\Big]^{\frac{1}{\theta}}\right).$$

The Gumbel copula reduces to C_0 if $\theta = 1$, as the generator function reduces to e^{-u} whereas, if $\theta \to \infty$ it holds that $C_G \to C_+$. However, unlike the Frank copula, C_G cannot converge to the counter-monotonicity copula C_-. Examples of Frank and Gumbel copula densities are displayed in Figure 1.2. Clearly, the Gumbel copula contains a greater amount of probability mass in its tails than the Frank copula. In particular, the Gumbel copula is skewed towards the upper tail and exhibits an asymmetric shape.

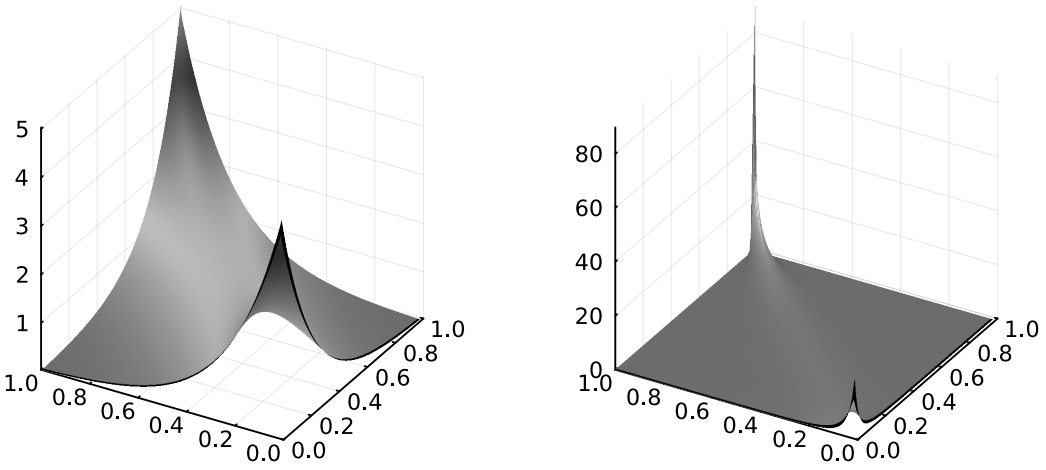


Figure 1.2: Frank copula density with parameter $\theta = 5$ (left) and Gumbel copula density with parameter $\theta = 4$ (right).

Other Archimedean Copulas

As described above, the difference between Archimedean copulas is solely determined by the generator function. The following table displays the generator functions and copula expressions for four additional commonly used copulas.

	Generator Function	Copula	Parameter Space
Clayton	$(u+1)^{-\frac{1}{\theta}}$	$\left(\sum_i^d (u_i^{-\theta}) + 1 - d\right)^{-\frac{1}{\theta}}$	$\theta \in (0, \infty)$
Joe	$1-(1-\mathrm{e}^{-u})^{\frac{1}{\theta}}$	$1-\left(1-\prod_{i=1}^d 1-(1-u_i)^{\theta}\right)^{\frac{1}{\theta}}$	$\theta \in [1, \infty)$
BB7 (Joe-Clayton)	$1-\left(1-(u+1)^{-\frac{1}{\delta}}\right)^{\frac{1}{\theta}}$	$1-\left(1-\left(\sum_{i=1}^d (1-(1-u_i)^{\theta})^{-\delta} - d + 1\right)^{-\frac{1}{\delta}}\right)^{\frac{1}{\theta}}$	$\theta \in [1, \infty), \delta \in (0, \infty)$
Ali-Mikhail-Haq	$\frac{1-\theta}{\mathrm{e}^u - \theta}$	$(1-\theta)/\prod_{i=1}^d \left(\frac{1-\theta}{u_i} + \theta\right)$	$\theta \in [0, 1)$

Table 1.1: Generator functions, copula expressions and parameter spaces for other Archimedean copulas.

Similar to the Frank and Gumbel copulas it is possible to obtain convergence results for the four copulas in Table 1.1, however, these are not presented in this thesis. From the table, it is evident that the BB7 copula is a composition of the Clayton and Joe copulas. Specifically, the generator function of the Joe copula is modified by substituting the term e^{-u} with the generator function of the Clayton copula. The BB7 copula is of particular interest, as it includes two parameters, which allow the model to be more flexible.

1.4 Tail Dependence

In the context of modeling financial data using copulas, where the actual copula is often unknown, it is important to choose a copula that accurately reflects the joint behavior of the data. In this regard, tail dependence, which relies solely on the associated copula, is used to quantify the relationship between extreme values of two random variables. This quantification renders copulas particularly useful in risk management, where small (large) values of tail dependence indicate a low (high) risk of simultaneous extreme returns. The lower and upper tail dependence can be defined as follows.

Definition 1.13 (Tail Dependence)**.**
Let X_1 and X_2 be random variables with continuous and strictly increasing distribution functions F_1 and F_2, respectively. Then the lower and upper tail dependence coefficients are, respectively, given by:

$$\lambda_l = \lim_{\alpha \to 0} \mathbb{P}\big(X_1 \leqslant F_1^{-1}(\alpha) \mid X_2 \leqslant F_2^{-1}(\alpha)\big), \tag{1.12}$$

$$\lambda_u = \lim_{\alpha \to 1} \mathbb{P}\big(X_1 \geqslant F_1^{-1}(\alpha) \mid X_2 \geqslant F_2^{-1}(\alpha)\big), \tag{1.13}$$

where $\alpha \in (0, 1)$. If λ_l or λ_u are non-zero, then X_1 and X_2 are said to have lower or upper tail dependence, respectively.

If we assume that $X_1 \perp\!\!\!\perp X_2$ then the conditional probability in (1.12) equals the unconditional probability $\mathbb{P}(X_1 \leqslant F_1^{-1}(\alpha))$. As $\alpha \to 0$, this probability approaches 0, indicating that $\lambda_l = 0$ implies independence in the extreme left tail. Similar inferences can be drawn from (1.13). When exploring copulas, these equations reduce to a limiting expression including only a copula. First, consider the result for the lower tail dependence.

Proposition 1.14 (Lower Tail Dependence)**.**
Let X_1 and X_2 be random variables with copula $C \in \mathscr{C}_2$ and continuous and strictly increasing distribution functions F_1 and F_2, respectively. Then the lower tail dependence coefficient is given by:

$$\lambda_l = \lim_{\alpha \to 0} \frac{C(\alpha, \alpha)}{\alpha}, \quad \alpha \in (0, 1).$$

Proof. Consider the definition of lower tail dependence:

$$\begin{aligned}
\lambda_l &= \lim_{\alpha \to 0} \mathbb{P}\big(X_1 \leqslant F_1^{-1}(\alpha) \mid X_2 \leqslant F_2^{-1}(\alpha)\big) \\
&= \lim_{\alpha \to 0} \frac{\mathbb{P}\big(X_2 \leqslant F_2^{-1}(\alpha), X_1 \leqslant F_1^{-1}(\alpha)\big)}{\mathbb{P}\big(X_2 \leqslant F_2^{-1}(\alpha)\big)} \\
&= \lim_{\alpha \to 0} \frac{\mathbb{P}(F_2(X_2) \leqslant \alpha, F_1(X_1) \leqslant \alpha)}{\mathbb{P}(F_2(X_2) \leqslant \alpha)} \\
&= \lim_{\alpha \to 0} \frac{C(\alpha, \alpha)}{\alpha},
\end{aligned}$$

where the penultimate step follows from the fact that F_1 and F_2 are continuous and strictly increasing. ■

Deriving the tail dependence coefficients for the previously discussed copulas is beyond the scope of this project. Nonetheless, results regarding the lower tail dependence of the Gaussian and t-copulas are presented. For bivariate Gaussian copulas $C_N \in \mathscr{C}_2$ it holds that $\lambda_l^N = 0$. For a bivariate t-copula $C_t \in \mathscr{C}_2$ with ν degrees of freedom and correlation coefficient ρ, the coefficient of lower tail dependence is given by:

$$\lambda_l^t = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{\rho+1}}\right), \tag{1.14}$$

where $t_{\nu+1}$ denotes the univariate distribution function of the t-distribution with $\nu + 1$ degrees of freedom. From (1.14) it is evident that $\nu \to \infty$ implies $\lambda_l^t \to 0$, which is desired, as $C_t \to C_N$ for $\nu \to \infty$. Next, we consider the result for the upper tail dependence coefficient corresponding to Proposition 1.14.

Proposition 1.15 (Upper Tail Dependence)**.**
Let X_1 and X_2 be random variables with copula $C \in \mathscr{C}_2$ and continuous and strictly increasing distribution functions F_1 and F_2, respectively. Then the upper tail dependence coefficient is given by:

$$\lambda_u = 2 - \lim_{\alpha \to 1} \frac{1 - C(\alpha, \alpha)}{1 - \alpha}, \quad \alpha \in (0, 1). \tag{1.15}$$

Proof. Consider the definition of upper tail dependence:

$$\begin{aligned}
\lambda_u &= \lim_{\alpha \to 1} \mathbb{P}\big(X_1 \geqslant F_1^{-1}(\alpha) \mid X_2 \geqslant F_2^{-1}(\alpha)\big) \\
&= \lim_{\alpha \to 1} \frac{\mathbb{P}(F_2(X_2) \geqslant \alpha, F_1(X_1) \geqslant \alpha)}{1 - \mathbb{P}(F_2(X_2) \leqslant \alpha)} \\
&= \lim_{\alpha \to 1} \frac{1 - \mathbb{P}(F_2(X_2) \leqslant \alpha) - \mathbb{P}(F_1(X_1) \leqslant \alpha) + \mathbb{P}(F_2(X_2) \leqslant \alpha, F_1(X_1) \leqslant \alpha)}{1 - \alpha} \\
&= \lim_{\alpha \to 1} \frac{1 - \alpha - \alpha + C(\alpha, \alpha)}{1 - \alpha}
\end{aligned}$$

$$= 1 + \lim_{\alpha \to 1} \frac{-1 + 1 - \alpha}{1 - \alpha} + \frac{C(\alpha, \alpha)}{1 - \alpha}$$
$$= 2 - \lim_{\alpha \to 1} \frac{1 - C(\alpha, \alpha)}{1 - \alpha},$$

where the third step follows from the fact that $\mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cup B^c) + \mathbb{P}(A^c \cap B^c)$. ■

For Gaussian and t-copulas it holds that $\lambda_l = \lambda_u$. This implies that the t-copula is much better suited to model tail dependence compared to the Gaussian copula. These tail dependence results are also clearly visible in Figure 1.1.

1.5 Copula Estimation

This section is based on [22] and [23].

There exist two distinct approaches to statistical analysis, namely the frequentist approach and the Bayesian approach. Although the frequentist approach and the use of maximum likelihood estimation is the most common method for copula estimation, this thesis will concentrate on using the Bayesian approach for making inferences. Therefore, a brief introduction to utilizing the Bayesian approach for copula parameter estimation is presented.

The typical procedure for conducting Bayesian inference involves the following steps. First, a density function $\pi(\theta)$, called the prior density, is chosen to express our beliefs regarding the parameter(s) θ before any data is observed. Next, a statistical model $\pi(x \mid \theta)$, often referred to as the likelihood, is selected to represent our beliefs about the data $x = (x_1, \ldots, x_n)$ given θ where $n \in \mathbb{N}$. Finally, upon observing data x, we revise our beliefs and compute the posterior density $\pi(\theta \mid x)$. The very essence of Bayesian inference lies in the posterior density, which encapsulates all that we can infer about the parameter(s) based on the observed data. Given a prior density and a likelihood function, the posterior density can be written as follows by Bayes formula:

$$\pi(\theta \mid x) = \frac{\pi(x \mid \theta)\pi(\theta)}{\pi(x)} \propto \pi(x \mid \theta)\pi(\theta), \tag{1.16}$$

where $\pi(x) = \int_{-\infty}^{\infty} \pi(x \mid \theta)\pi(\theta)\mathrm{d}\theta$ is a normalizing constant that does not depend on θ. Omitting this from the estimation procedure is common practice, as it reduces efficiency and can always be recovered afterward. Now that the posterior density is obtained, a Bayesian estimate of the unknown parameter(s) θ is simply the mean of the posterior density, that is:

$$\hat{\theta} = \mathbb{E}[\theta \mid x] = \int_{-\infty}^{\infty} \theta\pi(\theta \mid x)\mathrm{d}\theta. \tag{1.17}$$

Similarly, one can find the posterior variance in the following way:

$$\mathrm{Var}(\theta \mid x) = \mathbb{E}\big[(\theta - \mathbb{E}[\theta \mid x])^2 \mid x\big] = \int_{-\infty}^{\infty} (\theta - \mathbb{E}[\theta \mid x])^2\pi(\theta \mid x)\mathrm{d}\theta. \tag{1.18}$$

To provide a complete summary of the unknown parameter(s), an interval can be identified with a specific probability of containing the unknown parameter(s), which is commonly referred to as a credible interval. This interval serves as the Bayesian analogue of a confidence interval. A $1 - \alpha$ credible interval is defined as a subset $\mathcal{C} \subset \Theta$, where Θ denotes the parameter space, such that:

$$\mathbb{P}(\theta \in \mathcal{C} \mid x) = \int_{\mathcal{C}} \pi(\theta \mid x)\mathrm{d}\theta = 1 - \alpha, \quad \alpha \in (0, 1). \tag{1.19}$$

Once the posterior mean and variance, as well as the credible interval, have been determined, we possess a complete summary of the unknown parameter(s) θ. However, in many cases the integrals in (1.17), (2.1) and (1.19) may not be easy to compute. Therefore, Bayesian inference typically requires an alternative approach. One such alternative is to draw samples from the posterior and utilize them as a foundation for inference. This method remains effective even when we lack knowledge of the exact form of the posterior and only have information about its unnormalized version. By generating a sample $\theta_1, \dots, \theta_m$ where $m \in \mathbb{N}$ from the posterior, one can approximate the integrals in (1.17) and (2.1) as follows:

$$\int_{-\infty}^{\infty} \theta \pi(\theta \mid x) \mathrm{d}\theta \approx \frac{1}{m} \sum_{i=1}^{m} \theta_i = \bar{\theta},$$
$$\int_{-\infty}^{\infty} (\theta - \mathbb{E}[\theta \mid x])^2 \pi(\theta \mid x) \mathrm{d}\theta \approx \frac{1}{m} \sum_{i=1}^{m} (\theta_i - \bar{\theta})^2,$$

whereas credible intervals can be derived from the quantiles of the posterior samples. This allows us to approximate the posterior mean and variance, as well as credible intervals, in situations where computing the integral of the posterior density is not feasible.

To formulate our Bayesian approach for copula parameter estimation, we need to specify a suitable prior and likelihood. Since there is little knowledge about the parameter(s), a uniform (flat) prior will be used to represent an uninformative prior. As a consequence of (1.4), the likelihood of the observations $x = (x_1, \dots, x_n)$ is given by $\pi(x \mid \Psi, \theta) = \prod_{i=1}^{n} \pi(x_i \mid \Psi, \theta)$ where $x_i = (x_{1,i}, \dots, x_{d,i})$ for $n, d \in \mathbb{N}$, and:

$$\pi(x_i \mid \Psi, \theta) = c(u_i \mid \theta) \prod_{j=1}^{d} \pi_j(x_{j,i} \mid \psi_j).$$

Here $u_i = (u_{1,i}, \dots, u_{d,i})$ where $u_{j,i} = F_j(x_{j,i} \mid \psi_j)$, $\Psi = \{\psi_1, \dots, \psi_d\}$ are any parameters of the marginal models, $\pi_j(x_{j,i} \mid \psi_j) = \frac{\partial}{\partial x_{j,i}} F_j(x_{j,i} \mid \psi_j)$ is the marginal density of $x_{j,i}$, and $c(u_i \mid \theta)$ is the copula density of u_i. We now possess all the necessary components to determine the posterior density in (1.16). Nevertheless, this is one of the instances where it may be challenging to compute the integral of the posterior density. As mentioned, an alternative approach is to generate samples from the posterior and use them as the foundation for making inferences. To accomplish this, one can utilize MCMC methods, which create a Markov chain that has the posterior distribution as its equilibrium distribution.

1.6 Goodness-of-Fit Test

The following section is based on [3], [4] and [11].

We would like to determine whether a copula is properly specified. In order to do this, we introduce the Cramér-von Mises test which is a commonly used goodness-of-fit test. The null hypothesis of the test is that the estimated copula is the true underlying copula. Suppose that we have a random d-variate vector Z where $d \in \mathbb{N}$. The test is then based on a so-called pseudo-vector $V = (V_1, \dots, V_n)$ where $n \in \mathbb{N}$:

$$V_i = (V_{i1}, \dots, V_{id}) = \left(\frac{R_{i1}}{1+n}, \dots, \frac{R_{id}}{1+n} \right),$$

and R_{ij} denotes the rank of Z_{ij} amongst $(Z_{1j}, \dots, Z_{nj})$. Given a realization of Z, the realization of V can be considered to be a sample from the true underlying copula C. The

approach is further based on the empirical copula[2] defined by:

$$\hat{C}(u_1, \ldots, u_d) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}_{\{V_{i1} \leqslant u_1, \ldots, V_{id} \leqslant u_d\}},$$

where $(u_1, \ldots, u_d) \in [0,1]^d$. The empirical copula can be considered a consistent estimator of the true underlying copula. The idea of the test is then to compare the distance between the empirical copula and the estimated copula. This is done by the Cramér-von Mises test statistic which is given by:

$$\text{CvM} = \sum_{i=1}^{n} \left\{ \hat{C}(v_i) - C(v_i \mid \hat{\theta}) \right\}^2, \tag{1.20}$$

where $C(v_i \mid \hat{\theta})$ is the estimated copula and $v = (v_1, \ldots, v_n)$ is a realization of V. Approximate p-values can be deduced from the limiting distribution of (1.20). However, this limiting distribution depends on unknown parameter value(s). Therefore, the asymptotic distribution of the test statistics cannot be tabulated and approximate p-values can only be obtained via specially adapted Monte Carlo methods. In particular, a parametric bootstrap procedure can be used for this purpose. For more details on this, the interested reader is referred to [3].

1.7 Marginal Distribution Modelling

This section is based on [1], [11] and [18].

As Sklar's theorem states that any joint distribution can be expressed in terms of its marginals and a copula function, it is of particular importance to correctly specify the marginal distributions. When modelling marginals, or stochastic processes in general, econometricians often default to the ARMA-GARCH type models. These models are great for removing autocorrelation and heteroscedasticity, such that an *i.i.d.* innovation process can be obtained. Recall from Section 1.1 that a stochastic process $Y = (Y_t)_{t \in \mathbb{N}_0}$ is assumed to be decomposed in the following manner:

$$Y_t = \mu_t + \varepsilon_t. \tag{1.21}$$

First, we consider the mean process μ, which to account for possible autocorrelation is assumed to follow an ARMA model. That is, (1.21) becomes:

$$Y_t = \varepsilon_t + \underbrace{c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \kappa_j \varepsilon_{t-j}}_{=\mu_t},$$

where $1 < |\phi_i|$ for all $i \in \{1, \ldots, p\}$ is required to maintain stationarity and $c \in \mathbb{R}$. The error process ε is assumed to be decomposed by the use of a GARCH type framework:

$$Y_t - \mu_t = \varepsilon_t = \big(\underbrace{h(\mathscr{F}_{t-1}, \vartheta)}_{=h_t}\big)^{1/2} Z_t,$$

where the conditional variance h_t denotes a function of the information $\mathscr{F}_{t-1}$ gathered by past realizations $(Y_0 = y_0, \ldots, Y_{t-1} = y_{t-1})$, and GARCH model parameters ϑ. Furthermore,

[2] Note that although it is called the empirical copula it is not a copula.

for the innovation process Z we have that $Z_t \mid \mathscr{F}_{t-1} \overset{i.i.d.}{\sim} \mathcal{D}(0,1)$ and thus $Y_t \mid \mathscr{F}_{t-1} \overset{i.i.d.}{\sim} \mathcal{D}(\mu_t, h_t)$. Note that the possible candidates for the conditional distribution specification $\mathcal{D}(0,1)$ are defined in the following subsection. The conditional variance h_t is modelled using GARCH type models. This thesis will consider three GARCH type models, starting with the standard GARCH(p,q) model, abbreviated sGARCH(p,q):

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

where $p, q \in \mathbb{N}_0$, $\omega > 0$, $\alpha_i \geqslant 0$ for $i \in \{1, \dots, q\}$, and $\beta_j \geqslant 0$ for $j \in \{1, \dots, p\}$. Evidently, it is assumed that $\alpha_q > 0$ and $\beta_p > 0$. It should be noted that similar to the ARMA model, the model order of GARCH type models is indicated by the letters p and q because this is the standard notation in the literature.

