A Vine Copula Approach for Portfolio Optimisation

Exploring the Effect of Copulas and Vine Models on Optimal Investment Allocation of Stock Index Returns

Juliette Godard

Delft University of Technology

TU Delft

A Vine Copula Approach for Portfolio Optimisation

Exploring the Effect of Copulas and Vine Models on Optimal Investment Allocation of Stock Index Returns

by

Juliette Godard

To obtain the degree of Master of Science in Applied Mathematics at TU Delft,
to be defended publicly on Friday August 30^{th}, 2024 at 12:30 PM.

Student number:	5619467	
Thesis committee:	Prof. A. Derumigny	TU Delft, Daily Supervisor
	Prof. D. Kurowicka	TU Delft, Chair and Thesis Advisor
	Prof. G.F. Nane	TU Delft, External Examiner

Cover: `https://images.app.goo.gl/YvoV2SS1Vha17Vwq6`.

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.
Github : `https://github.com/AlexisDerumigny/MScProjectJulietteG`.

TU Delft

Acknowledgements

This Master thesis represents the end of my TU Delft journey, finalizing my degree in Applied Mathematics. Before beginning of thesis thesis, I feel necessary to make some acknowledgments.

First of all, a special thanks goes to both my supervisors, Dorota Kurowicka and Alexis Derumigny, for their precious advises, their availability and flexibility.

I would like to thank my beloved parents, my sister, and my grandmother, who have always believed in me and supported every choice I made, giving me strength to overcome all obstacles.

I am infinitely grateful to all my friends and colleagues for being able to brighten my days and for always being there for me. Thank you to all my friends in Delft, for making me feel home even if far away from home, for sharing the tough times, and the library days. A special mention goes to Luca for his precious advises, and support. And lastly, but not for importance, thank you to my dear Italian friends, Maria and Irene, who were able to feel close, even if geographically distant.

Juliette Godard
Delft, August 2024

Disclaimer: in the process of refining the text quality of this thesis, the use of ChatGPT has been used.

Summary

This thesis explores the growing complexity of contemporary financial markets, which is a consequence of a world that is increasingly interconnected and correlated. This evolution highlights the necessity of understanding and accurately modeling these underlying relationships, which translates into the need of incorporating more complex models into portfolio optimization, breaking away from Harry Markowitz's foundational Portfolio Optimization Theory. While Markowitz's model has been effective, the complexity of modern financial instruments demands more sophisticated approaches. This study focuses on the application of copulas and vine models to portfolio optimization, aiming to understand how these advanced models can enhance the optimization process by accurately capturing dependencies among financial assets. In particular, this thesis investigates the benefits of integrating copula-GARCH models, a combination of time series modelling where the residuals are modelled using copulas or vine models, into portfolio theory. Through this approach, the research aims to extend existing knowledge and highlight the specific advantages provided by these models in portfolio optimization.

Contents

List of Figures

Table of Notation

Notation	Definition	Reference
Ω	Sample space	Sect. E.1
$A, B \in \Omega$	Events	Sect. E.1
$\mathbb{P}(A)$	Probability of the event A	Def. 27
$\mathbb{P}(A\|C)$	Conditional probability of the event A, given C	Def. 28
X	Random variable	Def. 29 and 30
$\mathbf{X} = (X_1, ..., X_d)$	Random vector	Def. 32
f	Probability density function of X	Def. 28
F	Cumulative distribution function of X	Def. 31
$F(x_1, ..., x_d)$	Joint distribution function of the random vector $\mathbf{X}$	Def. 33
$F_i(x)$	Marginal distribution function of the random variable X_i	Def. 34
$f(x_1, ..., x_d)$	Joint density function of the random vector $\mathbf{X}$	Def. 35
$f_i(x_i)$	Marginal density function of the random variable X_i	Def. E.1
$\mathbb{E}(X)$	Expected Value of X	
$Var(X)$	Variance of X	
$\mathcal{M}$	Statistical model	Def. 36
$\mathcal{M}_\theta = \{\mathbb{P}_\theta : \theta \in \Theta\}$	Parametric model	Def. 37
Θ	Set of parameters	Def. 37
$\theta \in \Theta$	Parameter	Def. 37
θ^*	"true" parameter	Sect. F.1
$\hat{\theta}$	Estimator of the parameter θ	Def. 38
$\hat{\theta}^{ML}$	Maximum Likelihood estimator of θ	Sect. 40
$\hat{\theta}^{CMLE}$	Canonical Maximum Likelihood estimator of θ	Eq. (3.16)
$\hat{\theta}^{SSP}$	Stepwise Semiparametric Estimator of θ	Sect. 4.5
$MSE(\hat{\theta})$	Mean squared error of the estimator $\hat{\theta}$	Def. 39
$L(\theta)$	Likelihood function of the parameter θ	Def. 40
$l(\theta)$	Log-Likelihood function of the parameter θ	Def. 40
AIC	Akaike's Information Criterion	Def. 41
BIC	Bayesian Information Criterion	Def. 41
H_0, H_1	Null and Alternative test hypothesis	Sect. F.3
α	Significance Level of an Hypothesis Test	Sect. F.3.2
Φ	Cumulative Distribution function of a 1D standard normal	Sect. E.2.1
ϕ	Probability density of a 1-dimensional standard normal	Sect. E.2.1
(l_d, u_d)	Confidence interval of a parameter θ	Def. 42
$\hat{F_N}(x), \hat{F_{i,N}}(x)$	Empirical Cumulative distribution function	Sect. F.4.1
$KLIC(f_0, f, \theta)$	Kullback–Leibler Information criterion	Def. 12
$C(u_1, ..., u_d)$	Multivariate distribution function of a d-dimensional copula	Def 3
$c(u_1, ..., u_d)$	Density function of a d-dimensional copula	Def 3.2
$\varphi^{[-1]}, F^{[-1]}, \Phi^{[-1]}$	Pseudo-inverses	Theo 4
$\tau, \hat{\tau}_n$	Kendall's Tau and Estimated Kendall's Tau	3.2.1 & (3.6)
$\rho_s, \hat{\rho}_s$	Spearman's Correlation and Estimated Spearman's	Def. 5
$\mathcal{C} = \{\mathbb{P}_\rho : \theta \in \Theta\}$	Copula model	Def. 11
$C_{I\|J}(\mathbf{u}\|\mathbf{X}_J = \mathbf{x})$	Conditional copula	Def. 13
$C_{I\|J}^{A_J}(\mathbf{u}\|\mathbf{X}_J \in A_J)$	Conditional copula on the event $\mathbf{X}_J \in A_J$	Def. G.1
$\hat{C}(\mathbf{u})$	Survival copula	Def. 9

Notation	Definition	Reference
$W_d(\mathbf{u}), M_d(\mathbf{u})$	Fréchet-Hoeffding Bounds	Def. 7
$S = (X_1, ..., X_d)$	Initial sample of n independent observations	Sect. G.1
$S^* = (X_1^*, ..., X_d*)$	Constructed sample of n observations	Sect. G.1
$\bar{\mathcal{A}}_J = \{A_{1,J}, ..., A_{m,J}\}$	Partition of the original set into m disjoint subsets	Prop. 2
$T = \{N, E\}$	Tree	Def. 15
$e = \{a, b\}$	Edge of a tree	Def. 4.5
$\mathcal{M}_V = (\mathcal{F}, \mathcal{V}, \mathcal{B})$	Vine model	Def. 17
ψ_j	Log pair-copula density	Eq. (4.8)
$\hat{l}$	Pseudo-likelihood	Eq. (4.9)
μ_t	Mean Value	Def. 18
$\gamma_X(t, s)$	Autocovariance Function	Def. 18
$corr(X_{t+h}, X_t)$	Autocorrelation of X_t at lag h	Def. 26
$ACF(s, t)$	Autocorrelation Function	Def. 25
$\widehat{ACF}(h)$	Sample Autocorrelation	Sect. F.4.2
$PACF(s, t)$	Partial Autocorrelation Function	Def. 26
$WN(0, \sigma^2)$	White Noise	Def. 19
$rRMSE$	Relative Root Mean Squared Error	Def. 5.4
$\gamma_X(h)$	Autocovariance of a time series	Def. 19
$AR(p)$	Autoregressive Model	Def. 20
$MA(q)$	Moving Average Model	Def. 21
$ARIMA(p, d, q)$	ARIMA processes	Sect. 5.1.1
$ARCH(p)$	Autoregressive Conditional Heteroskedasticity model	Sect. 5.1.2
$GARCH(p, q)$	Generalized Autoregressive Conditional Heteroskedasticity model	Sect. 5.1.2
$\hat{X}_t(m)$	Forecasted Time Series value, m-days-ahead	Sect. 5.6
Ω_T	Information set up to time T	Sect. 7.1
$VaR_\alpha(X)$	Value-at-Risk	Def. 1
$CVaR_\alpha(X)$	Conditional Value-at-Risk	Def. 2
$P(t)$	Portfolio at time t	Eq. (2.1)
$S = \{s_1, \ldots, s_n\}$	Set of Securities	Eq. (2.1)
$r_i(t)$	Logarithmic Returns	Eq. (2.3)
$\mathbf{w}(t) = (w_1(t), \ldots, w_N(t))$	Vector of Weights of the portfolio at time t	Eq. (2.2)
$\mathbf{R}(t) = (R_1(t), \ldots, R_N(t))$	Vector of Discounted Returns of portfolio at time t	Eq. (2.4)
$R(t)$	Portfolio's Return at time t	Eq. (2.5)
$\mu_P(R)$	Expected return of the portfolio	Eq. (2.6)
$\sigma_P(R)$	Standard Deviation of the Portfolio	Eq. (2.7)
$L(w, R)$	Loss Function of the portfolio	Eq. (2.8)
$\rho(\mathbf{w}, \mathbf{R})$	Risk Measure of the Portfolio	Sect. 2.2
r_f	Risk-free Rate	Eq. (2.31)
SR	Sharpe Ratio	Eq. (2.31)
$STARR$	STARR Ratio	Eq. (2.32)
e_t	Standardized Residuals	Def. 5.3

1

Introduction

The contemporary financial market is a dynamic and constantly-evolving landscape, resulting in financial instruments becoming increasingly complex to understand and model. Understanding this complexity is crucial for investors seeking to make informed and rational decisions in their pursuit of financial success. The foundation of modern Portfolio Optimisation Theory, laid by Harry Markowitz with the publication of "Portfolio selection" in 1952 ([1]), provided an initial framework for constructing optimal investment portfolios. However, given the continuous transformation of the financial world, one may question whether Markowitz's relatively simplistic model remains sufficiently robust in capturing the intricate dynamics at play. On the other hand, there is a question about whether introducing additional complexity to the models brings true advantages to the optimisation process. This thesis aims to address these critical questions by analyzing the impact of more sophisticated models, with a specific focus on copulas and vine models, applied to portfolio optimisation techniques.

Modern portfolio optimisation is the theory, developed by Markovitz, that focuses on the strategic allocation of wealth among different financial assets available [2], [3]. The main idea behind Markovitz's theory is to construct an *efficient frontier*, by finding a balance between maximising the expected return and minimizing the variance of the portfolio assets: the exact solution on the frontier is then decided subjectively depending on individual risk preferences. The model in its simplicity has proved to be effective. However, over time many extensions of the original framework have been developed to keep up with the increasing complexity of modern financial instruments. Among these extensions, one particularly emerges as a useful instrument: *copula models* and their ability to accurately estimate and model the relationship between asset prices and stock indices.

Copulas were first introduced by Sklar in 1959 [4], offering a framework to characterize the joint dependence of multiple variables while separating it from the marginal distributions. By employing copulas, one can capture the underlying dependence structure independently of the single variables' distributions, allowing to set the focus on the dependence between the variables. A further advancement, introduced by Bedford and Cooke [5], is represented by the introduction of vine models, which represent a more flexible type of copulas. Vine models enable the construction of the joint distribution by decomposing it into bivariate copulas. This approach allows for a more structured representation

of the dependencies among financial assets and stock indices, enabling a deeper understanding of their collective behavior.

This master thesis therefore wants to reflect on the advantages of introducing copulas and vine models within the realm of portfolio theory, and specifically focus on the following questions:

What are the advantages of introducing copulas and vine models, in the context of portfolio optimisation theory? How can the properties of these models be used to improve and understand the optimisation process?

This will be the scope of this master thesis, which will investigate the advantages of applying vine copulas to portfolio optimisation, through the development of a time-dependant model that collects a set of financial assets, processes, models it, and construct an optimal portfolio. The output of the process will be a time varying set of weights, which indicates to what extent one needs to invest in a given asset at a given time period (for a clear explanation of the model refer to Section 7.2). Following the work of Barthel et al. [6] and Serban and Brockwell [7], we will extend the existing research on copula models for Portfolio optimisation, by focusing on the advantages of such models, and on the specific information that can be gained by their application.

1.1 Thesis Outline

This thesis follows a structured approach that allows the reader to comprehensively explore and analyze the concepts of portfolio theory, time series, copulas, and vine models, before diving into the core applications to financial modeling. We will provide all the necessary theoretical background in the first four chapters, and then move on to the practical applications. We can divide the structure in three macro parts, which are outlined as follows.

1.1.1 Theoretical Background

The first chapters of this thesis will be focused on providing the reader with all the necessary instruments for understanding the more practical aspects that will follow. Each chapter will provide the needed theoretical information, along with a series of simulation studies aimed at clarifying the main concepts, at underlying the practical use of each mathematical model described, and at highlighting possible up and downsides of each of them. We will start by introducing Portfolio Optimisation Theory in Chapter 2, where the main risk measures and their use will be introduced, along with a comparison of different methodologies that can be adopted to select an optimal portfolio. Chapter 3 will focus on the concept of copulas and Sklar's theorem. It will also discuss the most used dependence measures associated with copulas, and the extension to conditional copulas. Subsequently, we will extend copulas to vine copulas in Chapter 4. Additionally, the latter will offer an overview of the methods for estimating the parameters, and for selecting the most suitable family and structure for vine models. We will proceed to discuss time series in Chapter 5, along with a brief simulation study. Lastly, we will conclude by combining time series to vine copula models, and introduce the concept of copula-GARCH models in Chapter 6.

1.1.2 Application to Finance

Once all the theoretical basis will be set, we will collect all the information discussed in Chapter 7 , where we will discuss how to combine time series and copulas to construct optimal portfolios, delineating the procedure that we will adopt to analyze real financial data. In particular, we will discuss the transition to a time dependant model, and will illustrate how to leverage advanced techniques such as vine copula modeling and Monte Carlo simulation to capture dependencies and uncertainties in financial data. We will then explain how to use the gained information for portfolio optimisation. In Chapter 8 we will apply the methodology discussed in the previous Chapter to a set of international stock indexes. We will extract the closing prices, perform some data manipulation, and finally perform a portfolio optimisation with all the instruments provided. The obtained results will be carefully evaluated, and the first conclusions will be made. In addition, from the fitted vine model we will be able to draw some considerations about the market. In the concluding chapter, Chapter 9, we summarize the key findings of this thesis and discuss their implications within the word of financial modeling. In addition, we will examine some potential errors that could have been made throughout the process, discuss possible improvements, and extensions of the methodology created.

1.1.3 Appendix

Some of the results from the simulations provided in the theoretical Chapters will be collected in Appendix B and C; while in D some additional results from the application to real data of Chapter 8 are collected. For readers seeking additional theoretical background, we have included an Appendix (E, F) that presents fundamental statistical and probabilistic concepts underlying the theories discussed throughout the following chapters. In addition, G explores the main idea behind the Simplifying assumption, a key assumption that can be made in the context of vine models to remove some degrees of complexity. Lastly, Appendix A will provide some proofs of theorems and results discussed throughout this thesis.

2

Portfolio Optimisation

Portfolio optimisation theory, introduced by Harry Markovitz with the publication of "Portfolio selection" in 1952 ([1]), is centered around the strategic allocation of wealth among a selection of N different financial assets forming the portfolio basket. The primary goal is to find a balance between financial gain and risk [2], [3]. At the heart of Markowitz's contribution is the innovative concept of constructing an "efficient frontier." This frontier is created by seeking an optimal compromise between maximizing the expected return and minimizing the variance of the portfolio's returns. The specific solution on this frontier is then determined subjectively, taking into account individual risk preferences. A more in depth explanation of this approach, referred to as the *Mean-Var Approach*, will be provided in Section 2.3.1.

Since Markovitz, the theory around portfolio optimisation has been addressed using many instruments, ranging from statistical approaches, like Value at Risk (that will be discussed in 1) and Conditional Value at Riks, to machine learning-based methods such as Bayesian network (refer to [8] for more details), or Neural Networks. [9] offers a wide overview of some of the most well known methods, and also explores alternative and more recent methods: among them the use of reinforcement learning techniques and quantum-based approaches emerge.

Through the main sections of this Chapter, we will construct all the basis needed to understand the main concepts behind portfolio optimisation. Starting with an exploration of the foundational aspects in Section 2.1, where we will define some key concepts, we will then move to Section 2.3 where we will introduce risk measures, ranging from the classic Markowitz Mean-Variance Portfolio Optimization to more contemporary tools like Value-at-Risk and Conditional Value-at-Risk. Section 2.1 discusses the concept of efficient frontier, which is complemented by the introduction of the tangent portfolio in 2.4, incorporating metrics such as the widely-utilized Sharpe Ratio and the STARR Ratio, which are useful to select the optimal portfolio. Sections 2.5 and 2.6 bring analytical and simulation-based approaches into focus, respectively, providing readers with a comprehensive understanding of portfolio optimisation methods and their practical applications.