The sGARCH model is capable of replicating the majority of stylized facts of log return series. This includes the tendency for log returns to exhibit volatility clustering, no autocorrelation, and a leptokurtic distribution. However, because the sGARCH model only accounts for the squared past values of volatility, it does not consider the impact of past values' signs on volatility. As a result, the sGARCH model is unable to reproduce a critical stylized fact known as the leverage effect, which refers to the disproportionate increase in future volatility that results from negative returns compared to positive returns. In what follows, two model extensions designed to mitigate this shortcoming are presented. The first of these two models is the exponential GARCH(p,q) abbreviated eGARCH(p,q):

$$\begin{aligned} \ln(h_t^2) &= \omega + \sum_{i=1}^{q} \alpha_i g(Z_{t-i}) + \sum_{j=1}^{p} \beta_j \ln(h_{t-j}^2), \\ g(Z_t) &= \lambda Z_t + \gamma(|Z_t| - \mathbb{E}[|Z_t|]), \end{aligned}$$

where ω, α_i for $i \in \{1, \dots, q\}$, β_j for $j \in \{1, \dots, p\}$, γ, and λ, are all real numbers. The second of the two model extensions is the threshold GARCH(p,q) abbreviated tGARCH(p,q):

$$h_t^{1/2} = \omega + \sum_{i=1}^{q} \big(\alpha_i \mathbb{1}_{\{\varepsilon_{t-i} \geqslant 0\}} - \gamma_i \mathbb{1}_{\{\varepsilon_{t-i} < 0\}}\big)\varepsilon_{t-i} + \sum_{j=1}^{p} \beta_j h_{t-j}^{1/2},$$

where $\omega > 0$, $\alpha_i \geqslant 0$ and $\gamma_i \geqslant 0$ for $i \in \{1, \dots, q\}$, and $\beta_j \geqslant 0$ for $j \in \{1, \dots, p\}$. Both the eGARCH and tGARCH models incorporate a γ parameter that effectively captures the leverage effect.

It should be noted, that $Z_t \mid \mathscr{F}_{t-1}$ denotes a random variable conditioned on the set of information at time $t-1$ for all marginals. That is if $d = 2$ we have two stochastic processes Y^1 and Y^2, which means that $\mathscr{F}_{t-1}$ is the set of information gathered from $(Y_0 = y_0, \dots, Y_{t-1} = y_{t-1})$ where $Y_t = (Y_t^1, Y_t^2)$ and likewise $y_t = (y_t^1, y_t^2)$. However, the ARMA-GARCH specification of the marginal models above is only conditioned on the subset $\mathscr{F}_{t-1}^i \subseteq \mathscr{F}_{t-1}$ for $i = 1, 2$. When using such marginal models, the copula is a true copula if and only if:

$$Y_t^i \mid \mathscr{F}_{t-1}^i \overset{d}{=} Y_t^i \mid \mathscr{F}_{t-1}, \tag{1.22}$$

for $i = 1, 2$. That is, if (1.22) does not hold, then $Y_t \mid \mathscr{F}_{t-1}$ does not have the marginal models specified above. To test if (1.22) holds, a univariate ARMAX-GARCHX model is constructed for both stochastic processes, and a Wald test for the joint nullity of the additional explanatory variables is performed. In the case that (1.22) holds, we can model each marginal by an ARMA-GARCH type model before modelling their dependence using a copula. Furthermore, the specific choice of ARMA-GARCH type models for the marginal modelling, implies that we are in fact considering the conditional Sklar's theorem.

1.7.1 Distribution Specification

This subsection introduces the particular distribution specifications of the innovations encountered in this thesis. An independent standard Gaussian distribution provides a first specification for the conditional distribution of the innovations Z. The density function of a standard Gaussian distribution is given by:

$$f_{\mathrm{N}}(z_t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z_t^2}.$$

A second specification is given by an independent Student's t-distribution with $\nu > 0$ degrees of freedom and scale parameter s. The student's t-distribution has density function:

$$f_{\mathrm{S}}(z_t; \nu, s) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{(\pi\nu)^{1/2}\Gamma\left(\frac{\nu}{2}\right)} \frac{s^{-1/2}}{\left(1 + \frac{z_t^2}{s\nu}\right)^{(\nu+1)/2}},$$

where Γ denotes the Gamma function. The variance of Z is given by[3]:

$$\mathrm{Var}(Z_t) = \frac{s\nu}{\nu - 2}.$$

Hence, in order to ensure the existence of the variance of Z and that it equals one, it must hold that $\nu > 2$ and $s = (\nu - 2)/\nu$. Thus, the scale parameter s is completely specified by the degrees of freedom ν for a Student's t-distribution.

A third specification is given by an independent standardized generalized error distribution (GED). The density function of a GED is:

$$f_{\mathrm{GED}}(z_t; \nu) := \frac{\nu \exp\left(-\frac{1}{2}\left|\frac{z_t}{\lambda}\right|^{\nu}\right)}{\lambda 2^{(\nu+1)/\nu}\Gamma\left(\frac{1}{\nu}\right)}, \quad \lambda := \left(\frac{2^{(-2/\nu)}\Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)}\right)^{1/2},$$

where the parameter $\nu > 0$ determines the tail-thickness. If $\nu = 2$, the density function corresponds to the standard normal distribution. When $\nu < 2$, the distribution of Z exhibits thicker tails compared to the standard normal distribution, while a value of $\nu > 2$ indicates thinner tails.

[3] Note that for ease of notation, $Z_t \mid \mathscr{F}_{t-1}$ is abbreviated Z_t.

2 Risk Forecasting

This chapter is based on [2], [5], [6], [14], [15], and [17] unless explicitly stated otherwise.

The subsequent sections delve into the concept of risk forecasting, which is a critical aspect of risk management. We begin by providing an introduction to value-at-risk and expected shortfall, including an overview of the forecasting procedure for these measures. Next, we explore the topic of model evaluation and selection, which is based on the accuracy of risk forecasts, with the ultimate goal of identifying the model(s) that offer the most precise risk forecasts.

2.1 Risk Measures

One of the most commonly used risk measures in financial institutions is value-at-risk (VaR), which can be utilized to assess the risk exposure of a portfolio. The VaR measures the threshold value such that the probability of observing a loss larger or equal to it in a given time horizon is smaller or equal to a specific risk level $\alpha \in (0,1)$. The formal definition of VaR at time t for some realization process y given the information set $\mathscr{F}_{t-1}$ at risk level α is given by:

$$\mathrm{VaR}^{\alpha}(Y_t \mid \mathscr{F}_{t-1}) = \inf\{y_t \in \mathbb{R} : F_{Y_t}(-y_t \mid \mathscr{F}_{t-1}) \leqslant \alpha\} = -F^{-1}_{Y_t}(\alpha \mid \mathscr{F}_{t-1}). \tag{2.1}$$

Another well-known risk measure is the so-called expected shortfall (ES), which is also known as the conditional value-at-risk (CVaR). The ES measures the expected return below the negative VaR level and it is hence closely related to VaR. Many risk managers opt for ES over VaR as it captures tail risk and is a coherent risk measure. The ES given $\mathscr{F}_{t-1}$ at risk level α is obtained as:

$$\mathrm{ES}^{\alpha}(Y_t \mid \mathscr{F}_{t-1}) = \mathbb{E}[Y_t \mid Y_t \leqslant \mathrm{VaR}^{\alpha}(Y_t \mid \mathscr{F}_{t-1})] = -\frac{1}{\alpha}\int_0^{\alpha} \mathrm{VaR}^{\zeta}(Y_t \mid \mathscr{F}_{t-1})\mathrm{d}\zeta. \tag{2.2}$$

Hence, instead of setting a specific risk level α, one averages VaR over all levels $\zeta \leqslant \alpha$, thereby delving deeper into the tail of the distribution of Y_t. For ease of notation, we denote (2.1) and (2.2) by VaR^{α}_t and ES^{α}_t, respectively. The techniques employed to forecasts these risk measures are explained in the following section.

2.2 Risk Forecasting Procedure

Our objective is to forecast VaR and ES on a daily basis, with a one-day-ahead horizon, for an equally weighted portfolio composed of two stocks. Let the daily log return process of a stock be defined by:

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(P_t) - \log(P_{t-1}), \quad t = 1, \ldots, T, \tag{2.3}$$

for some $T \in \mathbb{N}$, where P_t is the price of the stock at day t. We denote by Y^1 and Y^2 the processes of daily log returns for two specific stocks with corresponding price processes P^1 and P^2, respectively. The return process of an equally weighted portfolio composed of those two stocks is given by:

$$\mathcal{P}_t = \frac{1}{2}P_t^1 + \frac{1}{2}P_t^2 - \left(\frac{1}{2}P_{t-1}^1 + \frac{1}{2}P_{t-1}^2\right) = \frac{1}{2}P_{t-1}^1(\mathrm{e}^{Y_t^1} - 1) + \frac{1}{2}P_{t-1}^2(\mathrm{e}^{Y_t^2} - 1). \tag{2.4}$$

In order to forecast the portfolio VaR and ES, the joint distribution of the vector (Y_t^1, Y_t^2) must be investigated. We accomplish this using the previously discussed copula and ARMA-GARCH frameworks. Because there are no simple and analytical formulas to convert conditional mean and variance to VaR and ES of a portfolio, Monte Carlo simulation is used. The exact procedure utilized to forecast one-day-ahead VaR and ES at risk level α based on copulas is given as follows:

1. Fit an ARMA-GARCH type model to each return process and estimate marginal distributions for the innovation processes using T realizations of Y^1 and Y^2.

2. Forecast one-step-ahead return means $(\hat{\mu}_{T+1}^1, \hat{\mu}_{T+1}^2)$ and variances $(\hat{h}_{T+1}^1, \hat{h}_{T+1}^2)$.

3. Use the conditional bivariate distribution modelled by ARMA-GARCH and copula models to simulate N Monte Carlo scenarios over the time horizon $[T, T+1]$.

   (a) Estimate copula parameters by the probability integral transforms u_t^1 and u_t^2 of the innovations z_t^1 and z_t^2 as described in Section 1.5.

   (b) Simulate two random variables $(u_{T+1}^{2,j}, u_{T+1}^{1,j})$ where $j = 1, \ldots, N$, from the copula function estimated in step (a)[1].

   (c) Obtain the simulated innovations $z_{T+1}^{1,j}$ and $z_{T+1}^{2,j}$ by using the estimated marginal distributions and Proposition 1.4:

   $$\left(z_{T+1}^{1,j}, z_{T+1}^{2,j}\right) = \left(F_{1,T+1}^{-1}(u_{T+1}^{1,j}; \hat{\psi}_1), F_{2,T+1}^{-1}(u_{T+1}^{2,j}; \hat{\psi}_2)\right).$$

   (d) Obtain simulated log returns by using the simulated innovations from step (c) and the forecasted means and variances from step 2:

   $$(y_{T+1}^{1,j}, y_{T+1}^{2,j}) = \left(\hat{\mu}_{T+1}^1 + \sqrt{\hat{h}_{T+1}^1} \cdot z_{T+1}^{1,j}, \hat{\mu}_{T+1}^2 + \sqrt{\hat{h}_{T+1}^2} \cdot z_{T+1}^{2,j}\right)$$

   (e) Repeat steps (b)-(d) for $j = 1, \ldots N$ and calculate the values of $\mathcal{P}_{T+1}^j$ using (2.4) for $j = 1, \ldots, N$.

   (f) Sort the N values of $\mathcal{P}_{T+1}^j$ in increasing order and calculate $\hat{\mathrm{VaR}}_{T+1}^{\alpha}$ as the absolute value of the $N\alpha$'th ordered value.

   (g) Calculate ES at risk level α as the average of the values of $\mathcal{P}_{T+1}^j$ that are smaller than or equal to the negative VaR at risk level α:

   $$\hat{\mathrm{ES}}_{T+1}^{\alpha} = \frac{1}{N\alpha}\sum_{j=1}^{N} \mathcal{P}_{T+1}^j \mathbb{1}_{\mathcal{P}_{T+1}^j \leqslant -\hat{\mathrm{VaR}}_{T+1}^{\alpha}}.$$

[1] See [21] for a discussion about copula simulation.

4. Repeat steps 1-3 on a rolling window basis: Let $y^1_{1:T+M} = (y^1_t)_{t=1}^{T+M}$ and $y^2_{1:T+M} = (y^2_t)_{t=1}^{T+M}$, $M \in \mathbb{N}$ denote some realizations with corresponding price processes. Then repeat steps 1-3 iteratively, using $(y^1_{1:T}, y^2_{1:T}), \ldots, (y^1_{M:T+M-1}, y^2_{M:T+M-1})$ along with the corresponding stock prices used in step (e). At each step, obtain the one-day-ahead VaR and ES.

Increasing the value of N leads to greater accuracy in estimating VaR and ES. However, larger values of N also result in longer simulation times. In this thesis, we choose $N = 100000$ to strike a balance between accuracy and efficiency.

2.3 Risk Forecasting Evaluation

We would like to evaluate the risk forecasts for the different copula models obtained by the procedure described in Section 2.2. This is done in two steps in this thesis: To determine which (if any) of the copula models produces reasonable risk forecasts, we first analyze the accuracy of the produced VaR and ES forecasts, separately, using two statistical tests. After that, by introducing the so-called model confidence set, we determine which copula model(s) produces the most accurate risk forecasts. This section will introduce the two tests as well as the model confidence set.

2.3.1 Accuracy of VaR and ES Forecasts

The obtained risk forecasts are not necessarily correctly specified. Therefore, we introduce two statistical tests for evaluating the accuracy of VaR and ES forecasts relative to the realized portfolio return series given by (2.4). This is in the literature usually referred to as backtesting. The employed tests for the VaR and ES forecast are the so-called dynamic quantile and expected shortfall regression tests, respectively.

Dynamic Quantile Test

First, we consider the dynamic quantile (DQ) test for evaluating the accuracy of VaR forecasts. To this end, suppose that $M \in \mathbb{N}$ out-of-sample VaR forecasts $\hat{\text{VaR}}^\alpha = (\hat{\text{VaR}}^\alpha_t)_{t=T+1}^{T+M}$ at level $\alpha \in (0,1)$ are given, and define the hit variable process $H^\alpha = (H^\alpha_t)_{t=T+1}^{T+M}$ as:

$$H^\alpha_t = \mathbb{1}_{\{\mathcal{P}_t \leqslant -\hat{\text{VaR}}^\alpha_t\}}. \tag{2.5}$$

If $\hat{\text{VaR}}^\alpha$ is correctly specified, then the hit variable has a mean value of α and is distributed independently over time. Hence, testing if $\hat{\text{VaR}}^\alpha$ is correctly specified is essentially equivalent to testing whether all elements in H^α follow an $i.i.d.$ Bernoulli distribution with parameter α, that is:

$$H^\alpha \overset{i.i.d.}{\sim} \text{Bern}(\alpha). \tag{2.6}$$

To test (2.6) we use the DQ test, which is based on the demeaned hit variable $Hit^\alpha_t = H^\alpha_t - \alpha$. The DQ test tests for joint nullity of all coefficients in the following linear regression:

$$Hit^\alpha_t = \delta_0 + \sum_{l=1}^{L} \delta_l Hit^\alpha_{t-l} + \delta_{L+1} \hat{\text{VaR}}^\alpha_{t-1} + w_t, \quad t \in \{T+L, \ldots, T+M\}, \tag{2.7}$$

where $\delta = (\delta_0, \ldots, \delta_{L+1}) \in \mathbb{R}^{L+2}$ with $L \in \mathbb{N}$, and w_t is some error term. It should be noted that in this thesis, the number of lags is chosen to be $L = 4$ as suggested in [15]. To test

for joint nullity, the following Wald test statistic is used:

$$W_{\mathrm{DQ}} = \frac{\hat{\delta}' G' G \hat{\delta}}{\alpha(1-\alpha)}, \tag{2.8}$$

where $\hat{\delta} = (\hat{\delta}_0, \ldots, \hat{\delta}_{L+1})$ denotes the OLS estimates, and G denotes the matrix of explanatory variables in (2.7). Then it can be shown that under the null hypothesis (2.6):

$$W_{\mathrm{DQ}} \xrightarrow{d} \chi^2(L+2), \quad \text{as} \quad M \to \infty.$$

Consequently, the null hypothesis is rejected at a given choice of significance level $a \in (0,1)$, if the observed DQ-statistic is in a critical region of the chi-square distribution with $L+2$ degrees of freedom.

Expected Shortfall Regression Test

Second, we consider the expected shortfall regression (ESR) test for evaluating the accuracy of ES forecasts. For this purpose, suppose that $M \in \mathbb{N}$ out-of-sample ES forecasts at level $\alpha \in (0,1)$, denoted by $\hat{\mathrm{ES}}^{\alpha} = (\hat{\mathrm{ES}}_t^{\alpha})_{t=T+1}^{T+M}$, are given. The general testing idea of the ESR test is to regress the realized portfolio returns on the ES forecasts and an intercept term by using the following regression:

$$\mathcal{P}_t = \gamma_1 + \gamma_2 \hat{\mathrm{ES}}_t^{\alpha} + u_t^e, \tag{2.9}$$

where $\gamma_1, \gamma_2 \in \mathbb{R}$ and $\mathrm{ES}^{\alpha}(u_t^e \mid \mathscr{F}_{t-1}) = 0$ almost surely. Since the forecasts $\hat{\mathrm{ES}}_t^{\alpha}$ are generated by using $\mathscr{F}_{t-1}$, this condition on the error term implies that:

$$\mathrm{ES}^{\alpha}(\mathcal{P}_t \mid \mathscr{F}_{t-1}) = \gamma_1 + \gamma_2 \hat{\mathrm{ES}}_t^{\alpha}.$$

The hypotheses of the test read:

$$\mathcal{H}_0 : (\gamma_1, \gamma_2) = (0,1), \qquad \mathcal{H}_A : (\gamma_1, \gamma_2) \neq (0,1). \tag{2.10}$$

Under the null hypothesis, the ES forecasts are correctly specified as it holds that $\hat{\mathrm{ES}}_t^{\alpha} = \mathrm{ES}^{\alpha}(\mathcal{P}_t \mid \mathscr{F}_{t-1})$ almost surely. In general, (2.9) is an example of a linear regression equation for the ES of the form $\mathcal{P}_t = W_t^{\top}\gamma + u_t^e$ for some vector of covariates W_t. As per [6], estimating γ alone by maximum likelihood-type (M) or generalized method of moments (GMM) estimation is infeasible. Therefore, one considers instead the joint regression technique:

$$\mathcal{P}_t = V_t^{\top}\beta + u_t^v, \qquad \mathcal{P}_t = W_t^{\top}\gamma + u_t^e, \tag{2.11}$$

where V_t and W_t are covariate vectors, and where $\mathrm{VaR}^{\alpha}(u_t^v \mid \mathscr{F}_{t-1}) = 0$ and $\mathrm{ES}^{\alpha}(u_t^e \mid \mathscr{F}_{t-1}) = 0$ almost surely. This setup calls for an estimation of the joint parameters (β, γ) which is indeed feasible. We choose to consider the strict ESR test where $V_t = W_t = (1, \hat{\mathrm{ES}}_t^{\alpha})$. The regression system in (2.11) then reads:

$$\mathcal{P}_t = \beta_1 + \beta_2 \hat{\mathrm{ES}}_t^{\alpha} + u_t^v, \qquad \mathcal{P}_t = \gamma_1 + \gamma_2 \hat{\mathrm{ES}}_t^{\alpha} + u_t^e.$$

To test the hypotheses in (2.10), the following Wald-type test statistic is taken into account:

$$\mathrm{W}_{\mathrm{MZ}} = M(\hat{\gamma} - \gamma_0)\hat{\Omega}_{\gamma}^{-1}(\hat{\gamma} - \gamma_0)',$$

where $\gamma_0 = (0,1)$, $\hat{\Omega}_{\gamma}$ is some consistent covariance estimator for the covariance of γ, and $\hat{\gamma}$ denotes the M-estimate. We will not go any further into the details of the M-estimator or the asymptotic result of the test. For more information, the interested reader is referred to [6].

2.3.2 Model Selection

Frequently, there exist various risk forecasting models that could be suitable, and the objective is then to determine the optimal one. While it can be difficult to pinpoint a single model that significantly outperforms all other models, it is possible to reduce the options to a smaller set of models known as a model confidence set (MCS), which comprises the top-performing model(s) with a predetermined level of confidence.