2.1 Definition

A financial asset is defined as a liquid asset with a value derived from a contractual right or ownership claim. Examples of such assets include stocks, bonds, mutual funds, cash and other form of commodities. Investors, each with their available wealth, can opt to invest in various assets, which will be forming what is known as a *portfolio*. These assets can be viewed as random variables, and the selection of an optimal portfolio will be determined by the behaviour of the said assets, and on the personal preferences of the investor according to the risks involved. A portfolio can be mathematically represented as follows:

$$P(t) = \sum_{j=1}^{N} w_j(t) s_j, \tag{2.1}$$

where $w_j(t)$ represents the weight associated to a certain security, hence the percentage of the total capital invested in the $j-th$ security at time t. $S = \{s_1, \ldots, s_N\}$ is the set of available securities [9]. The weights at each time point are therefore defined as:

$$w_j(t) = \frac{\text{Amount invested in } s_j}{\text{Total invested in } P(t)}. \tag{2.2}$$

The goal behind portfolio optimisation theory is to find an optimal set of weights $w_j(t)$, by maximising the returns, while minimizing risk. What we mean by risk will be clarified in the Section 2.2. The different problems and approaches used will be analyzed in the following Sections.

Each security s_i, is associated with a return $r_i(t)$ at each time point t. There are different ways of computing returns, in this thesis we will use the daily log-returns, which are defined as:

$$r_i(t) = \log \frac{p_i(t)}{p_i(t-1)}, \tag{2.3}$$

where $p_i(t)$ is the closing price at time t. The choice of log-returns offer numerous advantages: they are, for example, additive over time, hence they can be added together to obtain the total return over a longer period. Moreover, for each security we can define the discount rate $d_i(t)$ from each time point to the present. Consequently, the discounted anticipated return of s_i, can be therefore defined as:

$$R_i = \sum_{t=1}^{\infty} d_i(t) r_i(t). \tag{2.4}$$

$\{R_i\}_{i=1,\ldots,N}$ is a random variable, since it is a function of the random variables $r_i(t)$. Although portfolio theory was initially developed as a static framework for one-period models, it has since evolved into a more complex field of study, with ongoing research and innovation [3]. In this chapter, our focus remains on a single-period framework, deliberately excluding the consideration of time by neglecting the variable t in our notation from now on. Time will be introduced subsequently in our exploration of time-varying models.

By defining $\mathbf{R} = (R_1, \ldots, R_N)$, the random vector of the returns of the portfolio, and the vector of weight $\mathbf{w} = (w_1, \ldots, w_N)$ associated to the chosen portfolio, such that $w_i \geq 0, \forall i = \{1, \ldots, N\}$ and

$\sum_{i=1}^{N} w_i = 1$, we can formulate the discounted return of the portfolio as:

$$R = \mathbf{w}^T \mathbf{R} = \sum_{i=1}^{N} w_i R_i. \tag{2.5}$$

Since R is a random variable, obtained as a linear combination of random variables. Moreover, let $\boldsymbol{\mu} = (\mu_1, \dots \mu_N)^T$ be the vector of expected returns of $\mathbf{R}$, hence μ_i is the mean of the discounted return R_i, and let Σ_{ij} be the covariance between R_i and R_j, then from (2.5) we obtain the *expected return* of the chosen portfolio as:

$$\mu_P(R) = \sum_{i=1}^{N} w_i \mu_i, \tag{2.6}$$

and the *variance* of the portfolio as:

$$\sigma_P^2(R) = \sum_{i=1}^{N} \sum_{j=1}^{N} \Sigma_{ij} w_i w_j. \tag{2.7}$$

The mean and variance of each component, as well as the covariance matrix Σ, can be estimated using historical data, financial models, or other forecasting techniques [1]. In the following sections we will discuss different optimisation methods, and explore portfolio evaluation techniques, including risk assessment measures.

2.2 Portfolio Selection

2.2.1 Definition of the Problem

When seeking a way to compare and evaluate different portfolios, a commonly used strategy involves the use of risk measures. As the name suggests, a risk measure $\rho(\cdot)$ is statistical tool designed to quantify the risk associated with negative events, and it is expressed as a non-negative real number. Risk measures enable the comparison of different events, such as various investments in the financial domain, by assessing the associated risk of potential losses, which are characterized by a loss function $L(\mathbf{w}, \mathbf{R})$. Given a set of weights $\mathbf{w}$ and random vector of the returns $\mathbf{R}$, we will use as loss function the following:

$$L(\mathbf{w}, \mathbf{R}) = -\mathbf{w}^T \mathbf{R}. \tag{2.8}$$

This loss represents the negative of the returns of a portfolio with the chosen set of weights. The risk measures will then be a function of this loss.

Some widely used measures are Variance, Value at Risk (VaR_β), and Conditional Value at Risk ($CVaR_\beta$). These measures can be applied to a broad range of financial instruments, facilitating the quantification of risk, and will be discussed in Section 2.3.

With all the necessary elements in place, we are able to define the general portfolio optimisation problem, which involves a constrained double-objective problem, aiming to both maximize the expected return and minimize the associated risk. In general, given a specific risk measure, the problem can be summarized as follows:

$$\begin{cases} \min \rho(\mathbf{w}, \mathbf{R}) \\ \max \mathbf{w}^T \boldsymbol{\mu} \\ s.t. \ \sum_{i=1}^{N} w_i = 1. \end{cases} \tag{2.9}$$

To adopt a more practical approach, we can reformulate problem (2.9) into a single-objective one, making the problem more straightforward to solve, by selecting from two alternative formulations. The first approach involves setting a set of target risks, denoted as ρ_0, and solving to maximize the return for each risk. Alternatively, one can establish a set of target expected returns, r_0, and solve by minimizing the risk measure while constraining the portfolio's expected value to match the specified target return.

The latter methodology, as suggested by [7], transforms the problem into the following optimisation:

$$\begin{cases} \min \rho(\mathbf{w}, \mathbf{R}) \\ s.t. \ \mathbf{w}^T \boldsymbol{\mu} = r_0 \\ s.t. \ \sum_{i=1}^{N} w_i = 1, \end{cases} \tag{2.10}$$

where r_0 represents the target expected return. Throughout this thesis, we will adopt this second method, although it's worth noting that both methods provide equivalent results. The uniqueness of the solution to problem (2.10) depends on the choice of the risk measure $\rho(\mathbf{w}, \mathbf{R})$. To ensure uniqueness, $\rho(\mathbf{w}, \mathbf{R})$ needs to be convex, as we will discuss in the upcoming Sections.

It is important to notice that the solution to the problem stated in (2.9) is not unique; instead, it forms a set of points within the risk-expected return space, defining what we refer to as the *efficient frontier*, which will be elaborated in the following Section 2.2.2. Each feasible solution point on the frontier is associated with a specific set of weights $\mathbf{w}$. In practical terms, the construction of the frontier involves systematically selecting a vector of target returns.

2.2.2 Efficient frontier

After having chosen a proper risk measure, hence the objective function of the optimisation problem (2.9), we will identify a set of optimal points among all possible solutions to the problem, which compose the *feasible set* - the set of solutions that satisfy the constraints. The optimal points are known as the *Efficient Frontier*, and are of the form (ρ_i, μ_i): each point corresponds to an optimal portfolio of N assets with a set of specific weights $\mathbf{w}$. Figure 2.1 provides a visual representation, with each data point denoting a unique portfolio of assets. It is worth noting that there is no single, definitive solution: we will discuss methods for selecting portfolio in Section 2.4.

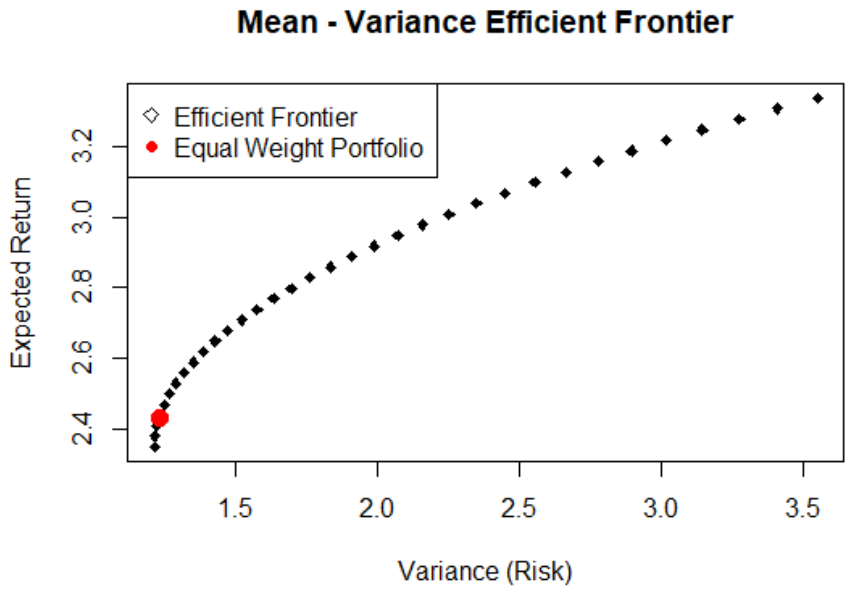


Figure 2.1: *Example of efficient frontier for a Mean-Variance problem, computed optimising a portfolio of two normally distributed asset's returns.*

2.3 Risk Measures

In this Section we analyse the main Risk measures that can be used for problem (2.10). As already anticipated, we will discuss Variance, Value at Risk (VaR_β), and Conditional Value at Risk ($CVaR_\beta$).

2.3.1 Variance

The Mean-Variance Markowitz Portfolio Selection approach is still widely used in finance and investment management, as it provides a framework easy to construct. The chosen risk measure is variance, setting the risk of problem (2.9) to be:

$$\rho(\mathbf{w}, \mathbf{R}) = \sigma_P^2(\mathbf{w}, \mathbf{R}) = \mathbf{w}^T \Sigma(\mathbf{R}) \mathbf{w}.$$

In its simplicity, it provides fast solutions that are still valid today. However, it is important to note that the model relies on certain assumptions, such as static parameters, that may not fully capture the complexities of real-world markets. Another drawback of the method is that it is not able to fully capture the relationships between variables if characterized by complex relationships (e.g. polynomial relationships) [9]. Various extensions and modifications have been developed to address these limitations, such as incorporating additional risk measures, considering transaction costs, and accounting for parameter uncertainty.

In addition, some limitations of adopting variance as a risk measure include its inability of taking into account asymmetries, and skewness, which are a common trait of financial data. In fact, because of how it is defined, variance is sensitive to the magnitude of deviations from the mean, but does not consider the direction of those deviations. In other words, it treats positive and negative deviations from the mean identically. Moreover, it lacks subadditivity, a crucial property for a coherent risk

measure (a coherent measure, is a risk measure that satisfies the four axioms of translation invariance, sub-additivity, positive homogeneity, and monotonicity, see [10] for more details). The variance of the sum of two random variables can, in fact, be expressed as

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2). \tag{2.11}$$

While a coherent risk measure, denoted as $\rho(\cdot)$, should adhere to the subadditivity criterion:

$$\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2), \tag{2.12}$$

property that is not met in the case of negative covariance between the random variables. The subadditivity property is directly associated with a diversified portfolio.

Additionally, another drawback of selecting variance as a risk measure is its lack of convexity, leading to the potential presence of multiple local extrema, which is an undesirable property, especially in optimisation contexts. A risk measure is considered convex if it adheres to the following relationship:

$$\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2), \tag{2.13}$$

while from the extended formula (2.11), we find that the convexity relationship is satisfied only if $Cov(X_1, X_2) \leq \frac{1}{2}(Var(X_1) + Var(X_2))$, condition that is not always met. To address these limitations, alternative risk measures like Value at Risk and $CVaR_\beta$ have gained importance.

2.3.1.1 Analytical Computation of the Mean-Variance Efficient Frontier

In the Markovitz's scenario, we are able to derive the analytical expression of the Efficient frontier. This represents a big advantage for the method, simplifying the resolution of the problem.

If we consider the mean-variance optimisation version of the problem introduced in (2.10), setting the Risk equal to the variance of the portfolio, the efficient frontier can be derived analytically by restating the problem with the corresponding Lagrangian function. If we consider:

$$\begin{cases} a = \boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu} \\ b = \boldsymbol{\mu}^\top \Sigma^{-1} \mathbb{1} \\ c = \mathbb{1}^\top \Sigma^{-1} \mathbb{1}. \end{cases} \tag{2.14}$$

we can derive the points of the parabola of the efficient frontier (r_0, Var), from the following theorem.

Theorem 1. *Let Σ be non singular and the returns are non identical, hence $R_i \neq R_j$, for some $i, j \in \{1, \ldots, N\}$. Then the region of efficient portfolios is described by the parabola:*

$$\mathrm{Var}(r_0) = \frac{cr_0^2 - 2br_0 + a}{\Delta}, \tag{2.15}$$

where Δ is defined as $\Delta = ac - b^2$.

The corresponding weight vector is obtained as:

$$\boldsymbol{w}(r_0) = \frac{\left(c\Sigma^{-1}\boldsymbol{\mu} - b\Sigma^{-1}\mathbb{1}\right) r_0 + \left(a\Sigma^{-1}\mathbb{1} - b\Sigma^{-1}\boldsymbol{\mu}\right)}{\Delta}. \tag{2.16}$$

The proof is taken from [3], and is available in Section A.1.2 of the Appendix.

2.3.1.2 Example for Mean-Variance

To have a better understanding of how the frontier is constructed, [1] proposes an example with three securities. We simulated a similar example with a portfolio of two assets, which serves purely for illustrative purposes.

Our objective is to delineate two graphs within the space of weights $\mathbf{w} = (w_1, w_2)$. The first corresponds to a set of *isomean curves*, which are lines characterized by the same expected return. They are constructed by setting a target return, defined as r_0. Similarly, an *isovariance curve* is a set of points with the same variance, the target variance, Var_0. The results are shown in Figure 2.2, where the reader can identify the set of feasible solutions, as well as the set of isomean curves, and the isovariance curves. The center of the isovariance ellipses corresponds to the point with lower variance.

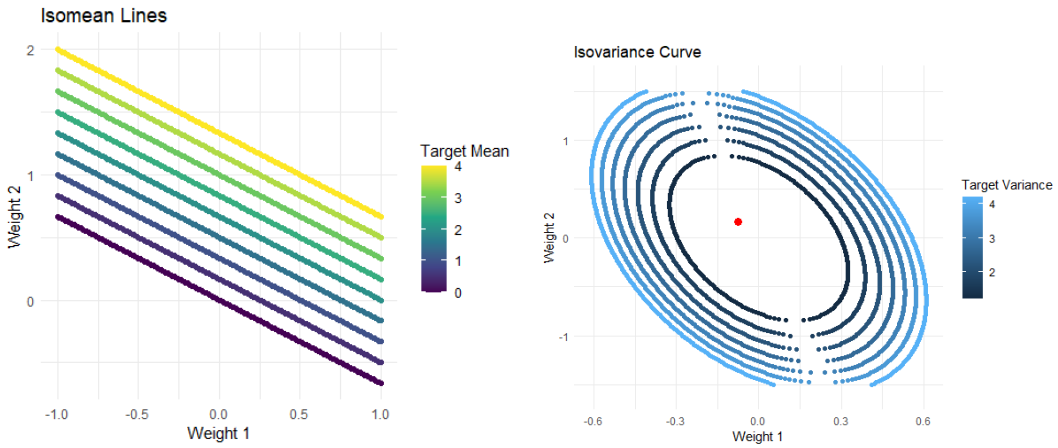


Figure 2.2: *Isomean and Isovariance curves obtained from two normally distributed returns. The red dot represents the weights associated with unconstrained minimum variance. The isomean lines are obtained from* $w_2 = \frac{r_0}{3} - \frac{2}{3}w_1$*, while the ellipses* $1.5w_1^2 + 1.4w_1w_2 + 2w_2^2 - Var_0 = 0$*. With* $r_0 \in \{0, \ldots, 4\}$*, and* $Var_0 \in \{1.19, \ldots, 4.19\}$*.*

More specifically, the isomean curve is obtained by finding the weights that satisfy $r_0 = \mathbf{w}^T\boldsymbol{\mu}$, for a specific target return. In the case of two assets, this corresponds to equation:

$$w_2 = \frac{r_0}{\mu_2} - \frac{\mu_1}{\mu_2}w_1,$$

which is a line with slope $-\frac{\mu_1}{\mu_2}$.

Equivalently, in the 2-dimensional case, the isovariance curves are a set of concentric ellipses, obtained by solving the quadratic equation $Var_0 = \mathbf{w}^T\Sigma\mathbf{w}$. In this case the weights compose the ellipse:

$$w_1^2\Sigma_{11} + 2w_1w_2\Sigma_{12} + w_2^2\Sigma_{22} - Var_0 = 0,$$

where Σ_{ij} represents the component ij of the covariance matrix Σ.

Among all the possible solutions represented in Figure 2.2, we can select the ones corresponding to the efficient frontier by solving (2.16) for a set of chosen target returns r_0.