To construct an MCS, we consider a set $\mathcal{M}^0$ that contains a finite number of risk forecasting copula models indexed by $i = 1, \dots, m_0$. To evaluate the models, we utilize different loss functions and introduce the notion of a loss differential, which is defined as follows:

$$d_{ij,t} = L_{i,t} - L_{j,t}, \quad \text{for all } i, j \in \mathcal{M}^0,$$

where $L_{i,t}$ denotes some loss function for model i at time $t \in [T+1, T+M]$. We will consider two loss functions in this thesis. To construct these, hit variables similar to that in (2.5) are defined as:

$$H_{i,t}^{\alpha} = \mathbb{1}_{\{\mathcal{P}_t \leqslant -\hat{\text{VaR}}_{i,t}^{\alpha}\}}.$$

Here $\hat{\text{VaR}}_{i,t}^{\alpha}$ denotes the VaR forecast at risk level α for model i at time t. The first loss function used is the so-called quantile loss (QL) function at risk level α:

$$\text{QL}_{i,t}^{\alpha} = (\alpha - H_{i,t}^{\alpha})(\mathcal{P}_t + \hat{\text{VaR}}_{i,t}^{\alpha}).$$

Hence, the QL function gives larger weights to returns that violate the VaR estimates. The second is the so-called Fissler Ziegel loss (FZL) function at risk level α:

$$\text{FZL}_{i,t}^{\alpha} = \frac{1}{\alpha \hat{\text{ES}}_{i,t}^{\alpha}} H_{i,t}^{\alpha}(\mathcal{P}_t + \hat{\text{VaR}}_{i,t}^{\alpha}) - \frac{\hat{\text{VaR}}_{i,t}^{\alpha}}{\hat{\text{ES}}_{i,t}^{\alpha}} + \ln(-\hat{\text{ES}}_{i,t}^{\alpha}) - 1.$$

where $\hat{\text{ES}}_{i,t}^{\alpha}$ denotes the ES forecast at risk level α for model i time t. Note that the QL and FZL are non-negative functions. In order to rank models, we evaluate their expected loss and prefer model i over model j when $\mu_{ij} < 0$ where $\mu_{ij} = \mathbb{E}[d_{ij,t}]$. The set of the superior model(s) is defined by:

$$\mathcal{M}^* = \{i \in \mathcal{M}^0 : \mu_{ij} \leqslant 0 \text{ for all } j \in \mathcal{M}^0\}. \tag{2.12}$$

The main goal of the MCS procedure is to identify a subset of $\mathcal{M}^0$ that contains all of $\mathcal{M}^*$ with a specified probability, known as the coverage probability. This is done through a sequence of significance tests, where models that are found to be significantly inferior to other models in $\mathcal{M}^0$ are eliminated. We assume that μ_{ij} is finite and does not depend on t for all $i, j \in \mathcal{M}^0$ for the tests to be valid and reliable. The null hypothesis that is being tested is as follows:

$$\mathcal{H}_{0,\mathcal{M}} : \mu_{ij} = 0, \quad \text{for all } i, j \in \mathcal{M}, \tag{2.13}$$

where $\mathcal{M} \subseteq \mathcal{M}^0$, meaning that under the null hypothesis, models i and j have equal performance for all $i, j \in \mathcal{M}$. As a consequence, the alternative hypothesis reads:

$$\mathcal{H}_{A,\mathcal{M}} : \mu_{ij} \neq 0, \quad \text{for some } i, j \in \mathcal{M}. \tag{2.14}$$

It is worth noting that $\mathcal{H}_{0,\mathcal{M}^*}$ is true by definition of $\mathcal{M}^*$, whereas $\mathcal{H}_{0,\mathcal{M}}$ is false if $\mathcal{M}$ includes elements from both $\mathcal{M}^*$ and its complement $\mathcal{M}^0 \backslash \mathcal{M}^*$.

The MCS procedure is based on two key components: an equivalence test denoted by $\xi_{\mathcal{M}}$, and an elimination rule represented by $e_{\mathcal{M}}$. The equivalence test $\xi_{\mathcal{M}}$ is employed to

test the null hypothesis $\mathcal{H}_{0,\mathcal{M}}$ for any $\mathcal{M} \subseteq \mathcal{M}^0$, whereas the elimination rule $e_{\mathcal{M}}$ identifies the model in $\mathcal{M}$ to be removed if the null hypothesis $\mathcal{H}_{0,\mathcal{M}}$ is rejected. We assign the values $\xi_{\mathcal{M}} = 0$ and $\xi_{\mathcal{M}} = 1$ to indicate the acceptance and rejection of the null hypothesis $\mathcal{H}_{0,\mathcal{M}}$, respectively. The following steps outline the procedure for the MCS procedure:

1. Set $\mathcal{M} = \mathcal{M}^0$.
2. Test $\mathcal{H}_{0,\mathcal{M}}$ using $\xi_{\mathcal{M}}$ at level $a \in (0, 1)$.
3. If $\mathcal{H}_{0,\mathcal{M}}$ is accepted, define $\hat{\mathcal{M}}^*_{1-a} = \mathcal{M}$; otherwise use $e_{\mathcal{M}}$ to eliminate a model from $\mathcal{M}$ and repeat the procedure from step 2.

The collection of surviving model(s), denoted as $\hat{\mathcal{M}}^*_{1-a}$, is referred to as the model confidence set. To create an MCS with the desired coverage probability a, it is important to make suitable choices of $\xi_{\mathcal{M}}$ and $e_{\mathcal{M}}$. However, determining such measures can be challenging. The following result specifies the assumptions that need to be met for $\xi_{\mathcal{M}}$ and $e_{\mathcal{M}}$ to be chosen appropriately.

Theorem 2.1.
For any $\mathcal{M} \subseteq \mathcal{M}^0$ and $a \in (0, 1)$, we assume the following about the equivalence test $\xi_{\mathcal{M}}$ and elimination rule $e_{\mathcal{M}}$:

(a) $\limsup_{M\to\infty} \mathbb{P}(\xi_{\mathcal{M}} = 1 \mid \mathcal{H}_{0,\mathcal{M}}) \leqslant a$, (c) $\lim_{M\to\infty} \mathbb{P}(e_{\mathcal{M}} \in \mathcal{M}^* \mid \mathcal{H}_{A,\mathcal{M}}) = 0$,

(b) $\lim_{M\to\infty} \mathbb{P}(\xi_{\mathcal{M}} = 1 \mid \mathcal{H}_{A,\mathcal{M}}) = 1$,

where $\mathcal{M}^*$, $\mathcal{H}_{0,\mathcal{M}}$, and $\mathcal{H}_{A,\mathcal{M}}$ are given by (2.12), (2.13), and (2.14), respectively. Then, it holds that:

(i) $\liminf_{M\to\infty} \mathbb{P}(\mathcal{M}^* \subseteq \hat{\mathcal{M}}^*_{1-a}) \geqslant 1 - a$, (ii) $\lim_{M\to\infty} \mathbb{P}(i \in \hat{\mathcal{M}}^*_{1-a}) = 0 \ \forall i \notin \mathcal{M}^*$,

where $\hat{\mathcal{M}}^*_{1-a}$ is the model confidence set.

The assumptions about $\xi_{\mathcal{M}}$ in Theorem 2.1 are standard requirements for hypothesis tests. In some cases, there may be several candidate models that perform similarly well, making the model selection problem challenging. However, in other cases, there may be a unique true model that provides the best risk forecasts. Identifying this model with high probability is crucial for making accurate predictions and informed decisions. The following result addresses this scenario.

Corollary 2.2.
For any $\mathcal{M} \subseteq \mathcal{M}^0$, suppose that the assumptions about the equivalence test $\xi_{\mathcal{M}}$ and elimination rule $e_{\mathcal{M}}$ in Theorem 2.1 holds and that $\mathcal{M}^*$, given by (2.12), is a singleton. Then:

$$\lim_{M\to\infty} \mathbb{P}(\mathcal{M}^* = \hat{\mathcal{M}}^*_{1-a}) = 1,$$

where $\hat{\mathcal{M}}^*_{1-a}$ is the model confidence set.

Proof. When $\mathcal{M}^*$ is a singleton, $\mathcal{M}^* = \{i^*\}$, it follows from Theorem 2.1 (ii) that i^* will be the last surviving element with probability approaching one as $M \to \infty$. The result now follows because the last surviving element is never eliminated. ■

This result has significant practical implications, as it ensures that the MCS procedure is a reliable tool for model selection in situations where a unique true model exists. The earlier asymptotic results do not depend on any explicit connection between the equivalence test $\xi_{\mathcal{M}}$ and the elimination rule $e_{\mathcal{M}}$. However, there is a benefit of having a degree of connection between them, called coherency, when the MCS procedure is implemented in finite samples. The result below provides a finite sample version of Theorem 2.1 (i) by assuming a particular coherence between $\xi_{\mathcal{M}}$ and $e_{\mathcal{M}}$.

Theorem 2.3.
For any $\mathcal{M} \subseteq \mathcal{M}^0$, let $\xi_{\mathcal{M}}$ and $e_{\mathcal{M}}$ be an equivalence test and elimination rule, respectively. Assume that $\mathbb{P}(\xi_{\mathcal{M}} = 1, e_{\mathcal{M}} \in \mathcal{M}^*) \leqslant a$, then:

$$\mathbb{P}(\mathcal{M}^* \subseteq \hat{\mathcal{M}}^*_{1-a}) \geqslant 1 - a,$$

where $a \in (0, 1)$, $\mathcal{M}^*$ is given by (2.12), and $\hat{\mathcal{M}}^*_{1-a}$ is the model confidence set.

The assumption $\mathbb{P}(\xi_{\mathcal{M}} = 1, e_{\mathcal{M}} \in \mathcal{M}^*) \leqslant a$ states that a bounds the probability that an element from $\mathcal{M}^*$ is eliminated. In practice, hypothesis tests often rely on asymptotic results that cannot guarantee that this assumption holds in finite samples. Therefore, a definition of coherency that is useful for situations where testing is based on asymptotic distributions is needed.

Definition 2.4 (Coherency)**.**
For any $\mathcal{M} \subseteq \mathcal{M}^0$, there is said to be coherency between the equivalence test $\xi_{\mathcal{M}}$ and elimination rule $e_{\mathcal{M}}$ if:

$$\mathbb{P}(\xi_{\mathcal{M}} = 1, e_{\mathcal{M}} \in \mathcal{M}^*) \leqslant \mathbb{P}(\xi_{\mathcal{M}} = 1 \mid \mathcal{H}_{0,\mathcal{M}}),$$

where $\mathcal{M}^*$ and $\mathcal{H}_{0,\mathcal{M}}$ are given by (2.12) and (2.13), respectively.

The combination of coherency and asymptotic control of the Type I error, expressed as $\limsup_{M\to\infty} \mathbb{P}(\xi_{\mathcal{M}} = 1 \mid \mathcal{H}_{0,\mathcal{M}}) \leqslant a$, leads to an asymptotic version of the assumption stated in Theorem 2.3.

Hence, coherency imposes limitations on the possible test and elimination rule combinations we can utilize, extending beyond the asymptotic conditions in Theorem 2.1 (a)-(c). In fact, coherency serves to reduce the reliance on asymptotic properties and prevents nonsensical combinations of tests and elimination rules that could lead to distorted outcomes in finite samples.

Construction of Equivalence Test and Elimination Rule

We will now consider a specific equivalence test and corresponding elimination rule based on t-statistics, that satisfies Definition 2.4 and the assumptions outlined in Theorem 2.1. To construct the t-statistics, the relative loss statistics are defined as $\bar{d}_{ij} = M^{-1} \sum_{t=T+1}^{T+M} d_{ij,t}$ and $\bar{d}_i = m^{-1} \sum_{j\in\mathcal{M}} \bar{d}_{ij}$, where $M \in \mathbb{N}$ denotes the number of samples and $m \in \mathbb{N}$ the number of models in $\mathcal{M}$. Here $\bar{d}_{ij}$ is the relative loss between model i and j, whereas $\bar{d}_i$ is the loss of model i relative to the average of all models in $\mathcal{M}$. From $\bar{d}_i$ the following statistic is constructed:

$$t_i = \frac{\bar{d}_i}{\left(\widehat{\mathrm{Var}}(\bar{d}_i)\right)^{1/2}}, \quad \text{for } i \in \mathcal{M},$$

where $\widehat{\mathrm{Var}}(\bar{d}_i)$ denotes some consistent estimate of the asymptotic variance of $\bar{d}_i$. These t-statistics have respective null hypotheses $\mathcal{H}_{0,i} : \mu_i = 0$, where $\mu_i = \mathbb{E}[\bar{d}_i]$, which generate

a basis for testing $\mathcal{H}_{0,\mathcal{M}}$. Moreover, $\mathcal{H}_{0,\mathcal{M}}$ is equivalent to $\{\mu_i \leqslant 0, \text{for all } i \in \mathcal{M}\}$, as we are only interested in excluding significantly poor performing models. This formulation of the null hypothesis maps into the test statistic:

$$T_{\mathcal{M}} = \max_{i \in \mathcal{M}} t_i.$$

It should be noted that the statistic $T_{\mathcal{M}}$ follows a non-standard asymptotic distribution, as it depends on nuisance parameters. To combat this issue, a bootstrap method is utilized to estimate the relevant distribution. For more information regarding the bootstrap method, the interested reader is referred to [17].

Next, we consider the construction of the corresponding elimination rule. The natural elimination rule associated with $T_{\mathcal{M}}$ is $e_{\mathcal{M},T} = \arg\max_{i \in \mathcal{M}} t_i$, because a rejection of the null hypothesis is identified as $\mu_i \neq 0$ for $i = e_{\mathcal{M},T}$. In this case, $e_{\mathcal{M},T}$ is the model contributing the most to the large test statistic. In particular, this model has the largest excess loss relative to the average across all models in $\mathcal{M}$ and is therefore eliminated. This combination of equivalence test and elimination rule meets the criteria outlined in Definition 2.4 and fulfills the assumptions stated in Theorem 2.1. For a proof of this, the interested reader is again refereed to [17].

Part II
Application

3 Data Presentation

In this applications part, we seek to forecast value-at-risk and expected shortfall, evaluate these forecasts, and determine the most suitable copula model(s). The forecasting is achieved by utilizing the presented ARMA-GARCH and copula models as described in the procedure in Section 2.2. The accuracy of the obtained forecasts is evaluated by the dynamic quantile and expected shortfall regression tests and the top-performing model(s) is identified by constructing model confidence sets.

This research is conducted for an equally weighted portfolio composed of the stocks of NextEra Energy (NEE) and British Petroleum (BP). NextEra Energy is a leading clean energy company that generates and distributes electricity from renewable sources such as wind and solar, whereas British Petroleum is a multinational oil and gas company. We consider the daily closing prices of each stock. The full samples for the two stocks start on the 1st of January 2015 and end on the 25th of April 2023. These two full samples are divided into in-sample and out-of-sample periods where the 31st of December 2020 serves as the last in-sample date. Therefore, the in-sample and out-of-sample periods consist of $T = 1510$ and $M = 580$ observations, respectively.

To justify the marginal distribution modelling choice of ARMA-GARCH type models, we have to examine the stylized facts, briefly noted in Section 1.7, for the full sample of both the NextEra Energy and British Petroleum data. To do this, an exploratory data analysis is carried out in the following. First, plots of the daily closing prices of the two stocks are considered.

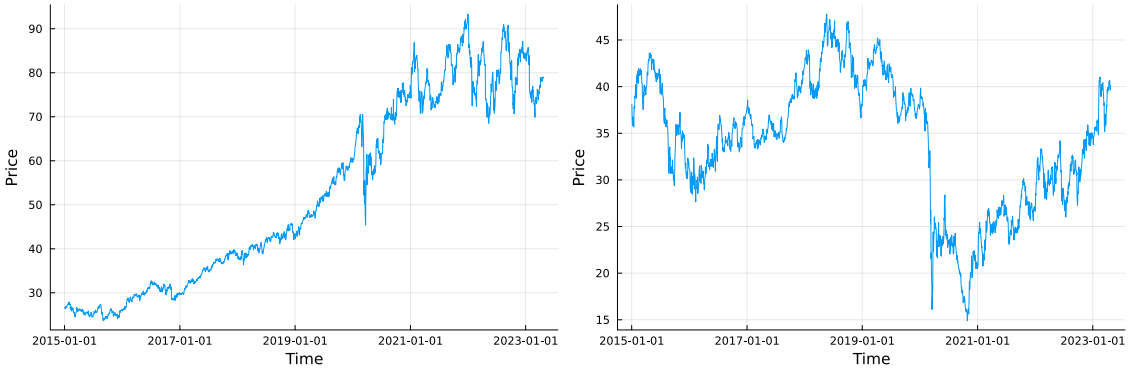


Figure 3.1: Daily closing prices of NEE (left) and BP (right) from 2015-01-01 to 2023-04-25.

The periods of both positive and negative trends in the two series indicate that they are both non-stationary. An augmented Dickey-Fuller test is used to more thoroughly determine whether this is the case. The null hypothesis of the augmented Dickey-Fuller test is that the given series is non-stationary. The p-values of the test for all lags $l \in \{1, \ldots, 10\}$ for both series are shown in Figure 3.2. For any tested lag, it is obvious that the null hypothesis of

non-stationarity cannot be rejected, indicating that both series are non-stationary.

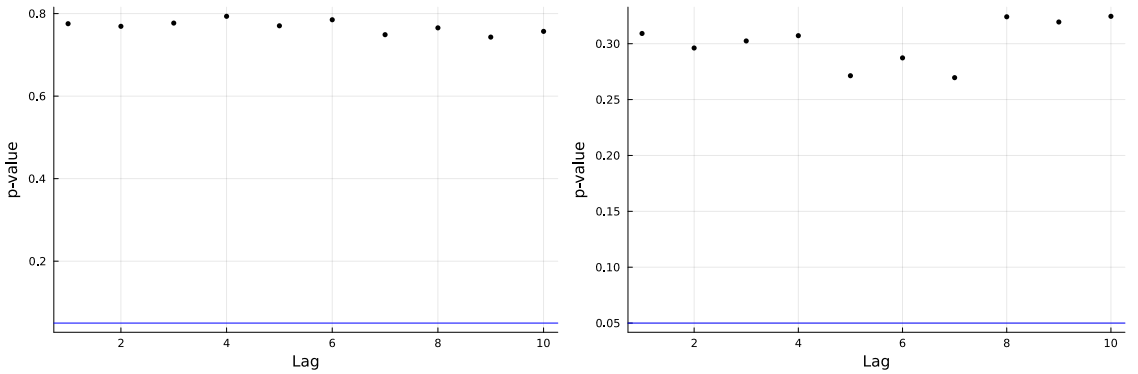


Figure 3.2: p-values for the ADF test for the daily closing prices of NEE (left) and BP (right).

Next, we consider the daily log return series of the full samples defined by (2.3). These are shown in Figure 3.3.

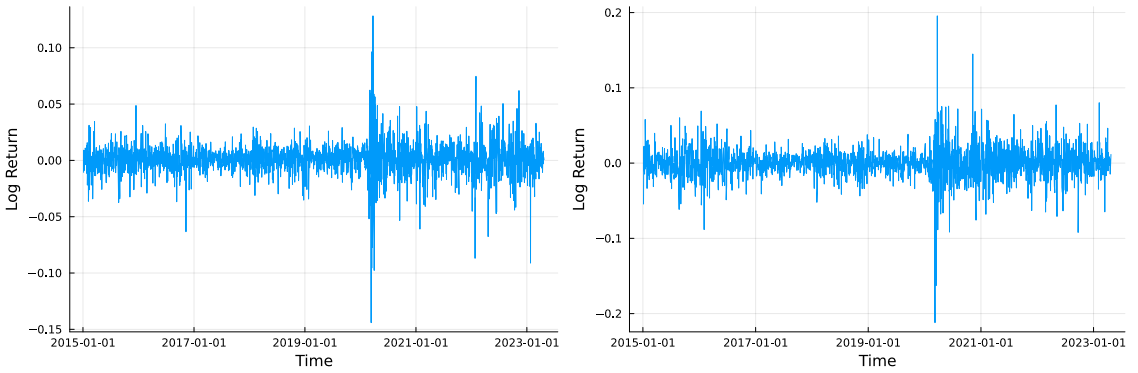


Figure 3.3: Log returns of NEE (left) and BP (right) from 2015-01-01 to 2023-04-25.

There are periods of low volatility and periods of high volatility for both stocks, which indicates that volatility changes over time and has a tendency to persist. The COVID-19 pandemic outbreak at the beginning of 2020 is clearly accompanied by a period of high volatility. The Russian invasion of Ukraine in 2022 also caused a spike in volatility for NEE, whereas the volatility of BP seems to be less affected.

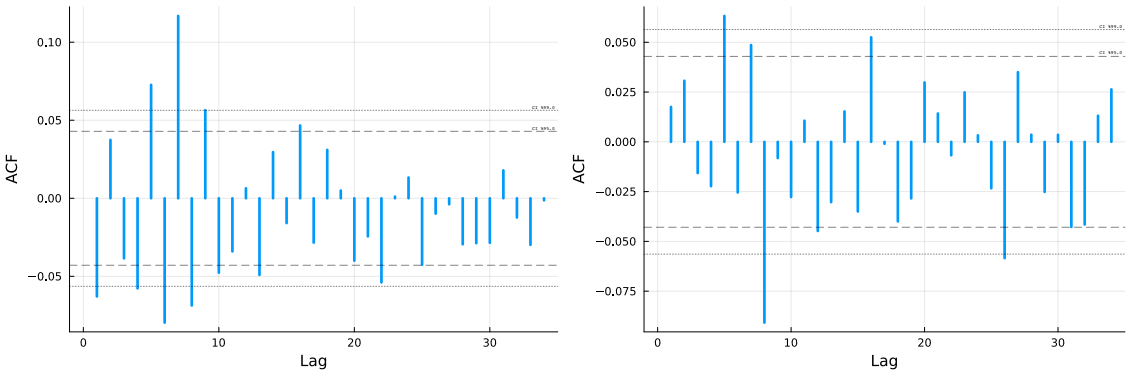


Figure 3.4: Correlogram of the log return series for NEE (left) and BP (right).

To ensure that the series are now stationary, we run the ADF test again. By doing this, we get that the p-values for all lags $l \in \{1, \ldots, 10\}$ are less than 0.05 for both stocks, indicating that they are both stationary at a 5% significance level. Furthermore, we consider whether the correlograms of the two return series show any autocorrelation. Figure 3.4 shows the correlograms along with the confidence bands calculated as $\pm\Phi^{-1}(1-\alpha^2)/\sqrt{T+M}$, where a significance level of $\alpha = 0.01$ and $\alpha = 0.05$ is chosen and Φ^{-1} signifies the quantile function of the standard normal distribution. For both stocks, there are several correlations rising above the confidence bands suggesting that there might be some autocorrelation in the log returns. In order to examine this more explicitly, the null hypothesis that there is no autocorrelation is tested using the Ljung-Box test. The p-values for all lags $l \in \{1, \ldots, 10\}$ are shown in Figure 3.5.

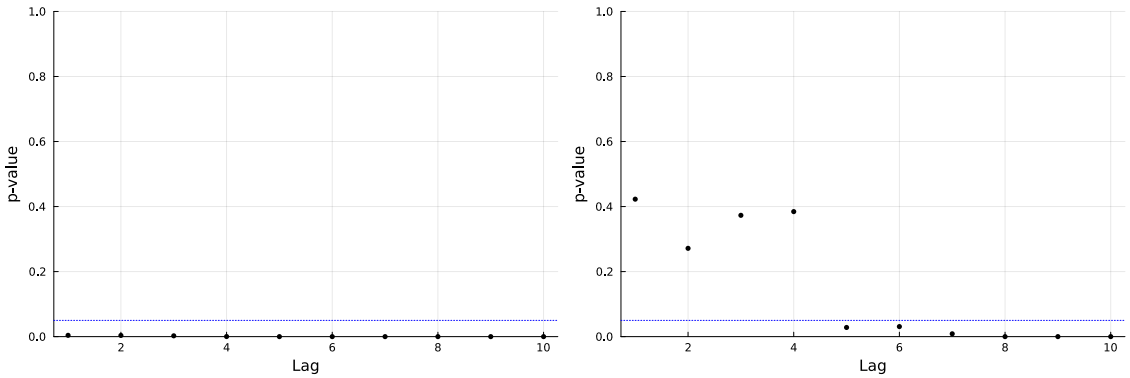


Figure 3.5: p-values for the Ljung-Box test for log return series of NEE (left) and BP (right).

For NEE, the plot shows that we reject the null hypothesis at a 5% significance level for all lags since the p-values are all below 0.05. This suggests that autocorrelation exists in this series. For BP, the p-values are above 0.05 for the first four lags and below for the remaining lags, which means that we cannot rule out the possibility that there is some autocorrelation present in this series as well. In light of the fact that ARMA models can be used to account for autocorrelation in stationary time series, they seem to be a suitable modelling choice for both the log return series of NEE and BP.

Subsequently, we examine whether the empirical distributions of both log return series are leptokurtic. To do this, the empirical histograms of the log returns of the two stocks are plotted in Figure 3.6 along with the normal density with mean and standard deviation given by the series.

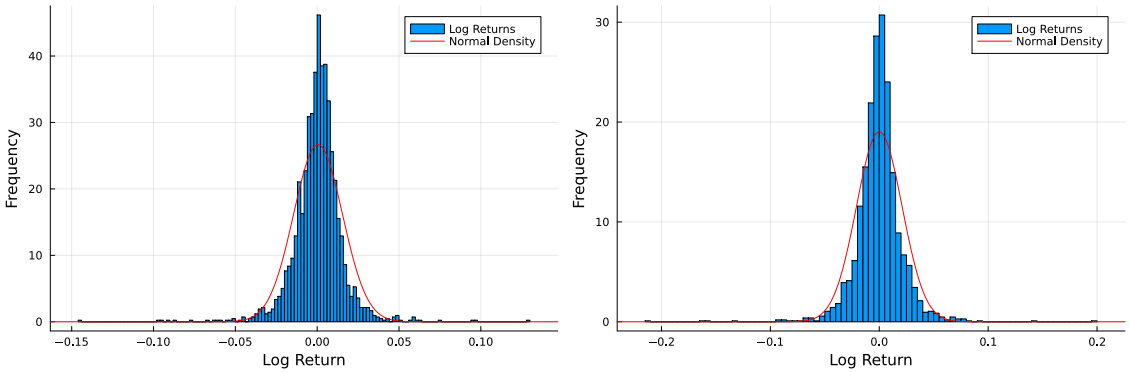


Figure 3.6: Histograms of log returns of NEE (left) and BP (right) with theoretical normal densities.

It is clear that both log returns' empirical distributions have higher peaks than their respective normal distributions. Additionally, it seems that the empirical distributions have heavier tails. To support this second claim, we consider QQ-plots in which the sample quantiles are compared against the quantiles of a normal distribution in Figure 3.7.

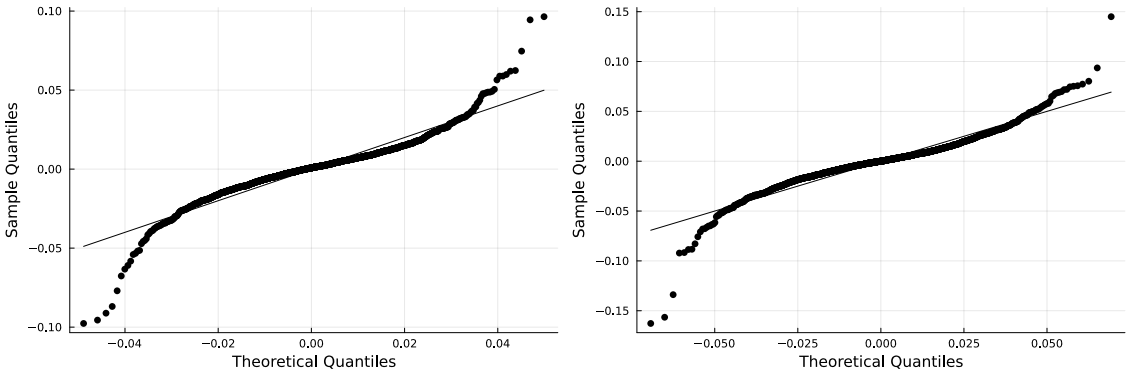


Figure 3.7: QQ-plots of sample quantiles against normal quantiles for log returns of NEE (left) and BP (right).

It is evident that both stocks' log returns have heavier tails than their respective normal distributions. When we calculate the sample kurtosis of the two stocks, we find values of 12.7 and 14.7 for NEE and BP, respectively. All of this points to a leptokurtic empirical distribution for both log return series.

Next, we examine the existence of leverage effects in the return series. To do this, we consider whether the absolute log returns $|y_t|$ are correlated with $y^+_{t-h} = \max\{y_{t-h}, 0\}$ and $y^-_{t-h} = \max\{-y_{t-h}, 0\}$ where $h \in \{1, \ldots, 30\}$. The results of this are seen in Figure 3.8.

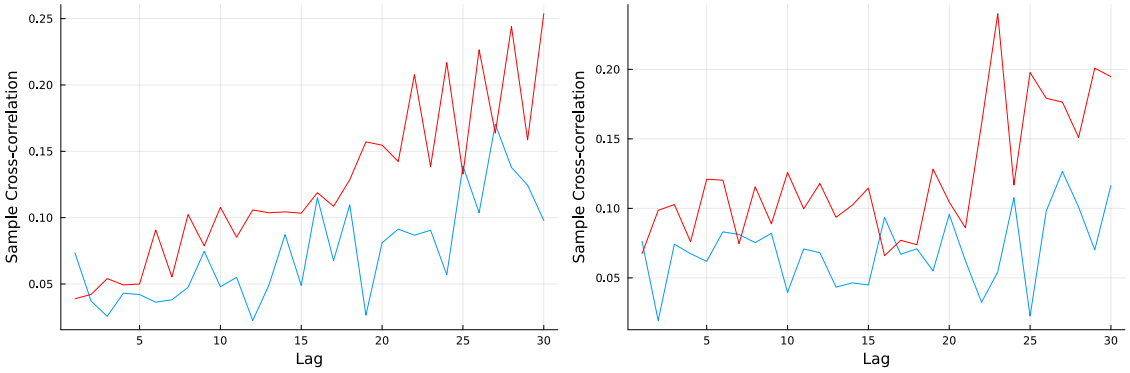


Figure 3.8: Sample cross-correlations of NEE (left) and BP (right) between $|y_t|$ and y^+_{t-h} (blue), and $|y_t|$ and y^-_{t-h} (red).

For both stocks, we see that there is a larger correlation between $|y_t|$ and y^-_{t-h} than between $|y_t|$ and y^+_{t-h} for almost all lags. This suggests that the leverage effect exists for both log return series because negative log returns tend to increase volatility by a greater amount than positive log returns of the same magnitude.

Finally, an ARCH-LM test is used to determine whether heteroscedasticity is present, by fitting an ARCH(l) model to the log return series and testing for joint nullity of the ARCH parameters. The null hypothesis of the test is that the innovations are homoscedastic, while the alternative hypothesis is that they exhibit heteroscedasticity. When performing the

test, we get p-values approximately equal zero for any specified lag $l \in \{1, \ldots, 30\}$. Hence, the test indicates that the innovations exhibit heteroskedasticity. This feature, along with the fact that both series' leverage effects exist and that their distributions are leptokurtic, suggests that the GARCH framework is a suitable option for modelling the log returns of NEE and BP. The general conclusion of this chapter is hence that the ARMA-GARCH type models seem to be suitable modelling options for the log return series of NEE and BP.

4 In-Sample Model Diagnostics

The objective of this chapter is to examine the empirical results from modelling the in-sample parts of the log returns for the stocks of NextEra Energy and British Petroleum using the discussed copula and marginal modelling frameworks. Several packages developed in the `Julia` programming language have been used to obtain the results. Therefore, we will first provide a brief overview of the key packages that have been used before presenting the in-sample model diagnostics.

4.1 Introduction to the Main Packages Used in `Julia`

The risk forecasting procedure described in Section 2.2 is first and foremost performed by use of the `BivariateCopulas` package in `Julia`. More specifically, the package is used to estimate the copulas and to sample from them. As described in Section 1.5, Bayesian inference is used for the copula estimation. To implement this estimation procedure in `Julia`, we take into consideration the `Turing` package in addition to the `BivariateCopulas` package. The usage of these packages for copula estimation is demonstrated in the example below.

```
@model function fit_Clayton_copula(W; epsilon = 1e-6)
    gamma ~ Uniform(epsilon,1-epsilon)
    theta = -log(gamma)
    for i in 1:length(W[1,:])
        W[:,i] ~ Clayton(theta)
    end
end

Clayton_chain = sample(fit_Clayton_copula(W), NUTS(), 1000)
theta =  -log.(vec(Clayton_chain[:gamma]))
Clayton_par = mean(theta)
```

Here, a bivariate Clayton copula is fitted to two uniform series contained in the matrix `W`. In doing this, 1000 samples are obtained from the posterior using the No-U-Turn Sampler (NUTS), which is a Hamiltonian MCMC method. For more information on this specific sampler, we refer to [9]. The prior is represented by the uniform distribution, and a parameter space transformation is performed using the log function. The mean of the transformed chain provides the parameter estimate for the copula. The following example shows how to sample from the copula using the `BivariateCopulas` package.

```
C = Clayton(Clayton_par)
rand(C,100000)
```

In this case, 100000 samples are collected from the Clayton copula using the previously discovered parameter. In addition to being used to estimate copulas and sample from them in the risk forecasting procedure, the `BivariateCopulas` package can be used for a variety of other tasks, e.g., evaluating the copula density given by (1.3). To do this, the function

`pdf(C,[u,v])` is used, where `C` is some specified copula and `[u,v]` is some evaluation point. The figures in Section 1.3 have copula densities that have been evaluated using this function.

The marginal modelling procedure, which is carried out using the ARMA-GARCH framework, is another crucial step in the risk forecasting procedure. We use the `ARCHModels` package to implement this in `Julia`. The usage of this package is demonstrated in the example below.

```
fit = selectmodel(GARCH, returns; meanspec=ARMA, criterion=aic, maxlags=3,
    dist=StdT)
residuals(fit)
predict(fit, :return)
predict(fit, :volatility)
```

In this example, the function `selectmodel` fits and chooses the optimal ARMA-sGARCH model based on AIC for a certain return series where $p, q \in \{0, \ldots, 3\}$[1] and the innovations are t-distributed. Additionally, the `residuals` function returns the innovation series. Using `selectmodel`, we can consider all the model specifications considered in Section 1.7. As a result, these functions provide us with the innovation series and the corresponding distribution function, both of which are necessary for the copula estimation. The `predict` function forecasts the means and variances of one-step-ahead return values based on `fit`, which is also used in the risk forecasting procedure.

4.2 Marginal Modelling

To model the log return series of NextEra Energy and British Petroleum, we consider the ARMA-GARCH framework, which was the subject of Section 1.7. We limit ourselves to the case where $p, q \in \{0, \ldots, 3\}$ for all ARMA-GARCH type models. In doing so, we take into account a total of 2304 models per series. These models are fitted to the two series using maximum likelihood estimation. We utilize the Akaike information criterion (AIC) to select one model out of all the model specifications for each series. If the AIC cannot be determined, the corresponding model is disregarded.

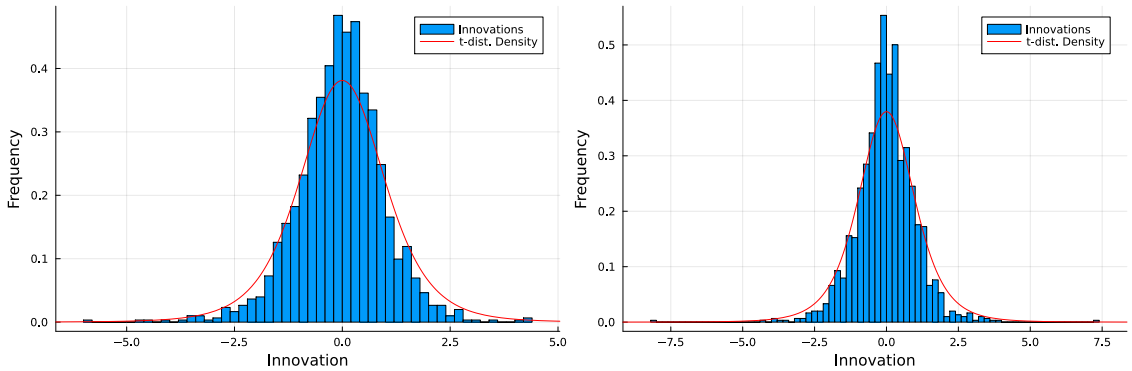


Figure 4.1: Histograms of innovations for NEE (left) and BP (right) along with estimated t-distribution densities.

We discover that the ARMA$(3,3)$-sGARCH$(1,1)$ model with t-distributed innovations is the most suitable model for the log return series of NEE, whereas the ARMA$(2,2)$-tGARCH$(1,1)$ model with t-distributed innovations is the most suitable model for the log

[1] Note that p and q describes the order of both the ARMA and GARCH models. Hence we restrict both the ARMA and GARCH models to have a maximum of 3 lags in this case.

return series of BP. The parameters for these models are seen in Table B.1. In Figure 4.1, the histograms of the two innovation series are shown, plotted against their respective estimated t-distributions. Evidently, the t-distributions do not fit the respective innovation series well. In particular, the theoretical distributions are less leptokurtic and have heavier tails than those of the innovation series. Finding suitable marginal models is critical when using copulas, as wrongly specified distributions of the innovations may lead to misspecification of the copula model when transforming the innovations into uniform realizations by the probability integral transform.

In light of these observations, we elect to also investigate the ARMA-GARCH type models with GED innovations. Specifically, this is done because the GED includes a shape parameter that allows for a more leptokurtic distribution. In this case, the number of possible model specifications amounts to a total of 768 per series. Based once again on AIC, we choose the most suitable ARMA-GARCH type models with GED innovations. With this alternative approach, we find that the ARMA$(3,3)$-sGARCH$(1,1)$ model is the most suitable model for NEE, and that the ARMA$(2,2)$-tGARCH$(1,2)$ model is the most suitable model for BP. For the remainder of this section, the results for the GED innovation models will be presented, whereas the results for the t-distributed innovation models are found in Appendix B. The parameters of the GED innovation models are shown in Table 4.1.

	NextEra Energy	British Petroleum
	ARMA$(3,3)$-sGARCH$(1,1)$	ARMA$(2,2)$-tGARCH$(1,2)$
	Mean equation parameters	
c	-4.1e-07 (3.9e-07)	-0.0002 (0.0005)
ϕ_1	0.6992 (0.1286)	0.5062 (0.0346)
ϕ_2	0.7108 (0.1138)	-0.9672 (0.0347)
ϕ_3	-0.4178 (0.0555)	-
κ_1	-0.7771 (0.1240)	-0.4909 (0.0367)
κ_2	-0.7386 (0.1044)	0.9551 (0.0423)
κ_3	0.5071 (0.0351)	-
	Volatility parameters	
ω	4.7e-06 (1.6e-06)	4.6e-06 (2.1e-06)
α_1	0.0968 (0.0221)	0.0409 (0.0217)
β_1	0.8692 (0.0270)	0.5759 (0.3182)
β_2	-	0.3195 (0.2924)
γ_1	-	0.1017 (0.0394)
	Distribution parameters	
ν	1.3013 (0.0690)	1.1968 (0.0769)

Table 4.1: Parameter estimates and standard errors of the most suitable models with GED innovations for the log return series of NEE and BP.

We will not delve into the particular parameter estimates. We only mention that the AIC suggests a model that takes into account the leverage effect for the log returns of BP which is reflected by the γ_1 parameter. Additionally, the GED parameters are somewhat similar since they only differ by approximately 0.1.

Next, the histograms of the innovation series associated with the models in Table 4.1 are plotted against their respective estimated GEDs in Figure 4.2. It is evident that compared to the estimated t-distributions in Figure 4.1, the estimated GEDs seems to provide a better fit to their respective innovation series.