2.3.2 Value-at-Risk

In 1994, financial institution JP Morgan proposed a new instrument to evaluate risk, whose objective was to "establish a benchmark for credit risk measurement" [11]: the Value at Risk, or VaR_β. It was lately adopted as the standard measure of risk for financial regulations by the Basel Committee [12]. This new tool's aim was to evaluate the risk, by quantifying how much of the capital invested could be lost with a certain degree of confidence β.

If we consider a decision vector w, which corresponds to the optimal set of weights taken from the feasible set of solutions, and considering the loss $L(\mathbf{w}, \mathbf{R})$ defined in Eq. 2.8, we can provide the mathematical definition of Value at Risk for the random variable L:

Definition 1 (Value at Risk (VaR))**.** *Value-at-Risk (VaR) of the random variable that expresse the loss $L(\boldsymbol{w}, \boldsymbol{R})$, with confidence level $\beta \in (0, 1)$ is:*

$$VaR_\beta(L) = \min\{z \ s.t. \ F_L(z) \geq \beta\},$$

where $F_L(z) = \mathbb{P}(L \leq z)$ is the cumulative distribution (see Def. 31) of the loss.

VaR_β represents the lower $\beta-$percentile of $L(\mathbf{w}, \mathbf{R})$, hence the lowest value such that the probability of the loss being less or equal to that value is equal to β.

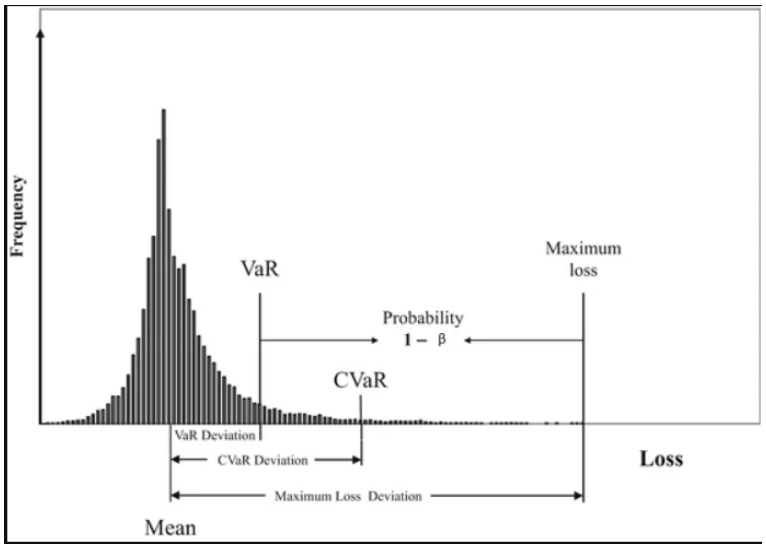


Figure 2.3: *Value at Risk and Conditional Value at Risk, [13].*

The intuition behind Value-at-Risk becomes clear by observing Figure 2.3, where the $VaR_\beta(L)$ is represented by a specific value in the loss distribution: the region under the curve. The risk measure can be computed at different confidence levels, typically $\beta = \{0.90, 0.95, 0.99\}$ are considered.

One advantage of using VaR_β is that it is characterized by stability in the estimation, in fact "it is not affected by very high tail losses" [13]. By not explicitly considering these very high tail losses, VaR avoids being overly influenced by extreme events that might be difficult to measure accurately or predict reliably. Value at Risk, can be estimated using different methods, including parametric approaches like historical data analysis, or Monte Carlo simulations [12].

However, as [12] points out, VaR_β presents some drawbacks that make its use not ideal in the context of portfolio optimisation. In fact, VaR_β is not a coherent measure since it does not satisfy sub-additivity (see 2.12). As already discussed for the Mean-Var case, sub-additivity is essential for ensuring a diversified portfolio, which is translated into a minimization of risk. Value-at-Risk fails to meet this criterion, with the only exception being represented by variables that exhibit an elliptic joint distribution [14]. A formal proof of this violation involves showing that there exist scenarios where the VaR_β of a portfolio is greater than the sum of the VaR_β of its individual components, hence by providing a counter example. A proof is discussed by [15] in Example 2.25, where a discrete scenario is analyzed.

Furthermore, VaR_β exhibits non-convexity (see (2.13)), implying the potential existence of multiple local extrema. This non-convex nature can pose challenges in the optimisation process [12]. In addition, since it is only dependant on the chosen value of β, its value might be strongly affected by the choice of the confidence level [13]. Given these limitations, an alternative risk measure, introduced by [16] in 1999, offers a viable choice: the Conditional Value at Risk (CVaR), [12], [14], [16].

2.3.2.1 Analytical Computation of VaR: Elliptical case

If we have normally distributed set of returns (see E.2.1), we can analytically compute the Value at Risk of the portfolio considered. If we set a confidence level β, the VaR_β of a normal random variable can be computed as:

$$\mathrm{VaR}_\beta(R) = \Phi^{-1}(\beta)\sigma_p + \mu_P, \tag{2.17}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal CDF. The proof is simple, and is provided in Section A.1.4 of the Appendix.

More in general, we are able to derive analytically the Value-at-Risk of elliptical distributed random vectors $\mathbf{X} = (X_1, \dots, X_N) \sim \mathrm{El}(\boldsymbol{\mu}, \Sigma, g)$, with mean vector $\boldsymbol{\mu}$, covariance matrix Σ, and g as the density generator. The density is therefore:

$$f(x) = |\Sigma|^{-\frac{1}{2}} g\left((x-\boldsymbol{\mu})^\top \Sigma^{-1}(x-\boldsymbol{\mu})\right).$$

The VaR_β of elliptically distributed variables is:

$$VaR_\beta = -\mathbf{w}^\top \boldsymbol{\mu} + q^g_{\beta,N}\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}, \tag{2.18}$$

where $q^g_{\beta,N}$ is the solution to:

$$G(s) = \frac{\pi^{\frac{N-1}{2}}}{\Gamma\left(\frac{N-1}{2}\right)} \int_s^{-\infty} \int_{z_1^2}^{+\infty} \left(u - z_1^2\right)^{\frac{N-3}{2}} g(u) du dz_1.$$

The proof is available at [17]. The result is coherent with the one obtained in (2.17), since a normal variable is a special case of elliptical with density generator:

$$g(z) = (2\pi)^{-\frac{N}{2}} e^{-\frac{z}{2}}.$$

2.3.3 Conditional Value-at-Risk

The Conditional Value-at-Risk (CVaR) is defined as "the expected value of losses exceeding VaR_β" [14]. If we consider a vector $\mathbf{w}$ of the portfolio's weights, a random vector $\mathbf{R} \in \mathbb{R}^N$ of returns, then the formal definition of $CVaR_\beta$ is the following:

Definition 2 (Conditional Value at Risk (CVaR))**.** *The Conditional Value-at-Risk (CVaR) of the random vector $\boldsymbol{R}$, with confidence level $\beta \in (0,1)$ is the expected loss in the $\beta-$tail with respect to the joint distribution of returns:*

$$CVaR_\beta(L) = \frac{1}{1-\beta} \int_{L(\boldsymbol{w},\boldsymbol{R}) \geq VaR_\beta(L)} L(\boldsymbol{w}, r) f(r) dr, \tag{2.19}$$

where $L(\mathbf{w}, \mathbf{R})$ is the loss function, and $f(r)$ is the probability density function of $\mathbf{R}$ (see Def. 30), [14]. An alternative formulation of the $CVaR_\beta$ is also provided by [14], which makes the interpretation of the measure being an expected value evident:

$$CVaR_\beta = VaR_\beta + \mathbb{E}[L(\mathbf{w}, \mathbf{R}) - VaR_\beta | L(\mathbf{w}, \mathbf{R}) > VaR_\beta]. \tag{2.20}$$

In contrast to Value-at-Risk, $CVaR_\beta$ possesses some interesting properties. It is, in fact, continuous with respect to the confidence level β, and is jointly convex in (L, β) [13]. As already highlighted, it is a convex risk measure, which ensures that efficient algorithms can reliably find the optimal solution [18]. However, Definition 2.20 still necessitates the computation of VaR_β to derive the $CVaR$ measure, which is not ideal for the previously outlined reasons. This has led to the development of an alternative formulation that is simpler, yet equivalent to $CVaR_\beta$ in optimisation problems. We can, in fact, define a function $F_\beta(L, \alpha)$, which depends on a new variable α:

$$F_\beta(L, \alpha) = \alpha + \frac{1}{1-\beta} \int_{L(\mathbf{w},\mathbf{R}) \geq \alpha} (L(\mathbf{w}, r) - \alpha) f(r) dr. \tag{2.21}$$

This new formulation has a set of desirable properties that are outlined by the following two theorems, derived by Rockafellar and Uryasev [16].