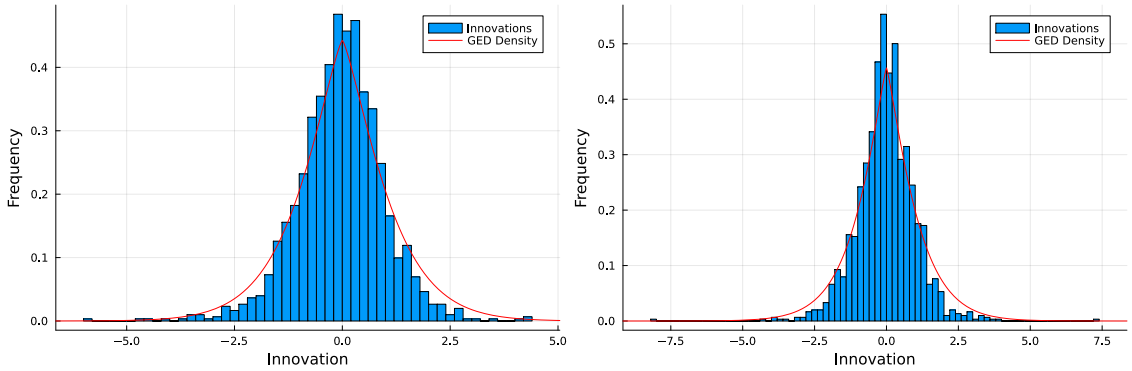


Figure 4.2: Histograms of innovations for NEE (left) and BP (right) along with estimated GED densities.

We complement these findings with Ljung-Box tests of the innovations and squared innovations in order to ensure that no autocorrelation and heteroskedasticity are left in the GED innovations. The p-values of the tests are seen in Figure 4.3.

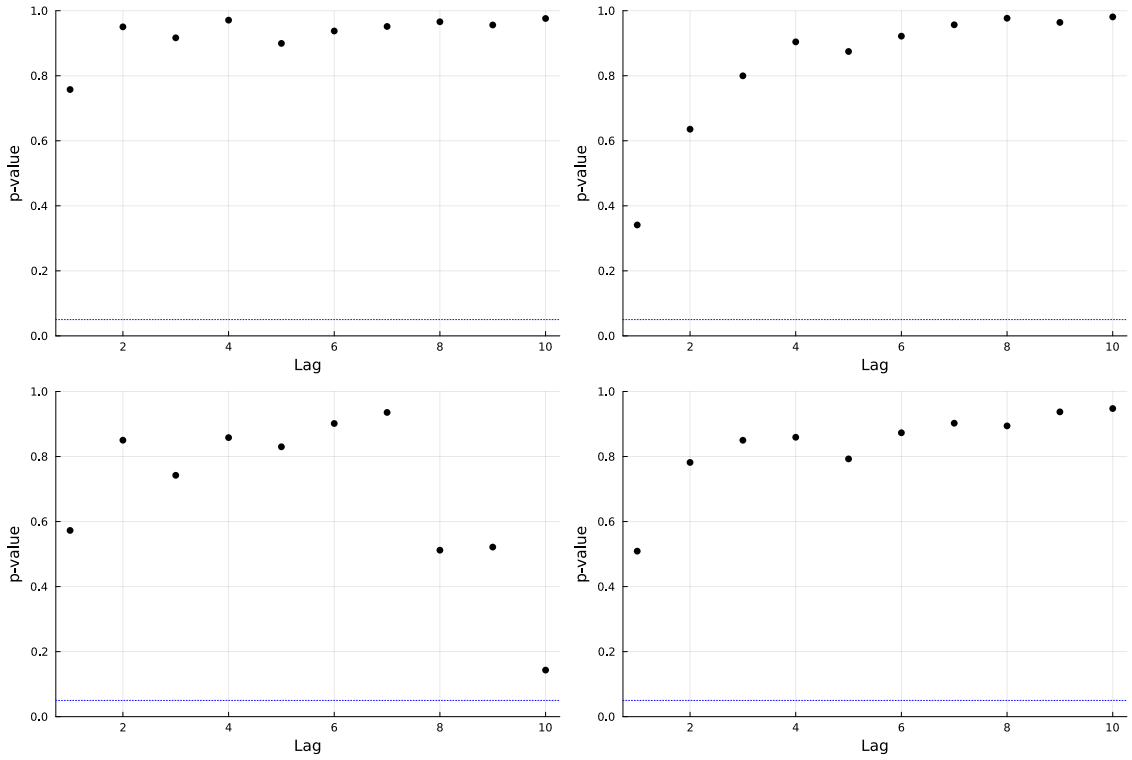


Figure 4.3: Ljung-Box tests of the GED innovations (top) and squared innovations (bottom) for NEE (left) and BP (right).

Evidently, there is no serial autocorrelation present in the innovations and squared innovations for both innovation series, as the null hypothesis of no autocorrelation cannot be rejected. The same conclusions are drawn for the t-distributed innovations from Figure B.1. The lack of autocorrelation in the squared innovations implies homoscedasticity. This is tested more formally by an ARCH-LM test on each innovation series which produces p-values above 0.05 for any specified lag $l \in \{1, \ldots, 30\}$ for both the GED and the t-distributed innovations. These results indicate that the innovation series exhibit no autocorrelation

and heteroscedasticity.

Lastly, we note that since we condition on different information sets when modelling the marginals, it should be tested whether lagged values of one series affect the other as described in Section 1.7. This is done by first fitting an ARMAX$(3,3)$ and ARMAX$(2,2)$ model to the log returns of NEE and BP, respectively. We test the significance of cross-equation effects by performing a Wald test for the joint nullity of the additional explanatory variables. The parameters of the ARMAX models are fixed as in Table 4.1. The tests yield p-values of 0.43 and 0.92 for NEE and BP, respectively, indicating that there are no significant cross-equation effects. The same conclusion is drawn when considering the ARMA parameters for the t-distributed innovations shown in Table B.1. Thus, it is reasonable to assume that the marginals can be modelled by univariate ARMA-GARCH type models. Note that, as described in Section 1.7 the proper test would be to fit an ARMAX-GARCHX to each log return series and test for the joint nullity of all additional explanatory variables, however, due to time constraints this was not possible.

Nonetheless, an overall conclusion is that, regardless of whether we are considering GED or t-distributed innovations for the univariate ARMA-GARCH framework, the models provide reasonable marginal modelling choices.

4.3 Copula Modelling

The goal of this section is to use Bayesian inference to estimate copula parameters and to do a simulation study where we compare samples of the estimated copula with the innovation series. Since we have no initial knowledge of the parameter estimates, we use a uniform prior for all copula parameters. It should be noted that, as the different copulas have different parameter spaces, these uniform prior realizations should be transformed by some suitable transformations. E.g. for the Clayton copula, the log function is used as shown in Section 4.1. The No-U-Turn Sampler is used to generate posterior samples of the copula parameters given the innovations and the prior distribution. The sampler is run with 1000 iterations and the samples are used to compute posterior estimates of the means and variances alongside the credible intervals of the parameters.

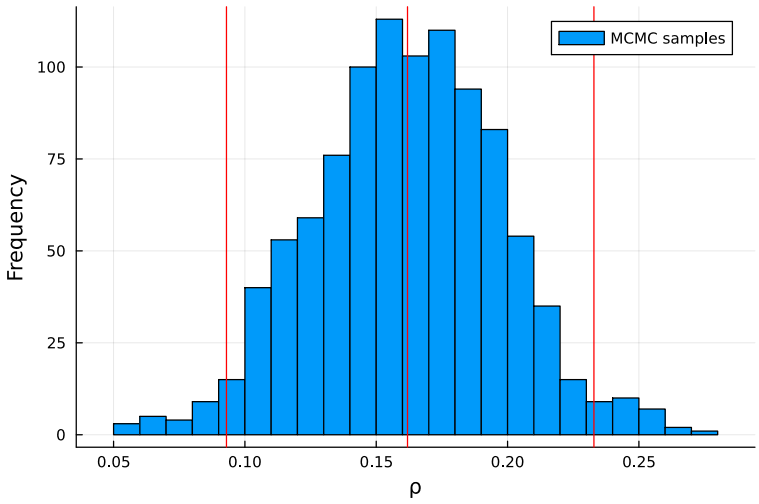


Figure 4.4: MCMC samples of posterior distribution, credible intervals, and estimate of ρ for the Gaussian copula with GED innovations.

The mean estimate and credible intervals for the parameter of the Gaussian copula with GED innovations are shown in Figure 4.4. The histogram shows the distribution of the MCMC samples, the outer lines represent the credible intervals, and the middle line depicts the mean of the distribution. This procedure is performed for all the specified copulas, and the resulting histograms are found in Figure B.2 and B.3 for GED and t-distributed innovations, respectively. It should be noted that parameter estimation of the t- and BB7 copulas are performed by maximum likelihood estimation, as the `BivariateCopulas` package does not support Bayesian estimation of these copulas as of writing this thesis.

Table 4.2 shows a comprehensive statistical summary for all the specified copulas with GED innovations. The Bayesian and maximum likelihood estimation summary is shown, along with the p-values of the Cramér-von Mises goodness-of-fit test[2], and the tail dependencies. When inspecting the correlation parameters of the Gaussian and t-copula, it is evident that the correlation between the two innovation series is estimated to be low. Although these two copulas somewhat agree on the level of correlation, the t-copula is estimated to have approximately 7 degrees of freedom, which is too low to represent a Gaussian copula. The estimate of degrees of freedom is supported by the Cramér-von Mises test, which could suggest that higher degrees of freedom provide a worse fit since the t-copula converges to the Gaussian copula for $\nu \to \infty$.

When inspecting the p-values of the Cramér-von Mises test, we note that the Frank, Joe, Ali-Mikhail-Haq, and t-copulas are all well-specified at a 5% significance level, according to the null hypothesis of the Cramér-von Mises test described in Section 1.6. Hence, any of these four copula models exhibit a reasonable fit with the marginal modelling choice of GED innovations. Out of these four well-specified copulas, only two exhibit tail dependence, namely, the Joe and t-copulas. This indicates there is no clear connection between the well-specified copulas with GED innovations and copulas with tail dependence.

Copula		Parameter estimate	Variance	Credible interval	Cramér-von Mises p-value	Lower tail dependence	Upper tail dependence
Gaussian	$\hat{\rho}$	0.1619	0.0013	[0.0930, 0.2329]	0.0058	0.0000	0.0000
t	$\hat{\rho}$	0.1367	-	-	0.1132	0.0411	0.0411
	$\hat{\nu}$	6.8450	-	-			
Frank	$\hat{\theta}$	0.6887	0.0391	[0.2882, 1.0641]	0.0573	0.0000	0.0000
Gumbel	$\hat{\theta}$	1.1228	0.0007	[1.0713, 1.1819]	0.0305	0.0000	0.1461
Clayton	$\hat{\theta}$	0.1637	0.0020	[0.0802, 0.2550]	0.0115	0.0145	0.0000
Joe	$\hat{\theta}$	1.1758	0.0014	[1.1023, 1.2528]	0.2398	0.0000	0.1969
BB7	$\hat{\theta}$	1.1404	-	-	-	0.0020	0.1636
	$\hat{\delta}$	0.1112	-	-			
Ali-Mikhail-Haq	$\hat{\theta}$	0.1467	0.0008	[0.0895, 0.2023]	0.1983	0.0000	0.0000

Table 4.2: Estimation summary for all specified copulas with GED innovations.

In Table 4.3 the estimation summary for all specified copulas with t-distributed innovations is shown. When inspecting the p-values of the Cramér-von Mises test, we note that as in Table 4.2, the Frank, Joe, Ali-Mikhail-Haq, and t-copulas are well-specified at a 5% significance level. Therefore, any of these four copula models exhibit a reasonable fit with the marginal modeling choice of t-distributed innovations as well. The large p-value of the Ali-Mikhail-Haq is especially noticeable, as it indicates that this copula is particularly well specified. These results indicate that there is also no clear connection between the well-specified copulas with t-distributed innovations and copulas with tail dependence.

[2]Note that the p-value of the Cramér-von Mises test is missing for the BB7 copula due to incompatibility with the package used. Due to time constraints, this test was not implemented manually.

Copula		Parameter estimate	Variance	Credible interval	Cramér-von Mises p-value	Lower tail dependence	Upper tail dependence
Gaussian	$\hat{\rho}$	0.1805	0.0015	[0.1063, 0.2570]	0.0012	0.0000	0.0000
t	$\hat{\rho}$	0.1435	-	-	0.1735	0.0783	0.0783
	$\hat{\nu}$	4.8961	-	-			
Frank	$\hat{\theta}$	0.7758	0.0453	[0.3662, 1.2256]	0.1480	0.0000	0.0000
Gumbel	$\hat{\theta}$	1.1443	0.0009	[1.0850, 1.2020]	0.0009	0.0000	0.1674
Clayton	$\hat{\theta}$	0.2011	0.0022	[0.1174, 0.3007]	0.0257	0.0319	0.0000
Joe	$\hat{\theta}$	1.1940	0.0017	[0.0017, 1.1203]	0.1909	0.0000	0.1969
BB7	$\hat{\theta}$	1.1599	-	-	-	0.0068	0.1823
	$\hat{\delta}$	0.1390	-	-			
Ali-Mikhail-Haq	$\hat{\theta}$	0.1639	0.0009	[0.1061, 0.2177]	0.7212	0.0000	0.0000

Table 4.3: Estimation summary for all specified copulas with t-distributed innovations.

To visually compare the goodness-of-fit for the copulas with GED innovations, we conduct a simulation study which is shown in Figure 4.5. Given the parameter estimations in Table 4.2, we create 10000 samples from each copula, transform them via Proposition 1.4, and compare them to the 1510 in-sample innovations. Since the number of samples is larger than the number of innovations, it seems as if the variations in the samples are larger than in the innovation series. When considering Figure 4.5, the dispersion structure seems to be reasonable for all copulas, however certain copulas such as the Gumbel and BB7 copulas seem to overestimate the upper tail dependence. Hence, although the null hypothesis of the Cramér-von Mises test is rejected for the Gaussian and Clayton copulas, these copulas seem to reproduce the innovations reasonably well.

A similar simulation study is conducted for the copula models with t-distributed innovations, which is shown in Figure 4.6. For the Joe, Gumbel, and Ali-Mikhail-Haq copulas, there are a few outliers. Excluding these outliers, the Joe and Ali-Mikhail-Haq copulas seem to fit the t-distributed innovations well. In fact, the only models which seem to have poor fits are again the Gumbel and BB7 copulas. As a result, it is clear once again that even while the null hypothesis of the Cramér-von Mises test is rejected for the Gaussian and Clayton copulas, these copulas appear to rather well reproduce the innovations. Furthermore, the larger Cramér-von Mises p-value associated with the Ali-Mikhail-Haq copula is supported by the simulation study, as it seems to best resemble the dispersion structure of the innovations.

As was discussed in Section 4.2, the estimated t-distributions seem to poorly fit the associated innovations which may lead to misspecification of the copulas. This is reflected in the simulation study, as certain copulas seem to overestimate the extreme observations, which could be a sign of misspecification. However, when inspecting the Cramér-von Mises test results, there is no clear difference between the copula models with GED and t-distributed innovations in terms of goodness-of-fit. Therefore, although the GED seems to provide a better fit to its innovations than the t-distribution does, there seems to be no significant difference in the copula modelling performance on the in-sample data.

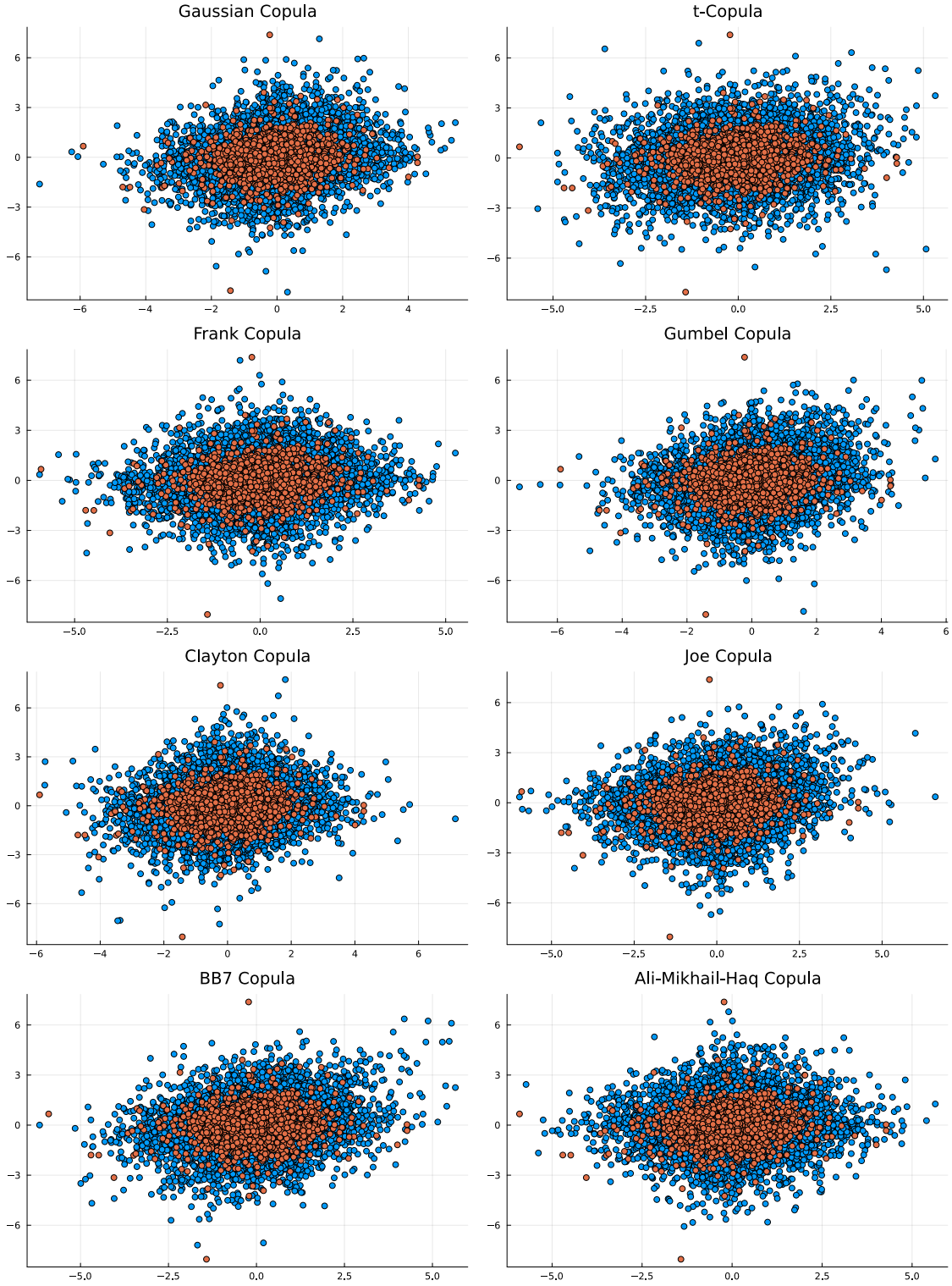


Figure 4.5: Scatter plots of copula samples (blue) and GED innovations (orange).

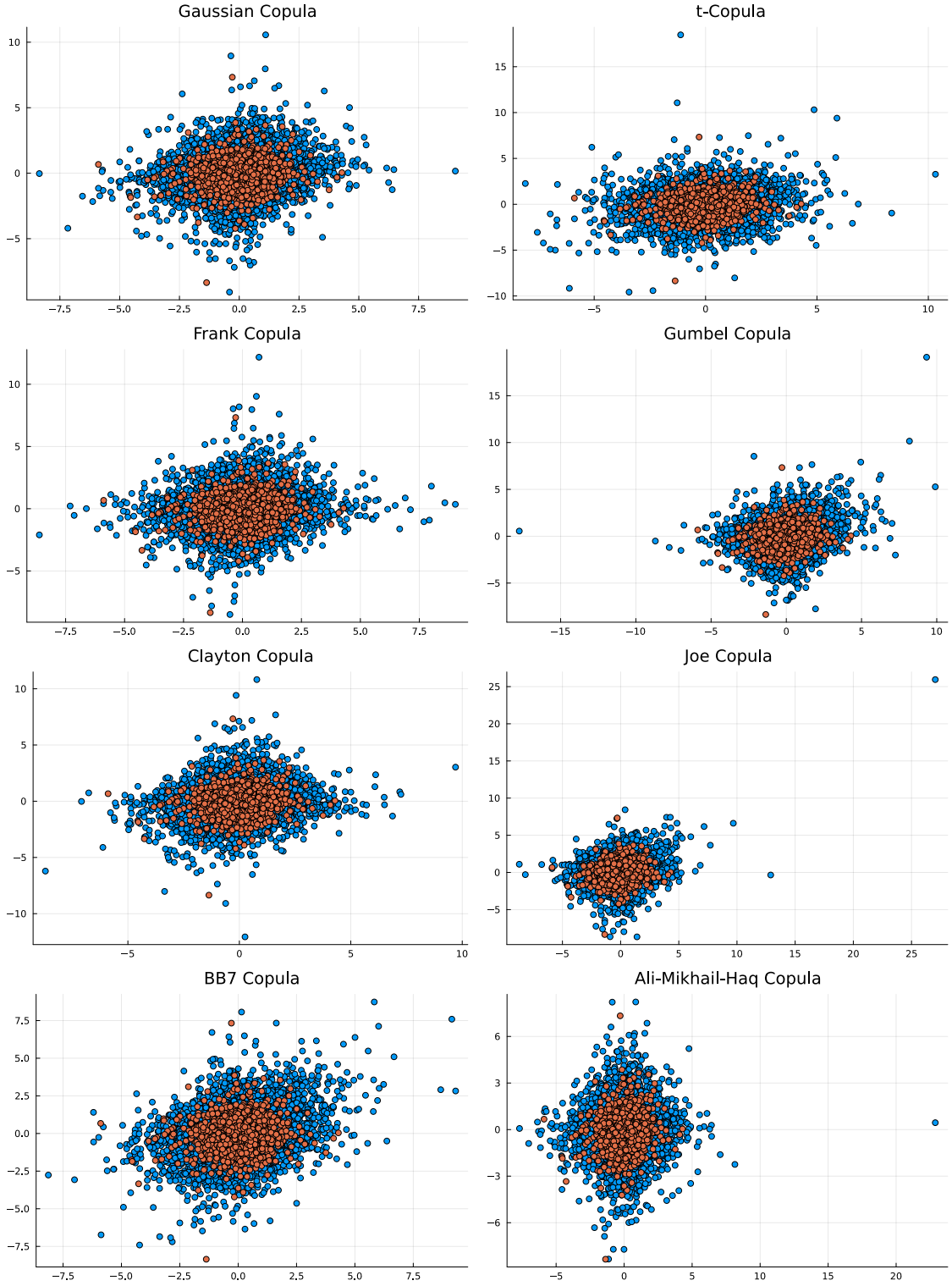


Figure 4.6: Scatter plots of copula samples (blue) and t-distributed innovations (orange).