Theorem 2. *As a function of α, $F_\beta(L, \alpha)$ is convex and continuously differentiable. The $CVaR_\beta$ of the loss associated with any return vector $\boldsymbol{R} \in \mathbb{R}^N$ can be determined from the formula:*

$$CVaR_\beta(L) = \min_{\alpha \in \mathbb{R}} F_\beta(L, \alpha). \tag{2.22}$$

In this formula the set consisting of the values of α for which the minimum is attained, namely

$$A_\beta(L) = \arg\min_{\alpha \in \mathbb{R}} F_\beta(L, \alpha), \tag{2.23}$$

is a nonempty, closed, bounded interval (perhaps reducing to a single point), and the VaR_β of the loss is given by

$$VaR_\beta(L) = \textit{left end point of } A_\beta(L). \tag{2.24}$$

In particular, one always has:

$$VaR_\beta(L) \in \arg\min_{\alpha\in\mathbb{R}} F_\beta(L,\alpha) \textit{ and } CVaR_\beta(L) = F_\beta(L, VaR_\beta(L)). \tag{2.25}$$

From Theorem 2 we gain the convexity and differentiability of the newly defined function F_β, which make the function ideal for minimization purposes. Computing the VaR_β becomes not necessary anymore, but the Value at Risk can still be derived from F_β. The proof of this theorem is taken from [16], and is discussed in Section A.1.1 of the Appendix.

Theorem 3. *Minimizing the $CVaR_\beta$ of the loss associated with the weights $\boldsymbol{w}$ over all $\boldsymbol{w} \in \mathcal{X}_w$, where $\mathcal{X}_w$ represents the feasible set, is equivalent to minimizing $F_\beta(L,\alpha)$ over all $(w,\alpha) \in \mathcal{X} \times \mathbb{R}$, in the sense that*

$$\min_{w\in\mathcal{X}_w} CVaR_\beta(L) = \min_{(w,\alpha)\in w\in\mathcal{X}_w\times\mathbb{R}} F_\beta(L,\alpha). \tag{2.26}$$

where moreover a pair $(\boldsymbol{w}^*, \alpha^*)$ achieves the second minimum if and only if $\boldsymbol{w}^*$ achieves the first minimum and $\alpha^* \in A_\beta(\boldsymbol{w}^*)$. Therefore, in circumstances where the interval $A_\beta(L^*)$ reduces to a single point (as is typical), the minimization of $F_\beta(L,\alpha)$ over $(\boldsymbol{w},\alpha) \in \mathcal{X}_w \times \mathbb{R}$ produces a pair $(\boldsymbol{w}^*;\alpha^*)$, not necessarily unique, such that $\boldsymbol{w}^*$ minimizes the $CVaR_\beta$ and α^* gives the corresponding VaR_β. Furthermore, $F_\beta(L,\alpha)$ is convex with respect to $(\boldsymbol{w},\alpha)$, and $CVaR_\beta(L)$ is convex with respect to w, when $L(\boldsymbol{w},\boldsymbol{R})$ is convex with respect to $\boldsymbol{w}$, in which case, if the constraints are such that $\mathcal{X}_w$ is a convex set, the joint minimization is an instance of convex programming.

From Theorem 3, whose proof is available in [16]'s Appendix, several interesting properties emerge. Firstly, it is important to note that optimizing $F_\beta(L)$ is equivalent to optimizing the Conditional Value at Risk, which significantly simplifies the problem. This problem falls under the category of convex stochastic optimisation, which has been extensively addressed in the literature [16]. Another observation is that the convexity of the loss function $L(\mathbf{w},\mathbf{R})$ implies the convexity of both $F_\beta(L)$ and, most importantly, $CVaR_\beta(L)$.

However, the $CVaR_\beta$ risk measure also presents some disadvantages. It is more sensitive to errors, and is highly affected by the accuracy of tail modelling [13].

2.3.3.1 Analytical Computation of Conditional VaR

Analogously to the VaR_β case, we can analytically derive the $CVaR_\beta$ of a of a portfolio composed by Gaussian returns [19]. Hence, the analytical formulation is:

$$\text{CVaR}_\beta = \frac{1}{1-\beta}\varphi\left(\Phi^{-1}(\beta)\right)\sigma_P - \mu_P, \tag{2.27}$$

where $\varphi(\cdot)$ is the density of a standard normal, and $\Phi^{-1}(\cdot)$ the inverse of the CDF of a standard normal. The proof is provided in the Appendix, in Section A.1.5.

2.3.3.2 CVaR's Problem Simplification

Differently from the Mean-Var case, finding solutions to the $CVaR_\beta$ optimisation problem in non Gaussian scenarios can be complex, since we cannot derive an analytical formulation of the efficient frontier.

Luckily, the $CVaR_\beta$ offers a simplification that allows to solve the problem with an equivalent formulation through linear programming optimisation. Starting from the equivalent formulation of (2.21), [16] introduced its discretized version:

$$\tilde{F}_\beta(L, \alpha) = \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} (L(\mathbf{w}, \mathbf{R}_k) - \alpha) \mathbb{1}_{L(\mathbf{w}, \mathbf{R}_k) \geq \alpha}, \tag{2.28}$$

which is obtained by sampling from the distribution of the return vector $\mathbf{R}$, using a sample set $\mathbf{R}_1, \ldots, \mathbf{R}_q$. $\tilde{F}_\beta(L, \alpha)$ can be further simplified into a linear expression using an auxiliary variable u_k, subject to the linear constraints $u_k \geq 0$ and $L(\mathbf{w}, \mathbf{R}_k) + \alpha + u_k \geq 0 \ \forall k = 1, \ldots, q$:

$$\alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} u_k,$$

becoming a linear programming problem. For each sampled vector $\mathbf{R}_k$, the variables u_k corresponds to the quantity $[L(\mathbf{w}, \mathbf{R}_k) - \alpha]^+$, since the constraints imposed guarantee that the variable is located in the space where the constraint $L(\mathbf{w}, \mathbf{R}_k) \geq \alpha$ is satisfied.

The new formulation is much easier to optimize, and can be used also for non-normal distributions. Problem (2.10) therefore becomes:

$$\begin{cases} \min_{w,\alpha,u} \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} u_k \\ s.t. \ L(\mathbf{w}, \mathbf{R}_k) + \alpha + u_k \geq 0 \ \forall k = 1, \ldots, q \\ s.t. \ u_k \geq 0 \ \forall k = 1, \ldots, q \\ s.t. \ \mathbf{w}^T \boldsymbol{\mu} = r_0 \\ s.t. \ \sum_{i=1}^{N} w_i = 1 \\ s.t. \ w_i \geq 0 \\ s.t. \ \alpha \in [0, 1]. \end{cases} \tag{2.29}$$

2.4 The Tangent Portfolio

The selection of a specific point among the collection of optimal ones of the efficient frontier, depends on the level of risk that an investor is willing to take. In this subsection, we will discuss different methods for choosing an optimal solution, aiming to find a balance between a satisfying expected return, and managing the level of risk to which the chosen portfolio is exposed.

The general approach is what is defined as the *reward-risk ratio*, hence a method that implies the choice of the portfolio based on two factors: the reward (the expected return in the scenario that we consider), and the risk. [20], [21] propose an overview of some of the ratios available. Among these, the Sharpe Ratio, introduced by Sharpe in 1966 in [22], stands out as one of the most widely adopted.

Given a risk measure $\rho(\cdot)$, the corresponding reward-risk ratio, can be defined as:

$$\frac{\mathbf{w}^T\boldsymbol{\mu} - r_f}{\rho(\mathbf{w}^T r)}, \tag{2.30}$$

where r_f is the risk free rate, defined as the return on an investment that is considered to have no risk of financial loss.

The portfolio on the efficient frontier corresponding to the maximum ratio is selected as the optimal portfolio. This portfolio is commonly referred to as the *ρ-tangent portfolio*, as geometrically, the solution is defined by the point where the efficient frontier intersects the line representing the maximum ratio. This line passes through the risk-free rate point, emphasizing the trade-off between risk and return in portfolio selection [20]. An example is shown in Figure 2.4, where the risk measure $\rho(\cdot)$ is the Variance, hence corresponding to the Sharpe ratio scenario.

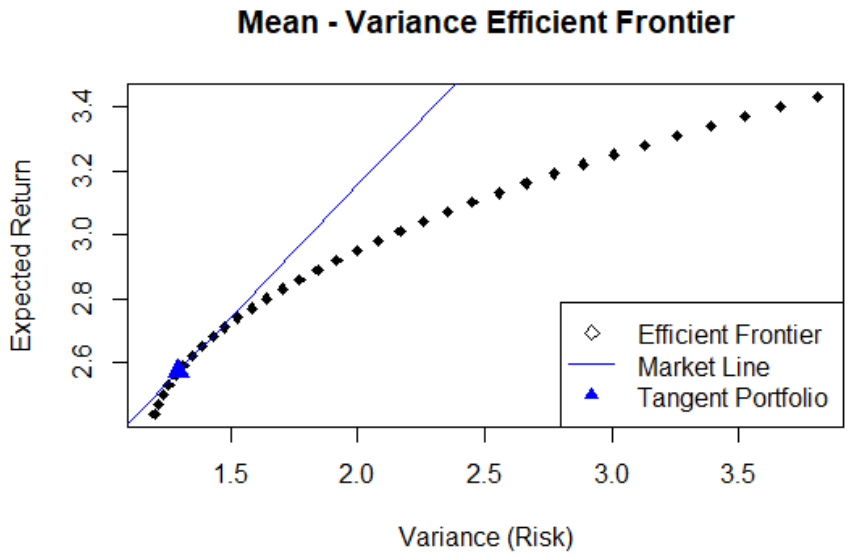


Figure 2.4: *Efficient frontier of a Mean-Variance problem, computed optimising a portfolio of two normally distributed returns. The tangent portfolio and the Capital Market Line, computed with a risk free rate $r_f = 1.5$.*

In the following Sections, we will explore the Sharpe ratio, generally associated to the Mean-VaR problem, and the STARR ratio, which finds utility in the context of Conditional Value at Risk (CVaR). For Value at Risk, there is not a specific ratio identified, but in practical application we will deploy the formula presented in (2.30).

2.4.1 Sharpe Ratio

Given a risk-free rate r_f, we can define the Sharpe ratio as:

$$SR = \frac{\mu_P - r_f}{\sigma_P}, \tag{2.31}$$

where μ_P is the expected return of the portfolio, and σ_P is its standard deviation [23]. The higher the value of the ratio, the better the performance of the portfolio.

The Sharpe Ratio can be used to select the best portfolio in the Mean-Variance scenario. The general approach is to find the tangent portfolio, hence the point of the efficient frontier that is tangent to the Capital Market Line (CML). The the CML is defined as the line that maximises the reward-risk ratio on a risk-free adjusted environment. The analytical derivation of this tangent portfolio will be discussed in 2.3.1.1.

2.4.2 STARR Ratio

Analogously to the Mean-Var context, we can define the $STARR$ Ratio in the context of Conditional value at risk minimization. Following the general formula discussed in (2.30), for a given confidence level β, the Ratio is obtained by computing:

$$STARR_\beta = \frac{\mu_P - r_f}{CVaR_\beta(\mathbf{w}, \mathbf{R})}, \tag{2.32}$$

where again μ_P is the expected return of the portfolio, and r_f the risk-free rate [20].
The objective is to identify the portfolio that maximizes the STARR Ratio, corresponding to the $CVaR-$tangent portfolio. It is noteworthy that different values of β lead to distinct results [21].

2.5 The Delta-Normal Method

The risk measures can be obtained through different methods. One such method is the delta-normal technique, which is based on the assumption that the underlying assets follow a multivariate normal distribution. This method makes the computation of VaR_β and $CVaR_\beta$ extremely straightforward (see Sections 2.3.2.1, 2.3.3.1), yet it relies on the assumption of data normality. The delta method is straightforward to compute, it is often less applicable in financial contexts where asset's returns behaviors generally are not correctly represented by a normal distribution. This considerations set the stage for the simulation study discussed in the following Section 2.6.

Alternatively, historical data can be used for simulation, assuming that future observations will mimic past behavior. Monte Carlo Simulations provide an alternative methodology, simulating data from a chosen statistical distribution believed to accurately describe the behaviour of the returns of an asset [24].

2.6 Simulation Study

Having discussed the theory and the computations behind portfolio theory, we propose a small simulation study.

In this study, we simulate $N = 1000$ data points, representing two asset's returns, from a Gaussian distribution: $R_1 \sim N\left(\mu_1, \sigma_1^2\right)$ and $R_2 \sim N\left(\mu_2, \sigma_2^2\right)$. Consequently, the portfolio obtained as a linear combination of the two, will have a return of $R = w_1 R_1 + w_2 R_2$, which will be again normally distributed. The goal of the optimisation process is to determine the value of the two weight, by deploying the different methodologies presented in the previous Chapters. The results obtained will then be compared, highlighting the differences between the different risk measures.

The selected parameters are the following $\boldsymbol{\mu}$ and Σ:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 15 \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} 10^2 & 0.7 * 15 * 10 \\ 0.7 * 15 * 10 & 15^2 \end{bmatrix}.$$

The choice of the parameters is dictated by a need of exploring two cases that should mimic the behaviour of a financial index: we generate an asset with low expected return, and low volatility, which corresponds to a less risky asset, and one that has higher expected value but also higher volatility. The portfolio obtained by combining the two assets is a normally distributed random variable with mean:

$$\mu_P = w^T \boldsymbol{\mu},$$

and variance

$$\sigma_P^2 = \mathbf{w}^T \Sigma \mathbf{w}.$$

The simulated data are visualized in the following Figure 2.5, along with the estimated Pearson Correlation between the two variables.

Figure 2.5: *Distribution of the two simulated asset's returns, drawn from two Gaussian distributions. A scatterplot of the data, highlighting their linear correlation is also provided (refer to section 3.2 for additional details on correlation).*

After having simulated the two returns, we started studying the implications of following an Equal-Weight approach (the weights are simply set at 50% for each asset). We proceeded by studying the frontiers obtained from the Mean-Variance approach. Lastly, we studied the effect of choosing risk measures like VaR_β and $CVaR_\beta$, discussed in Sections 2.3.2, and 2.3.3, to select the optimal set of weights. Reward-to-Risk Ratios are then applied to select the best portfolio from each method, and to compare the results.

The risk free rate was set to be $r_f = 6$, coherently with the mean and volatility value of the assets

considered.

2.6.1 Simulation-Based Calculation of the Mean-Variance Efficient Frontier

After having estimated the mean $\boldsymbol{\mu}$, and the covariance matrix Σ from the simulated data, following the steps discussed in Section 2.3.1, we started by computing the target return associated with minimum variance, which corresponds to the starting value of the vector of target returns r_0. This return, r_{min} is obtained from $r_{min} = \mathbf{w}_{min}^T \boldsymbol{\mu}$, where $\mathbf{w}_{min}$ is the vector of weights that solves the following quadratic minimization problem:

$$\begin{cases} \min_{\mathbf{w}} & \mathbf{w}^T \Sigma \mathbf{w} \\ \text{s.t.} & \mathbf{w}^T \mathbb{1} = 1. \end{cases}$$

Once we obtained the value of r_{min}, we created a vector of target returns which was used to generate the efficient frontier from the parabola that we derived in Equation 2.15. The results obtained are presented in Figure 2.6, where on the y-axis we find the target returns, and on the x-axis the corresponding Variance. The results are compared with the mean and variance of the Equal-Weight portfolio, hence the portfolio whose weights are $w_i = \frac{1}{N}$.

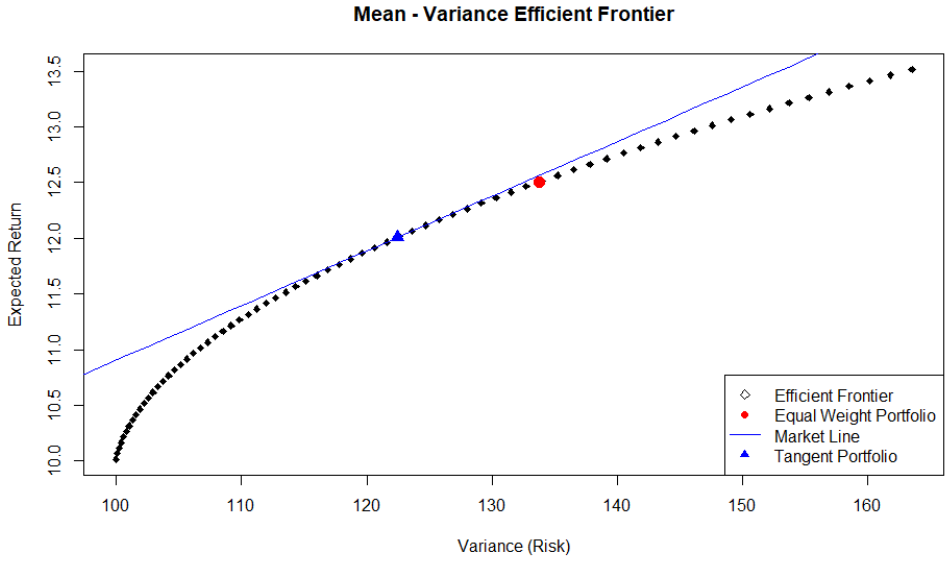


Figure 2.6: *Mean-Variance efficient frontier of the normally distributed asset's returns, where the tangent portfolio (blue triangle) is identified as the intersection of the frontier with the market line. The portfolio is compared to the equal weight portfolio (red dot).*

Subsequently, we set the risk free rate $r_f = 6$, and found the Tangent Portfolio by solving Equation (1). The portfolio is the point that maximises the Sharpe Ratio (formula 2.31) among the ones of the efficient frontier. For this example, the portfolio is presented as a blue triangle in Figure 2.6. We can, in fact, see how the point is located on the frontier, and is indeed obtained from the intersection of the tangent line (blue) with the curve of optimal points. The values of the tangent portfolio are:

$$\begin{cases} \mu_{tp} = 12.004 \\ \sigma_{tp}^2 = 122.486. \end{cases}$$

The results obtained will be compared to the ones of the other methods in Section 2.6.4. In plot 2.6 we compare the resulting portfolio to the equal weights one (red dot in the Figure): both points are located on the frontier, but Markovitz' method prefers a less risky combination.