5 Out-of-Sample Risk Forecasting

In this chapter, we seek to perform out-of-sample forecasting of value-at-risk and expected shortfall, evaluate these forecasts, and determine the optimal risk forecasting copula model(s). As mentioned, the risk forecasts are obtained for the equally weighted portfolio consisting of the stocks of NextEra Energy and British Petroleum. In order to perform risk forecasting, the presented ARMA-GARCH and copula models are utilized as described in the procedure in Section 2.2. We simulate $N = 100000$ scenarios of the portfolio return series and sample 1000 times from the posterior, similar to the examples in Section 4.1. To evaluate whether the specified copula models provide reasonable forecasts, the obtained forecasts are evaluated by the DQ and ESR tests described in Section 2.3.1. To determine which model specification(s) produces the most accurate risk forecasts, model confidence sets, introduced in Section 2.3.2, are constructed. Again, we will both consider the specified copula models with GED innovations and the models where the distributions of the innovation series are determined entirely by AIC.

First, we consider plots of the portfolio returns alongside the one-day-ahead VaR and ES forecasts for the t-copula with GED innovations at risk level $\alpha = 0.01$ and $\alpha = 0.05$ in Figure 5.1. These plots are included to provide a visual representation of the VaR and ES forecasts. The negative values of the VaR forecasts have been used in order to compare them to the returns and ES forecasts. By definition of VaR, the return series should subceed the negative forecasts of $\text{VaR}^{0.01}$ and $\text{VaR}^{0.05}$ for approximately 1% and 5% of the out-of-sample dates, respectively, if the VaR forecasts are properly determined. Although it can be challenging to see visually from the graphs below, this appears to approximately be the case. The ES is defined as the expected return below the negative VaR level. Therefore, it makes sense that the ES forecasts are lower than the negative VaR forecasts. Similar plots for the remaining copula models with GED innovations are found in Figure B.4-B.5. All of the plots resemble those in Figure 5.1 quite closely.

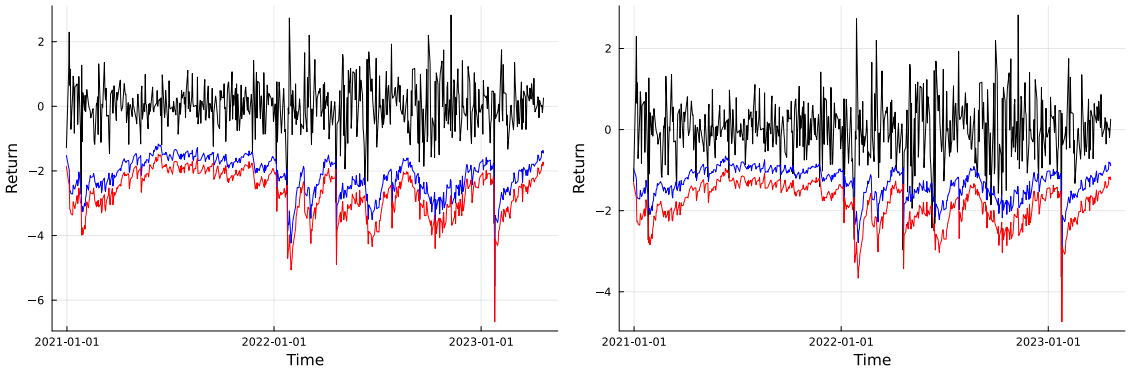


Figure 5.1: VaR (blue) and ES (red) forecasts at risk level $\alpha = 0.01$ (left) and $\alpha = 0.05$ (right) plotted against the portfolio return series (black) for the t-copula with GED innovations.

To consider more rigorously whether the obtained VaR and ES forecasts from the copula models with GED innovations are reasonable, we consider now the results of the DQ and ESR tests. The p-values are summarized in Table 5.1. If we consider first the DQ test of the VaR forecasts, we see that all p-values are above 0.05 at risk levels $\alpha = 0.01$ and $\alpha = 0.05$, indicating that the null hypothesis of the DQ test cannot be rejected for any model specification at a 5% significance level for both risk levels. As a result, it is implied that all copula models offer accurate VaR forecasts at both risk levels. The p-values for the VaR forecasts at risk level $\alpha = 0.01$ are all almost equal to 1. This indicates that these VaR forecasts are particularly accurate. Next, we observe that all p-values at both risk levels $\alpha = 0.01$ and $\alpha = 0.05$ are also above 0.05 for the ESR test of the ES forecasts. Therefore, the ESR test's null hypothesis cannot be rejected for any model specification at a 5% significance level for either of the risk levels. This suggests that all the specified copula models also provide accurate ES forecasts at both risk levels. Again, it should be noted that the p-values for the forecasts at risk level $\alpha = 0.01$ are noticeably greater than the p-values for the forecasts at risk level $\alpha = 0.05$, indicating that these forecasts are particularly accurate. This leads us to the conclusion that, according to the DQ and ESR tests, the copula models with GED innovations produce reasonable VaR and ES forecasts at both risk levels $\alpha = 0.01$ and $\alpha = 0.05$ at a significance level of 5%.

Copula	DQ 1% VaR p-value	ESR 1% ES p-value	DQ 5% VaR p-value	ESR 5% ES p-value
Gaussian	0.9483	0.4637	0.3956	0.0836
t	0.9963	0.5260	0.4061	0.0908
Frank	0.9479	0.4655	0.3951	0.0782
Gumbel	0.9532	0.4754	0.3952	0.0849
Clayton	0.9973	0.5199	0.3161	0.0990
Joe	0.9553	0.4631	0.3915	0.0787
BB7	0.9978	0.4878	0.3105	0.0958
Ali-Mikhail-Haq	0.9976	0.4566	0.3924	0.0779

Table 5.1: p-values for the DQ and ESR tests for VaR and ES forecasts at level $\alpha = 0.01$ and $\alpha = 0.05$ for the copula models with GED innovations.

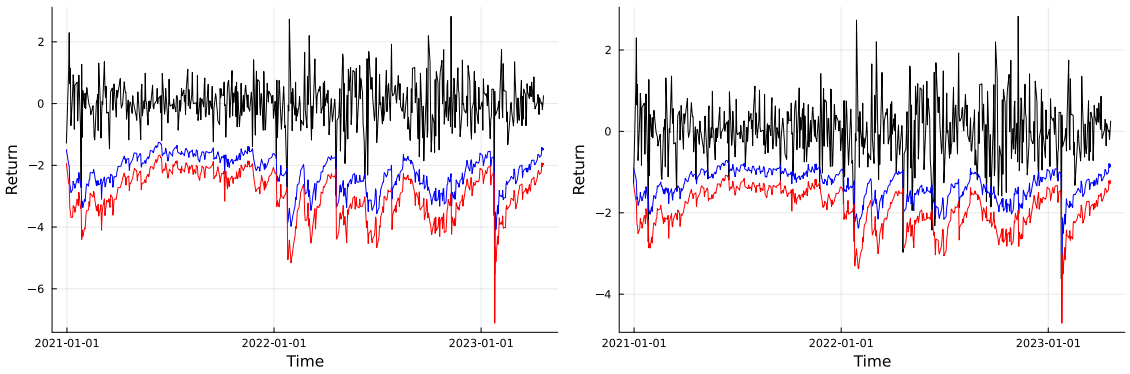


Figure 5.2: VaR (blue) and ES (red) forecasts at risk level $\alpha = 0.01$ (left) and $\alpha = 0.05$ (right) plotted against the portfolio return series (black) for the t-copula with distributions of innovations determined entirely by AIC.

Next, we consider plots of the portfolio returns alongside the obtained VaR and ES forecasts at risk level $\alpha = 0.01$ and $\alpha = 0.05$ for the t-copula where the distributions of the

innovations are determined entirely by AIC in each roll of the risk forecasting procedure. This is seen in Figure 5.2. Plots for the remaining copula models with distributions of innovations determined entirely by AIC are found in Figure B.6-B.7. It is still difficult to draw any conclusions from these plots, but visually, the VaR forecasts appear to be plausible. The ES forecasts are still lower than the VaR forecasts which, as previously mentioned, makes sense given the definition of ES.

Copula	DQ 1% VaR p-value	ESR 1% ES p-value	DQ 5% VaR p-value	ESR 5% ES p-value
Gaussian	0.9921	0.3850	0.2583	0.1651
t	0.9930	0.8347	0.4045	0.1871
Frank	0.9947	0.8199	0.2684	0.1593
Gumbel	0.9945	0.8383	0.2647	0.1785
Clayton	0.9908	0.1976	0.2693	0.1346
Joe	0.9948	0.8144	0.3384	0.1037
BB7	0.9928	0.8120	0.2707	0.1897
Ali-Mikhail-Haq	0.9946	0.8022	0.2730	0.1056

Table 5.2: p-values for the DQ and ESR tests for VaR and ES forecasts at level $\alpha = 0.01$ and $\alpha = 0.05$ for the copula models with distributions of innovations determined entirely by AIC.

The results of the DQ and ESR tests for the risk forecasts obtained from the copula models where the distributions of the innovations are determined entirely by AIC are seen in Table 5.2. The results are very similar to the results in Table 5.1. The DQ test of the VaR forecasts reveals that all p-values are above 0.05 at both risk levels, indicating that all copula models provide precise VaR forecasts at both risk levels at a 5% significance level. For the ESR test of the ES forecasts, we notice that all p-values at both risk levels are also above 0.05. As a result, it is also suggested that all of the copula models give precise ES forecasts at both risk levels. As in Table 5.1, the p-values for the VaR and ES forecasts at risk level $\alpha = 0.01$ are all considerably larger than the p-values for the forecasts at risk level $\alpha = 0.05$, demonstrating that these forecasts are again particularly accurate. We can therefore draw the conclusion that the copula models with innovation distributions determined entirely by AIC also give acceptable VaR and ES forecasts for both risk levels of $\alpha = 0.01$ and $\alpha = 0.05$ at a significance level of 5%.

Knowing that all copula models provide acceptable VaR and ES forecasts, it is of great interest to find the model(s) which provide the most accurate VaR and ES forecasts. To this end, the MCS described in Section 2.3.2 is employed. That is, we seek to determine the true set of the best model(s) $\mathcal{M}^*$ by iteratively removing the least suitable model from $\mathcal{M}$. In Table 5.3 the results regarding the copulas with GED innovations are shown. The table shows the MCS $\hat{\mathcal{M}}^*_{1-a}$ for $a \in \{0.01, 0.05, 0.25\}$ derived from the FZL and QL functions at risk levels $\alpha = 0.01$ and $\alpha = 0.05$. For a risk level of $\alpha = 0.05$, the MCS for both the FZL and QL functions consists of only the t-copula. This is evidence that the t-copula excels at VaR and ES forecasting compared to the other copula models. At risk level $\alpha = 0.01$, the results vary depending on which coverage probability and loss function the MCS is based on. First, we see that all copula models are included in the MCSs $\hat{\mathcal{M}}^*_{0.99}$ and $\hat{\mathcal{M}}^*_{0.95}$ based on the FZL function. This could indicate that the restrictions regarding which models are eliminated are too harsh. When considering the remaining MCSs, it is evident that the Frank, Joe, and Ali-Mikhail-Haq copulas all are selected as superior models when based on the QL and FZL functions. Thus, the conclusion to be drawn from Table 5.3 is that the t-copula is superior when forecasting VaR and ES at risk level $\alpha = 0.05$, whereas the

Frank, Joe, and Ali-Mikhail-Haq copulas are superior when forecasting VaR and ES at risk level $\alpha = 0.01$.

	FZL 1%	FZL 5%	QL 1%	QL 5%
$\hat{\mathcal{M}}^*_{0.99}$				
Best model(s)	All copulas	t-copula	Gaussian copula Frank copula Joe copula AMH copula	t-copula
$\hat{\mathcal{M}}^*_{0.95}$				
Best model(s)	All copulas	t-copula	Frank Copula Joe Copula AMH copula	t-copula
$\hat{\mathcal{M}}^*_{0.75}$				
Best model(s)	Frank copula Joe copula AMH copula	t-copula	Frank copula Joe copula AMH copula	t-copula

Table 5.3: Model confidence sets with coverage probabilities $a = 0.01$, $a = 0.05$, and $a = 0.25$ for the FZL and QL functions evaluated at risk level $\alpha = 0.01$ and $\alpha = 0.05$ for the copula models with GED innovations.

Now consider the MCSs from the copula models with distributions of innovations determined entirely by AIC in Table 5.4. Evidently, the MCSs again only consist of the t-copula for both loss functions at risk level $\alpha = 0.05$. This is again evidence that the t-copula is outperforming the other copula models.

	FZL 1%	FZL 5%	QL 1%	QL 5%
$\hat{\mathcal{M}}^*_{0.99}$				
Best model(s)	All copulas	t-copula	Frank copula Joe copula AMH copula	t-copula
$\hat{\mathcal{M}}^*_{0.95}$				
Best model(s)	Gaussian copula Frank copula Gumbel copula Joe copula BB7 copula AMH copula	t-copula	Frank copula Joe copula AMH copula	t-copula
$\hat{\mathcal{M}}^*_{0.75}$				
Best model(s)	Frank copula Joe copula AMH copula	t-copula	Frank copula Joe copula AMH copula	t-copula

Table 5.4: Model confidence sets with coverage probabilities $a = 0.01$, $a = 0.05$, and $a = 0.25$ for the FZL and QL functions evaluated at risk level $\alpha = 0.01$ and $\alpha = 0.05$ for the copula models with distributions of innovations determined entirely by AIC.

At risk level $\alpha = 0.01$, many of the same conclusions can be drawn as from Table 5.3. That is, for $\hat{\mathcal{M}}^*_{0.99}$ and $\hat{\mathcal{M}}^*_{0.95}$ based on the FZL function, the elimination restrictions seem

to be too harsh since all and almost all models are included, respectively. Again, the Frank, Joe, and Ali-Mikhail-Haq copulas are contained in all MCSs. This indicates that the Frank, Joe, and Ali-Mikhail-Haq copulas likely are the most suitable models for VaR and ES forecasting at risk level $\alpha = 0.01$.

Since the purpose of this thesis is to find the most accurate risk forecasting copula model(s), it is of particular interest to compare the models considered in Tables 5.3 and 5.4. The resulting MCSs are found in Table 5.5. For ease of notation, we let t-copula$_{\text{GED}}$ denote the t-copula with GED innovations. Similarly, we let t-copula$_{\text{AIC}}$ denote the t-copula with distributions of innovations determined entirely by AIC. This notation is also used for all other specified copulas in Table 5.5.

	FZL 1%	FZL 5%	QL 1%	QL 5%
$\hat{\mathcal{M}}^*_{0.99}$				
Best model(s)	All copulas	t-copula$_{\text{GED}}$	Frank copula$_{\text{GED}}$ Joe copula$_{\text{GED}}$ AMH copula$_{\text{GED}}$	t-copula$_{\text{GED}}$ t-copula$_{\text{AIC}}$
$\hat{\mathcal{M}}^*_{0.95}$				
Best model(s)	Gaussian copula$_{\text{GED}}$ Gaussian copula$_{\text{AIC}}$ t-copula$_{\text{GED}}$ Frank copula$_{\text{GED}}$ Frank copula$_{\text{AIC}}$ Gumbel copula$_{\text{GED}}$ Gumbel copula$_{\text{AIC}}$ Joe copula$_{\text{GED}}$ BB7 copula$_{\text{GED}}$ BB7 copula$_{\text{AIC}}$ AMH copula$_{\text{GED}}$ AMH copula$_{\text{AIC}}$	t-copula$_{\text{GED}}$	Frank copula$_{\text{GED}}$ Joe copula$_{\text{GED}}$ AMH copula$_{\text{GED}}$	t-copula$_{\text{GED}}$ t-copula$_{\text{AIC}}$
$\hat{\mathcal{M}}^*_{0.75}$				
Best model(s)	Frank copula$_{\text{GED}}$ Joe copula$_{\text{GED}}$ AMH copula$_{\text{GED}}$	t-copula$_{\text{GED}}$	Frank copula$_{\text{GED}}$ Joe copula$_{\text{GED}}$ AMH copula$_{\text{GED}}$	t-copula$_{\text{GED}}$

Table 5.5: Model confidence sets with coverage probabilities $a = 0.01$, $a = 0.05$, and $a = 0.25$ for the FZL and QL functions evaluated at risk level $\alpha = 0.01$ and $\alpha = 0.05$ for all the specified copula models.

It is evident that at risk level $\alpha = 0.05$ the MCSs based on the FZL function consist only of the t-copula$_{\text{GED}}$. This is supported by $\hat{\mathcal{M}}^*_{0.75}$ based on the QL function, however, $\hat{\mathcal{M}}^*_{0.95}$ and $\hat{\mathcal{M}}^*_{0.99}$ based on the QL function consist of two models, namely the t-copula$_{\text{GED}}$ and the t-copula$_{\text{AIC}}$. This suggests that the t-copula$_{\text{GED}}$ outperforms the t-copula$_{\text{AIC}}$ when the MCSs are based on the FZL function, whereas they are almost statistically similar in performance when the MCSs are based on the QL function. When considering the risk level $\alpha = 0.01$, the $\hat{\mathcal{M}}^*_{0.99}$ and $\hat{\mathcal{M}}^*_{0.95}$ based on the FZL function suggest that all and almost all copulas are not significantly different in performance, respectively. However, for $\hat{\mathcal{M}}^*_{0.75}$ based on the FZL function and all the MCSs based on the QL function, the superior models are the Frank copula$_{\text{GED}}$, Joe copula$_{\text{GED}}$, and AMH copula$_{\text{GED}}$. In fact, these models are part of all MCSs based on either the FZL or QL function, which indicates that they are the superior models when forecasting VaR and ES at risk level $\alpha = 0.01$.

Although no single copula model significantly outperforms the remaining copula models

when considering all risk levels and coverage probabilities, the t-copula is superior at risk level $\alpha = 0.05$, and the Frank, Joe, and Ali-Mikhail-Haq copulas are superior at risk level $\alpha = 0.01$. The most suitable copulas also tend to have GED innovations. This is in line with the in-sample model diagnostics, which showed that the GED appeared to fit its innovation series better than the distributions determined entirely by the AIC. This highlights the importance of proper marginal distribution modelling, as discussed in Chapter 4. It is important to note that, despite the fact that copulas with GED innovations tend to outperform those with distributions of innovations determined entirely by AIC, the latter nevertheless generate acceptable VaR and ES forecasts at both risk levels $\alpha = 0.01$ and $\alpha = 0.05$ according to the DQ and ESR tests. Furthermore, we note that the superior copula models, i.e. the Frank, Joe, Ali-Mikhail-Haq, and t-copulas are the same copula models that were deemed well-specified in Section 4.3 according to the Cramér-von Mises goodness-of-fit test. This provides evidence of these four copula models being superior in both goodness-of-fit and risk forecasting accuracy compared to the remaining copula models.

6 Conclusion

This master's thesis investigates how copulas can be employed for risk forecasting in stock portfolios, and how one can evaluate the accuracy of these forecasts and identify the most suitable copula model(s).

In order to provide answers to these questions, we begin by outlining all the necessary copula theory. Among other things, this includes the useful probability integral transforms, the specific copula examples that we will consider, and the crucial Sklar's theorem, which enables us to model each variable individually before modeling their dependence with a copula. In addition, we introduce the copula estimation method of Bayesian inference, the Cramér-von Mises goodness-of-fit test, and the marginal distribution modelling choice of ARMA-GARCH models. Based on this knowledge, we present the risk measures we would like to forecast, namely the value-at-risk (VaR) and expected shortfall (ES), as well as the procedure for doing so. In summary, the procedure involves fitting an ARMA-GARCH model to each return series under consideration, estimating distributions for the innovation series, and forecasting one-day-ahead return means and variances. The specified copulas are then estimated and Monte Carlo simulations of the return process for an equally weighted portfolio are performed. These results are used to obtain estimates for VaR and ES. Following that, this process is repeated on a rolling window basis. In addition to forecasting VaR and ES, our goal is to assess the forecasts' accuracy. In order to accomplish this, we present the dynamic quantile (DQ) and expected shortfall regression (ESR) tests, which test the accuracy of the VaR and ES forecasts, respectively. Both tests depend on a particular regression and a Wald-type test statistic. Finally, we introduce the theory of the model confidence set (MCS) in order to determine the most suitable copula model(s). The goal of the model confidence set is to identify a set of models which contain the best model(s) with a specified level of confidence. This is achieved through a sequence of significance tests, where the models that are significantly inferior to the other models are eliminated.

An equally weighted portfolio composed of the stocks of NextEra Energy and British Petroleum is then subjected to the theoretical solution to the problem statement described above. Prior to applying the risk forecasting procedure to this portfolio, we confirm that the ARMA-GARCH framework is an appropriate modeling choice. To do this, we take into account histograms, ADF tests, ACF plots, Ljung-Box tests, among other things. All of them point to the ARMA-GARCH framework as being reasonable. Marginal and copula model diagnostics are then performed for the in-sample data which consists of 1510 daily observations between the 1st of January 2015 and the 31st of December 2020. Here we find that ARMA-GARCH type models with t-distributed innovations have the lowest AIC values. However, it appears that the estimated t-distributions do not suit their innovation series well. We decide to also take into account ARMA-GARCH type models with exclusively GED innovations since the GED seems to provide a better fit for its innovations. In both instances of distribution specifications, the innovation series show no autocorrelation or heteroscedasticity, indicating that the aforementioned ARMA-GARCH type models are

appropriate options for marginal modelling. Following this, we fit the given copulas using the GED and t-distributed innovations and consider the Cramér-von Mises test results. From this, we find that the Frank, Joe, Ali-Mikhail-Haq, and t-copulas are all well-specified at a 5% significance level for both distribution choices.

Finally, we reach the out-of-sample risk forecasting part. Here, we apply the risk forecasting procedure to the 580 out-of-sample daily observations that fall between the 1st of January 2021 and the 25th of April 2023. Again, we take into account both the scenario where the distributions of innovations are determined entirely by AIC and the scenario where the distributions are fixed to be GED. We find, that both the VaR and ES forecasts are accurate for the risk levels $\alpha = 0.01$ and $\alpha = 0.05$ in both of the innovation distribution scenarios, according to the results of the DQ and ESR tests. As a result, the risk forecasting procedure provides reasonable risk forecasts for both distribution specification scenarios. For the model confidence set, we find that the t-copula is the most suitable model for risk assessment at risk level $\alpha = 0.05$, while at risk level $\alpha = 0.01$, the Frank, Joe, and Ali-Mikhail-Haq copulas appear to be the most suitable ones since they are included in all the MCSs. These four copula models are also the models that are deemed well specified by the Cramér-von Mises test. When all risk levels and coverage probabilities are taken into account, no single set of copula models significantly outperforms the other copula models. Therefore, we are unable to determine a single set of superior models at both risk levels, however, we can infer that the most suitable copula models typically have GED innovations.

To sum up, we first provide a theoretical response to the problem statement by introducing the necessary copula theory, the risk forecasting procedure, as well as the VaR and ES backtests and the model confidence set. All of this is then applied to a portfolio composed of the two stocks of NextEra Energy and British Petroleum. From this, we discover that the VaR and ES estimates are accurate regardless of whether the innovation distributions are determined entirely by AIC or fixed to be GED. Additionally, we find that no particular copula model outperforms the others in terms of risk assessment in general, nevertheless, the t-copula and the Frank, Joe, and Ali-Mikhail-Haq copulas appear to be the most suitable ones at risk levels of $\alpha = 0.05$ and $\alpha = 0.01$, respectively. Additionally, copula models with GED innovations typically outperform those with innovation distributions determined entirely by AIC.

7 Future Considerations

The research of this thesis could be extended in the following ways to provide a more nuanced answer to the problem statement.

We showed that the use of univariate ARMA-GARCH type models was justified when modelling the marginal distributions. As mentioned in Section 1.7 this is not the case in general, as the conditioning set is the information set of all considered series. Hence, if the series have cross-equation effects, multivariate ARMA-GARCH type models must be utilized. In this case, the proposed risk forecasting procedure in Section 2.2 should be modified to instead fit and forecast from a multivariate ARMA-GARCH type model. It is worth mentioning that we only tested cross-equation effects in the first moment (ARMA) and did not explore the second moment (GARCH) due to time constraints. Consequently, the implementation of an ARMAX-GARCHX type model was not performed, which presents an extension to be considered.

We chose to investigate Archimedean and parametric copulas. A second extension would be to incorporate other copula models available in the `BivariateCopulas` package, i.e. convex, discretized, non-parametric, rotated, and vine copulas. Additionally one could test for time-varying dependence between the investigated stocks. As suggested by [20], one can compute quantile dependence coefficients at different quantiles to obtain more information about the dependence structure. If necessary, copula models can be extended to capture time-varying dependence, often referred to as time-varying copulas in the literature.

When ARMA-GARCH type models were fitted to the marginals we elected to only consider the models with the lowest AIC values and $p, q \in \{0, \dots, 3\}$. A fourth extension would be to perform risk forecasting based on all ARMA-GARCH type models regardless of AIC values and for larger values of p, q. In doing so, the MCSs would be the only method of model selection, selecting among thousands of copula models. Note that this is likely infeasible, as the complexity of the procedure would be massively increased.

In Section 4.2, the marginal model diagnostics were presented. Initially, the ARMA-GARCH distribution specification was determined entirely by AIC. However, it was observed that this choice resulted in a poor fit to the associated innovation series. As we emphasized, a proper specification is of paramount importance as incorrectly specified marginals can lead to misspecification of the copula models. To address this issue, we enforced a generalized error distribution (GED) as the distribution specification, which exhibited a better fit to the associated innovations. Throughout the applications part of the thesis, we found that the copulas with GED innovations seemed to outperform the copulas with distributions of innovations determined entirely by AIC in terms of risk forecasting. This leads to the fifth extension of including several more flexible distributions e.g. the normal inverse Gaussian (NIG) distribution, which can account for the shape, skewness, and scale of the distribution through its parameters. With a flexible distribution such as the NIG distribution, it should be possible to fit distributions to the marginals that ensure the innovations properly adhere to the distribution specification.

In Tables 5.3-5.5 the most suitable risk forecasting copula model(s) are presented. By the construction of the model confidence set (MCS), see Section 2.3.2, the coverage probability of the MCS determines the confidence that the true best model(s) is not excluded from the MCS. We chose the coverage probabilities to be $a = 0.01$, $a = 0.05$, and $a = 0.25$, in accordance with [17]. As a consequence, the MCS for $a = 0.01$ is much more inclined to include an unsuitable model than to exclude a suitable one. Hence, a sixth extension to be considered would be to expand the range of a and to determine a sweet spot for the coverage probability.

In the risk forecasting procedure, we considered an equally weighted portfolio consisting of two stocks. An interesting final extension would involve portfolio allocation based on risk forecasting. This can be accomplished by considering $\mathbb{P}(\mathcal{P}_t < b)$ where $\mathcal{P}_t$ is an arbitrary portfolio return series and $b \in \mathbb{R}$. The copula models would then be used to determine the probability $\mathbb{P}(\mathcal{P}_t < b)$. Subsequently, any portfolio satisfying $\mathbb{P}(\mathcal{P}_t < b) < r$ where $r \in [0, 1]$ would be considered a favorable portfolio for appropriate values of b and r. It should be noted that some alterations in the risk forecasting procedure are required in this case.

Appendices

A The Fréchet-Hoeffding Bounds

Theorem A.1 (The Fréchet-Hoeffding Bounds)**.**
For every copula $C \in \mathscr{C}_d$, we have the bounds:

$$\max\left\{\sum_{i=1}^{d} u_i + 1 - d, 0\right\} \leqslant C(u_1, \ldots, u_d) \leqslant \min\{u_1, \ldots, u_d\}.$$

Proof. For the first inequality, we observe that:

$$\begin{aligned} C(u_1, \ldots, u_d) &= \mathbb{P}\left(\bigcap_{i=1}^{d}\{U_i \leqslant u_i\}\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{d}\{U_i > u_i\}\right) \\ &\geqslant 1 - \sum_{i=1}^{d}\mathbb{P}(U_i > u_i) = 1 - \sum_{i=1}^{d}(1 - u_i) = \sum_{i=1}^{d} u_i + 1 - d, \end{aligned}$$

where the third step follows from Boole's inequality. For the second inequality, we see that:

$$\bigcap_{j=1}^{d}\{U_j \leqslant u_j\} \subseteq \{U_i \leqslant u_i\}, \quad i = 1, \ldots, d,$$

which implies that:

$$C(u_1, \ldots, u_d) = \mathbb{P}\left(\bigcap_{j=1}^{d}\{U_j \leqslant u_j\}\right) \leqslant \min_{i\in\{1,\ldots,d\}}\{\mathbb{P}(U_i \leqslant u_i)\} = \min(u_1, \ldots, u_d).$$

■

B Miscellaneous Tables and Figures

	NextEra Energy ARMA(3, 3)-sGARCH(1, 1)	British Petroleum ARMA(2, 2)-tGARCH(1, 1)
	Mean equation parameters	
c	0.0003 (0.0002)	-1.9e-5 (0.0005)
ϕ_1	-0.5128 (0.0451)	0.5081 (0.0202)
ϕ_2	0.2845 (0.0638)	-0.9725 (0.0328)
ϕ_3	0.9032 (0.0452)	-
κ_1	0.4781 (0.0362)	-0.4958 (0.0258)
κ_2	-0.3321 (0.0509)	0.9616 (0.0410)
κ_3	-0.9482 (0.0364)	-
	Volatility parameters	
ω	4.8e-06 (1.6e-06)	3.5e-06 (1.3e-06)
α_1	0.0978 (0.0230)	0.0334 (0.0157)
β_1	0.9213 (0.0165)	0.9272 (0.0130)
γ_1	-	0.0712 (0.0223)
	Distribution parameters	
ν	5.4194 (0.7226)	5.0090 (0.7110)

Table B.1: Parameter estimates and standard errors of the most suitable models for the log return series of NEE and BP determined entirely by AIC.

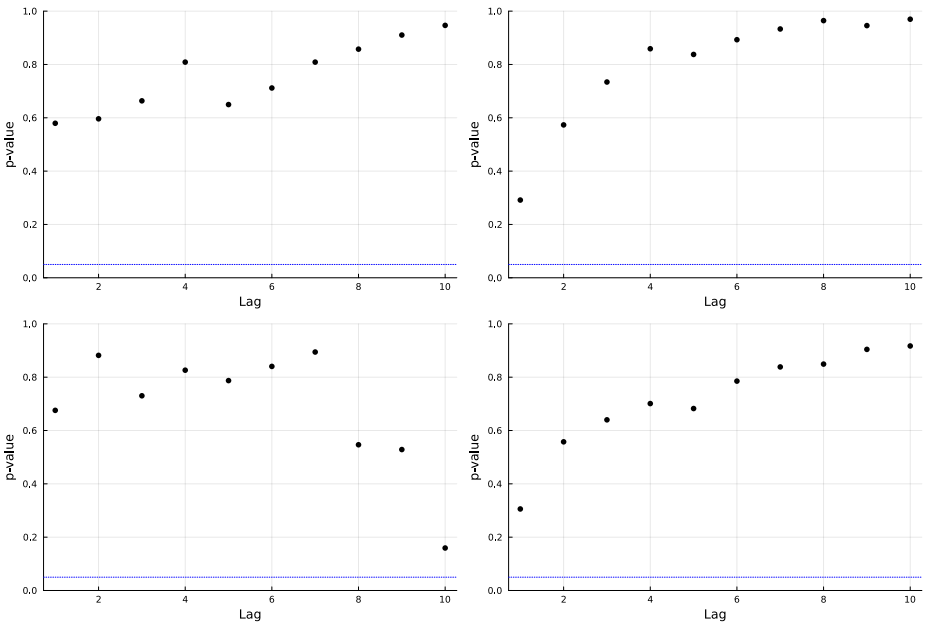


Figure B.1: Ljung-Box test of the t-distributed innovations (top) and squared innovations (bottom) for NEE (left) and BP (right).

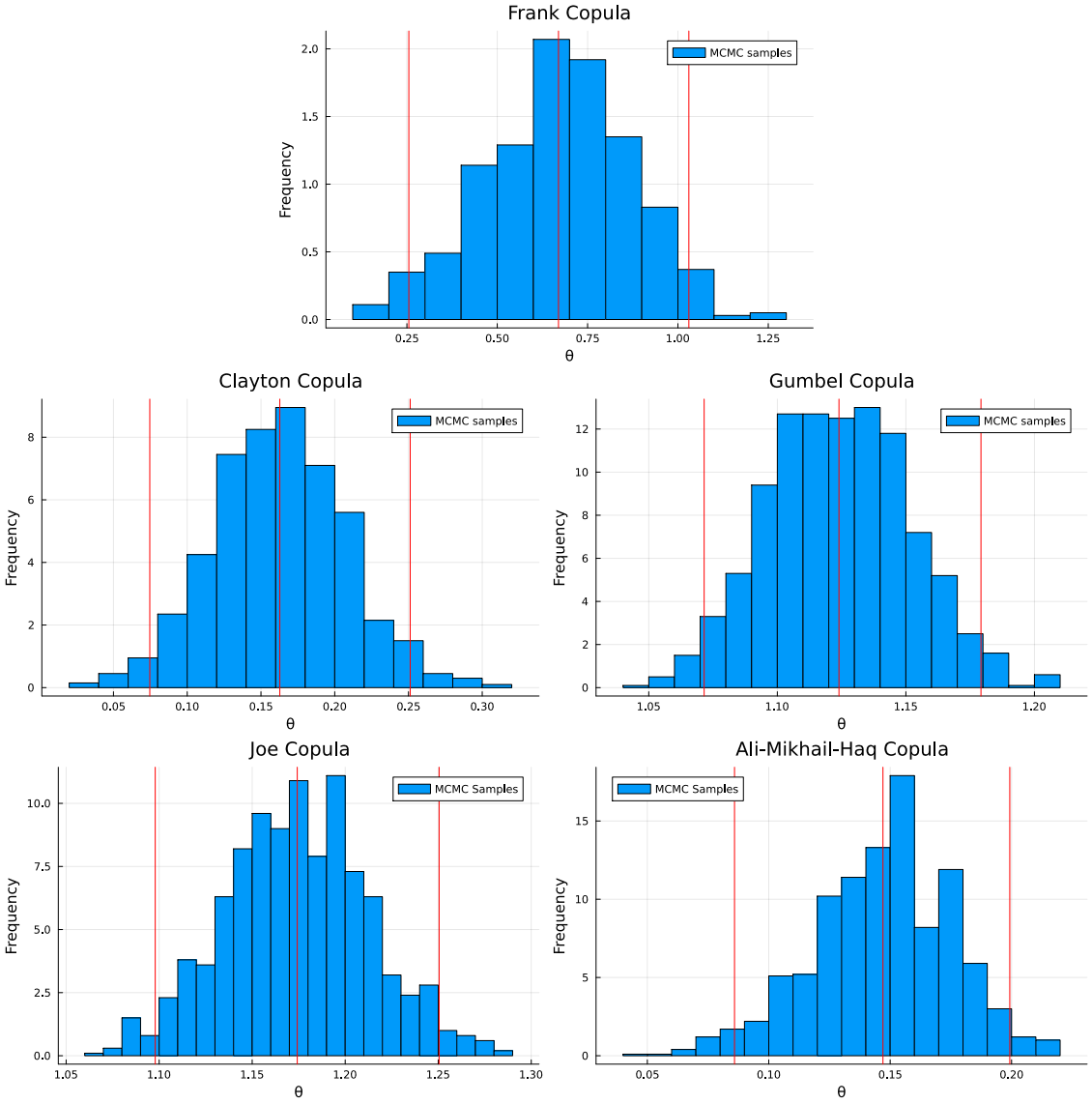


Figure B.2: MCMC samples of the posterior distributions, credible intervals, and parameter estimates for the remaining copulas with GED innovations.

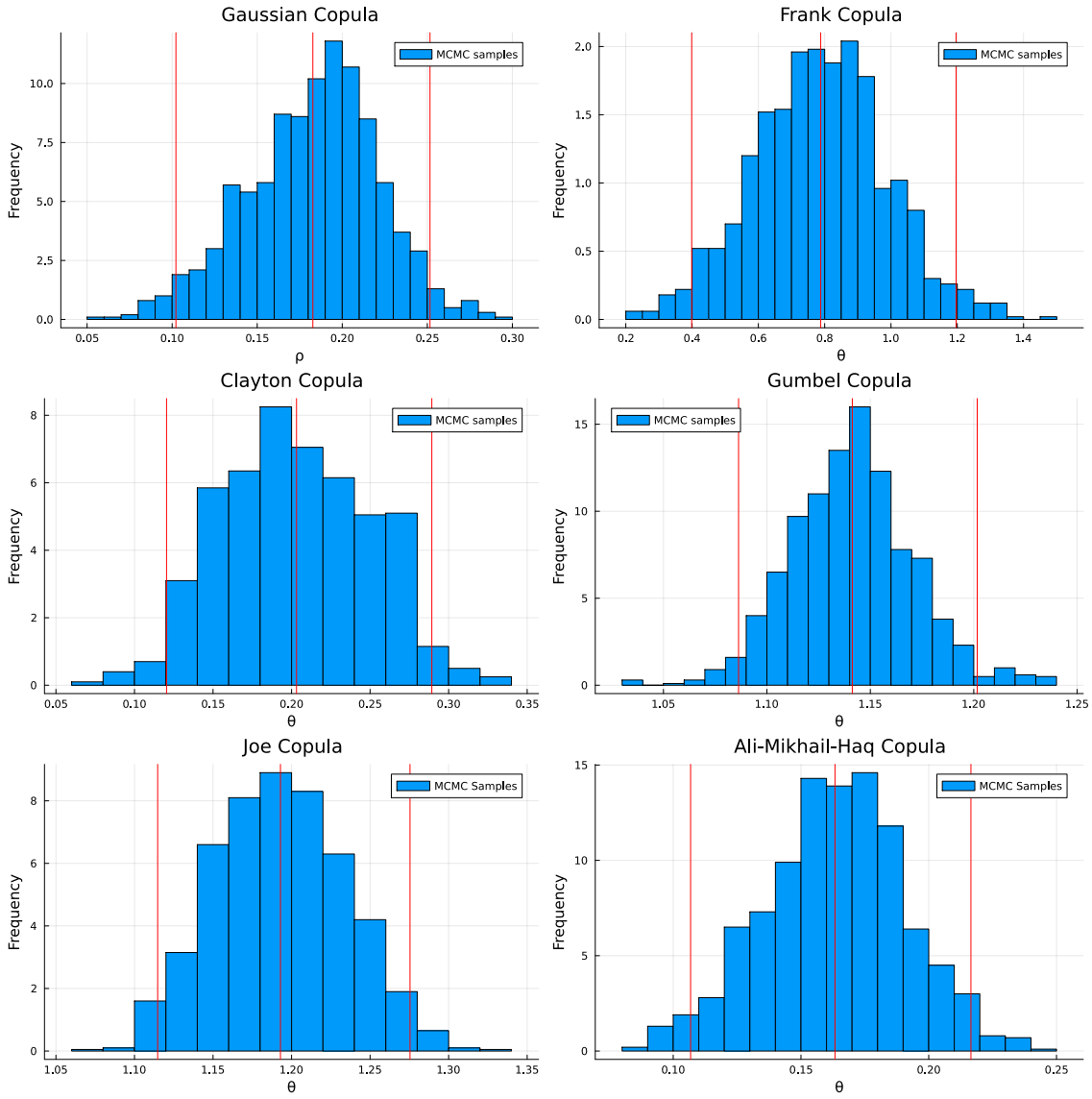


Figure B.3: MCMC samples of the posterior distributions, credible intervals, and parameter estimates for all specified copulas with t-distributed innovations.

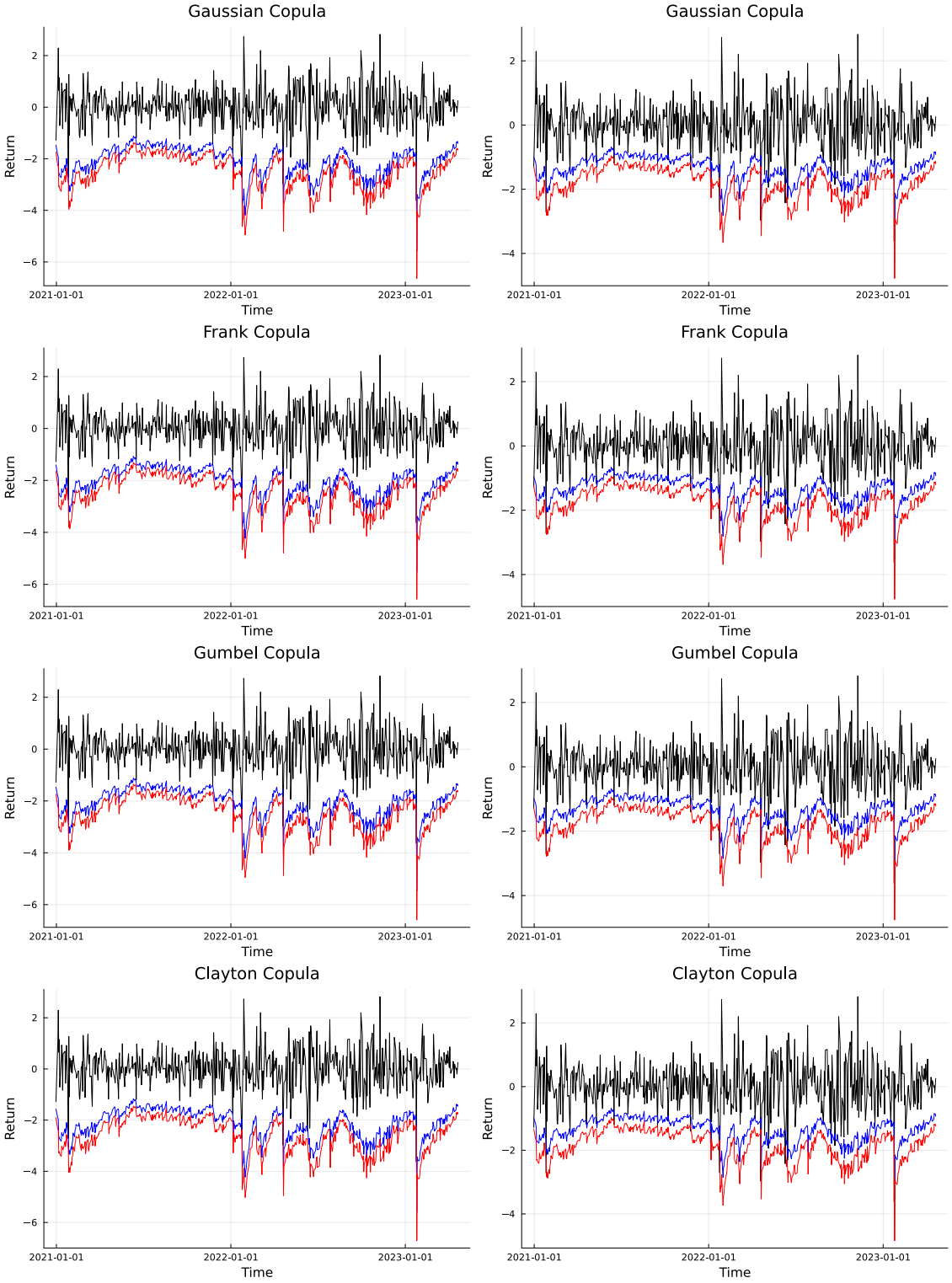


Figure B.4: VaR (blue) and ES (red) forecasts at risk level $\alpha = 0.01$ (left) and $\alpha = 0.05$ (right) plotted against the portfolio return series (black) for the Gaussian, Frank, Gumbel, and Clayton copulas with GED innovations.

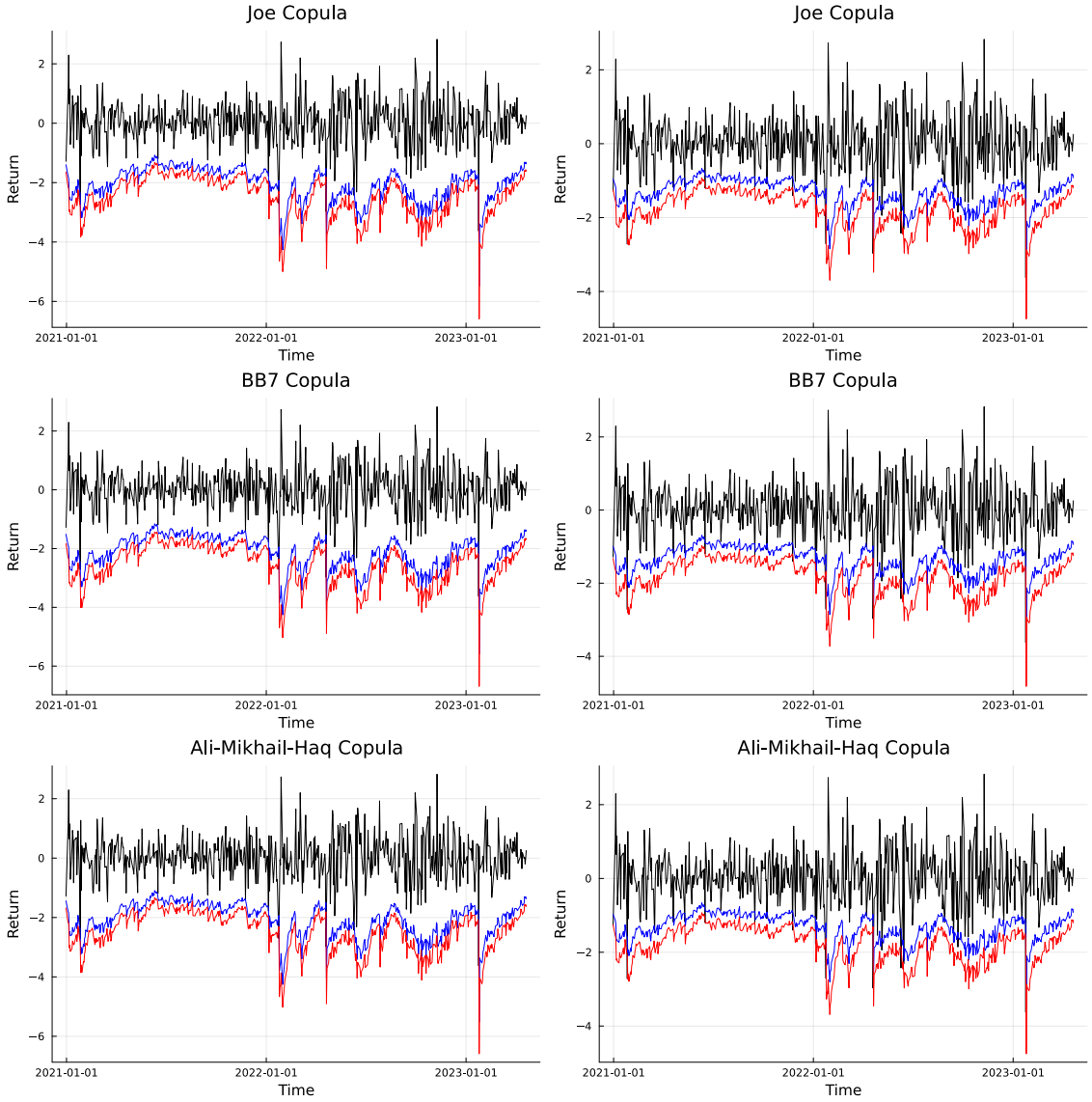


Figure B.5: VaR (blue) and ES (red) forecasts at risk level $\alpha = 0.01$ (left) and $\alpha = 0.05$ (right) plotted against the portfolio return series (black) for the Joe, BB7, and Ali-Mikhail-Haq copulas with GED innovations.

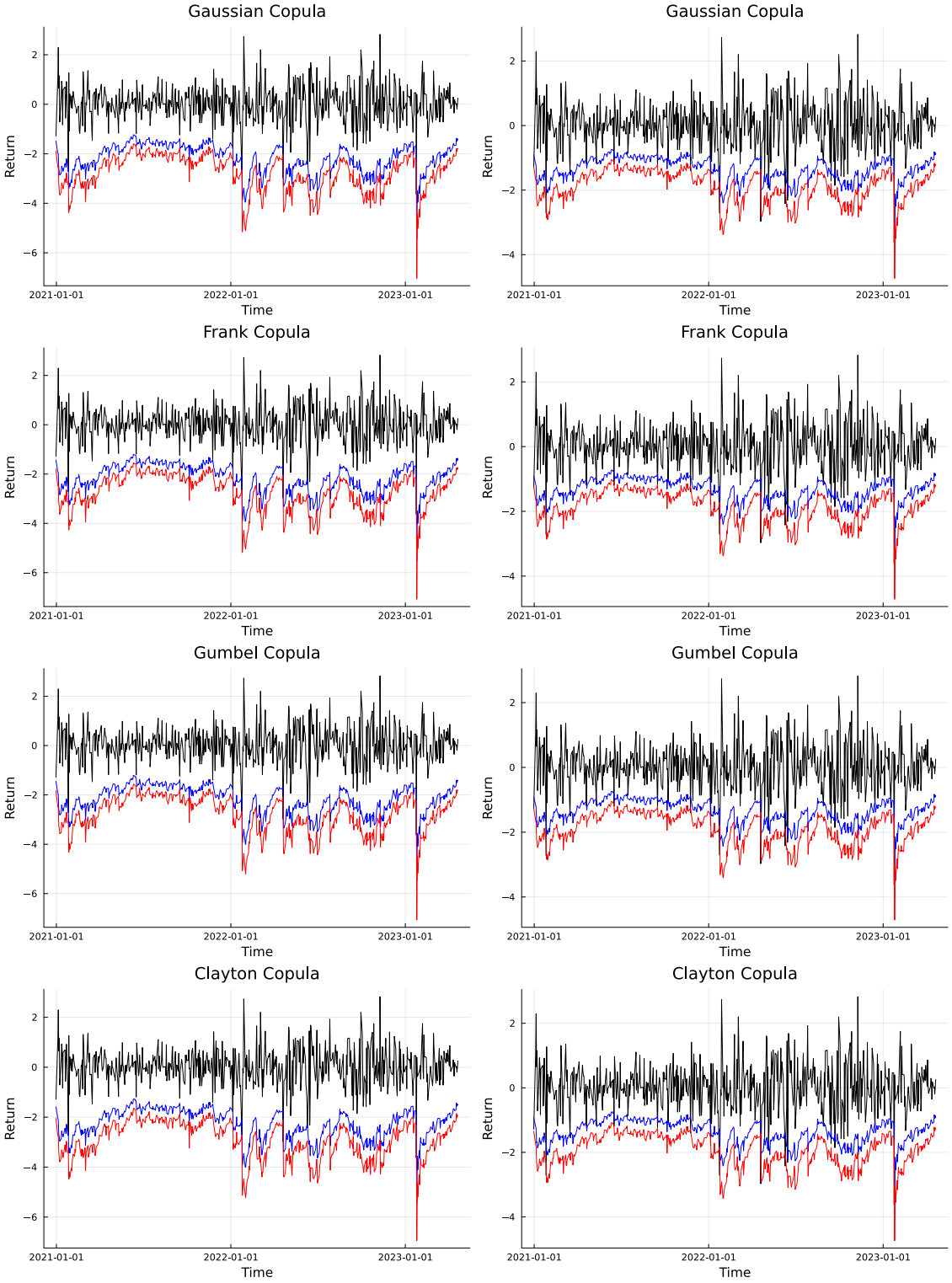


Figure B.6: VaR (blue) and ES (red) forecasts at risk level $\alpha = 0.01$ (left) and $\alpha = 0.05$ (right) plotted against the portfolio return series (black) for the Gaussian, Frank, Gumbel, and Clayton copulas with distributions of innovations determined entirely by AIC.

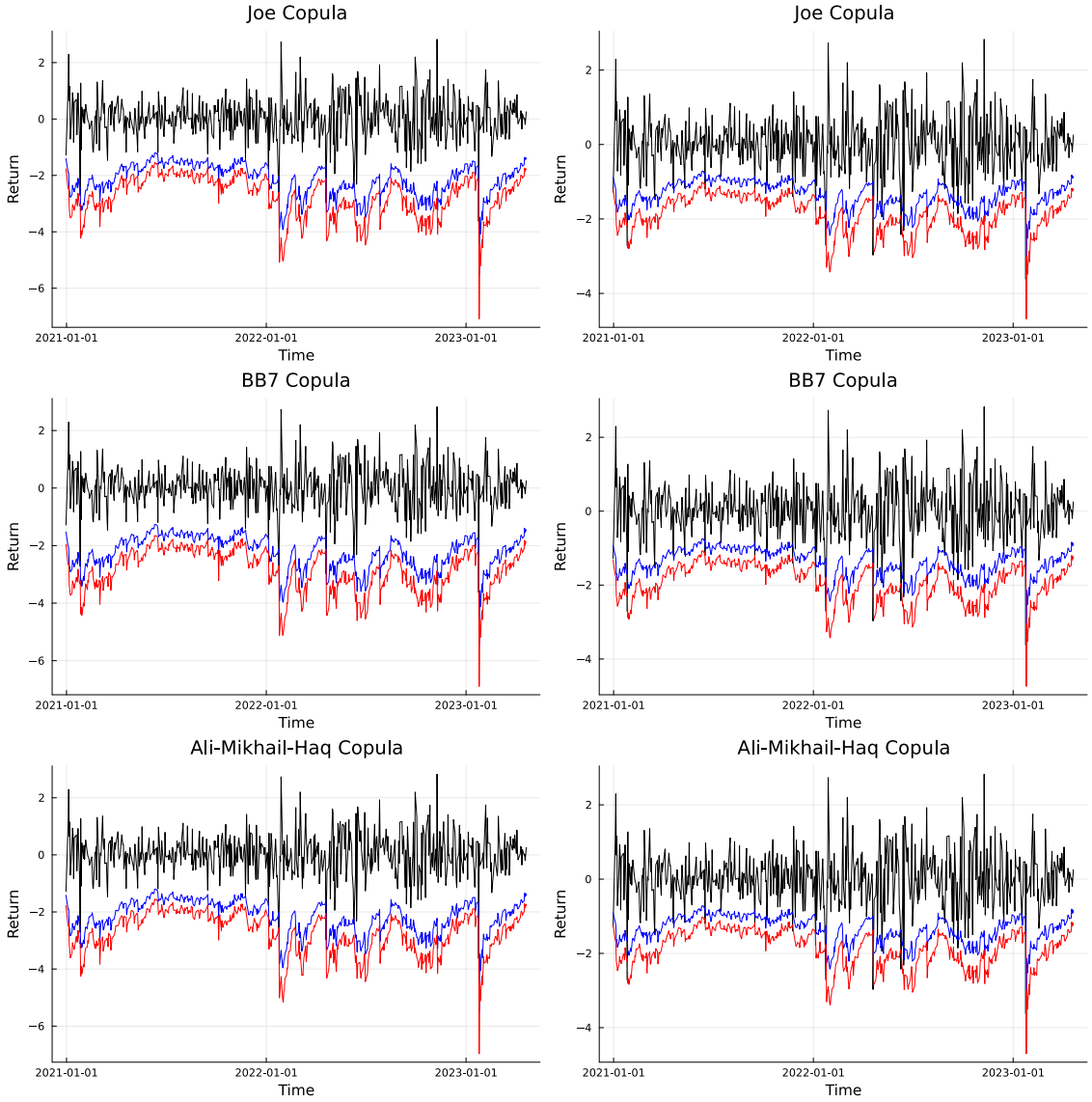


Figure B.7: VaR (blue) and ES (red) forecasts at risk level $\alpha = 0.01$ (left) and $\alpha = 0.05$ (right) plotted against the portfolio return series (black) for the Joe, BB7, and Ali-Mikhail-Haq copulas with distributions of innovations determined entirely by AIC

Bibliography

[1] Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control*, 18(5):931–955, 1994.

[2] Rüdiger Frey & Paul Embrechts Alexander J. McNeil. *Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition*. Princeton University Press, 1st edition, 2015.

[3] Christian Genesta & Bruno Rémillard & David Beaudoin. Goodness-of-Fit Tests for Copulas: A Review and a Power Study. *Insurance: Mathematics and Economics*, 44:199–213, 2009.

[4] Daniel Berg. Copula Goodness-of-Fit Testing: An Overview and Power Comparison. *Journal of Finance*, 15:675–701, 2009.

[5] David Ardia & Keven Bluteau & Kris Boudt & Leopoldo Catania. Forecasting Risk with Markov-switching GARCH Models: A Large-scale Performance Study. *International Journal of Forecasting*, 34(4):733–747, 2018.

[6] Sebastian Bayer & Timo Dimitriadis. Regression-Based Expected Shortfall Backtesting. *Journal of Financial Econometrics*, 20(3):437–471, 2022.

[7] Olivier P. Faugeras. Sklar's Theorem Derived Using Probabilistic Continuation and Two Consistency Results. *Journal of Multivariate Analysis*, 122:271–277, 2013.

[8] Gildas Mazo & Stéphane Girard & Florence Forbes. A Class of Multivariate Copulas Based on Products of Bivariate Copulas. *Journal of Multivariate Analysis*, 140:363–376, 2015.

[9] Matthew D. Hoffman & Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1351–1381, 2014.

[10] Martin B. Haugh. An Introduction to Copulas. 2016.

[11] Anca Pircalabu & Thomas Hvolby & Jesper Jung & Esben Høg. Joint Price and Volumetric Risk in Wind Power Trading: A Copula Approach. *Energy Economics*, 62:139–154, 2017.

[12] Courtney St. James. What Causes Market Volatility? Navigating Ups and Downs of an Unstable Market. *Public Holdings, Inc.*, 2022.

[13] Erik Kole, Kees Koedijk, and Marno Verbeek. Selecting Copulas for Risk Management. *Journal of Banking Finance*, 31(8):2405–2423, 2007.

[14] Xun Fa Lu & Kin Keung Lai & Liang Liang. Portfolio Value-at-Risk Estimation in Energy Futures Markets with Time-varying Copula-GARCH Model. *Springer Science+Business Media*, 219:333–357, 2011.

[15] Robert F. Engle & Simone Manganelli. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business Economic Statistics*, 22(4):367–381, 2004.

[16] David Ruppert & David S. Matteson. *Statistics and Data Analysis for Financial Engineering*. Springer Texts in Statistics, 2nd edition, 2015.

[17] Peter R. Hansen & Asger Lunde & James M. Nason. The Model Confidence Set. *Econometrica*, 79(2):453–497, 2011.

[18] Daniel B. Nelson. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2):347–370, 1991.

[19] Andrew J. Patton. Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula. *UCSD Discussion Paper*, (01-09), 2001.

[20] Andrew J. Patton. On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation. *Journal of Financial Econometrics,*, 2:130–168, 2004.

[21] Annalisa Di Clemente & Claudio Romano. Calibrating and Simulating Copula Functions in Financial Applications. *Frontiers in Applied Mathematics and Statistics*, 7(642210), 2021.

[22] Paul Damien & Petros Dellaportas & Nicholas G. Polson & David A. Stephens. *Bayesian Theory and Applications*. Oxford University Press, 1st edition, 2013.

[23] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics, 1st edition, 2003.