2.6.2 Simulation-Based Calculation of the Mean-VaR_β Efficient Frontier

As previously discussed in Section 2.2.2, the efficient frontier in the Mean-Value at Risk framework can be derived by solving the optimisation problem of 2.9, by setting VaR_β as the risk measure, and solving the problem for a set of target returns. Since we are in a Gaussian setting, we can simplify the computations by deploying the analytical formulation of the value at Risk introduced in Equation (2.17).

We determined the optimal set of weights that minimizes the corresponding VaR_β, for each target return, obtaining the frontier in Figure 2.7. The optimisation was carried out by applying a built in function of R's package "*fPortfolio*" [25], "*portfolioFrontier*", which uses a quadratic programming algorithm (or QP, a nonlinear programming algorithm used to optimize multivariate quadratic functions subject to linear constraints).

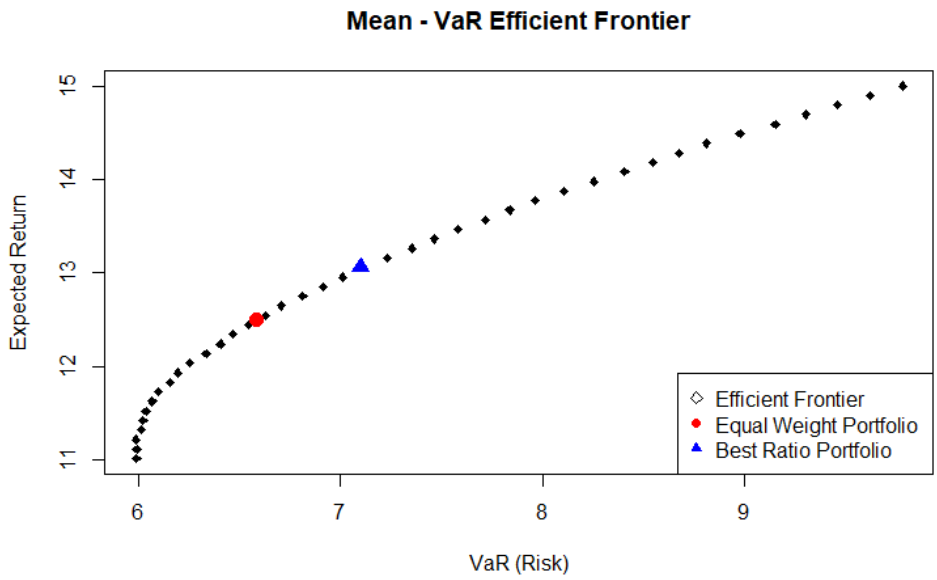


Figure 2.7: *Efficient frontiers for the Mean-Value-at-Risk case, with the optimal reward-to-risk portfolio (blue), and compared with the equal weight (red).*

Among the points obtained from the optimisation process, we selected the portfolio that maximised the Reward-to-Risk ratio, with a $r_f = 6$, hence the point i for which the following was maximised:

$$\frac{\mu_i - r_f}{VaR_\beta(\mathbf{w}_i, \mathbf{R})},$$

μ_i representing the return of the i-the portfolio, and $\mathbf{w}_i$ the associated set of weights.

The best portfolio is identified by a blue triangle, and has values:

$$\begin{cases} \mu_{tp} = 13.059 \\ VaR_{0.95} = 7.099. \end{cases}$$

Again, the EW point, which appears as a red dot in the plot, is located on the frontier, but this time is less risky than the one preferred by the mean-VaR method.

2.6.3 Simulation-Based Calculation of the Mean-$CVaR_\beta$ Efficient Frontier

Analogously to the VaR_β case, we can also optimize a portfolio using the Conditional Value at Risk (CVaR) risk measure. This optimisation involves solving the problem introduced in (2.29), as this discretized problem is equivalent to solving the original one, but can be solved with linear programming.

In Figure 2.8 we can observe the parabolic efficient frontier derived. Here, the x-axis represents the risk measure under optimisation, hence the values of the CVaR, and on the y-axis the corresponding target returns.

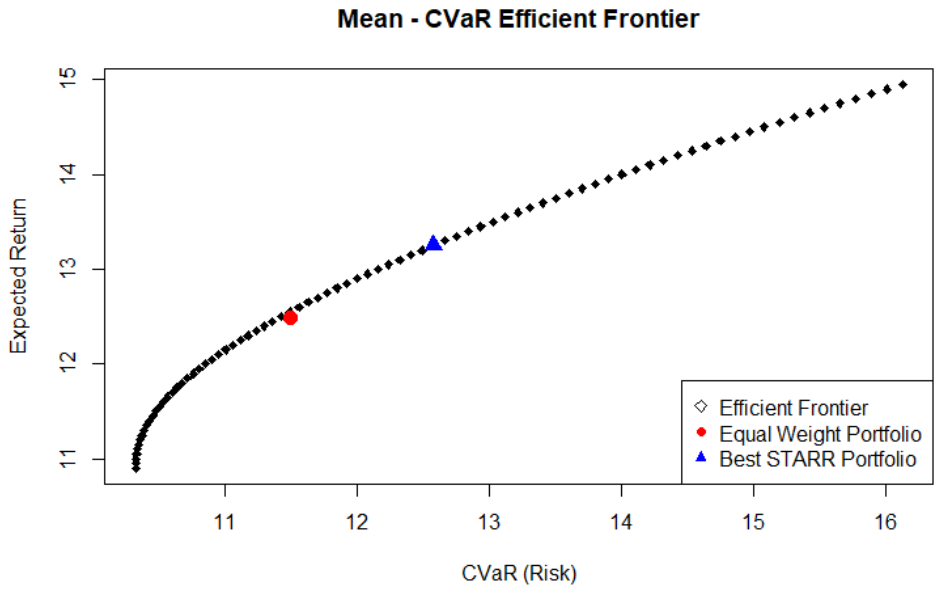


Figure 2.8: *Mean-Conditional Value-at-Risk efficient frontier, computed optimising a portfolio of two normally distributed returns. In blue we find the best STARR ratio portfolio, in red the EW portfolio.*

The optimal portfolio, highlighted as a blue triangle has the following values:

$$\begin{cases} \mu_{tp} = 13.209 \\ CVaR_{0.95} = 12.546. \end{cases}$$

The method does not provide an efficient frontier that is better than the EW portfolio (in red), since the portfolio is located on the frontier itself. However, this might be due to the fact that we considered only two assets. Similarly to the Mean-VaR, the chosen portfolio is riskier than the one obtained from setting the weights equal.

2.6.4 Results comparison

In the following two tables we will compare the portfolios obtained by solving the optimisation problem with the different methods discussed in Subsections 2.6.1, 2.6.2 and 2.6.3. The portfolios are also compared to an equal weight portfolio. Furthermore, we will examine two portfolios exposed to maximum and minimum risk, achieved by setting the weights equal to $w_1 = 100\%$ and $w_2 = 0\%$, and vice versa.

To compare the results, we compute the variance, the Value at Risk, and the Conditional Value at Risk of the portfolios selected from each method. Using the corresponding weights, we derive these risk measures through numerical methods, the results are provided in Table 2.1. For each method we computed the Sharpe and STARR ratios (eq. (2.31) and (2.32)). We also provide the time that the optimisation process took to compute the optimal portfolios.

In addition, given the Gaussian nature of the simulations, we are able to compute all the values discussed above with analytical formulas, as discussed in Section 2.5. Consequently, we present the analytical values in Table 2.2: this will allow to discuss the efficacy of applying numerical approximations to the risk measures. To give further meaning to this comparison, in Table 2.2 we used the actual parameters used to simulate the portfolios, hence the values of $\boldsymbol{\mu}$ and Σ provided at the beginning of this Section, instead of using values estimated from the simulated data.

Method	Weights	Mean	Var	VaR	CVaR	Sharpe	STARR	Time
EW	$(50\%, 50\%)$	12.495	133.779	6.574	11.488	4.8%	0.565	-
Mean-Var	$(59.9\%, 40.1\%)$	11.997	122.476	6.223	10.922	**4.9%**	0.549	**0.14 s**
Mean-VaR_{95}	$(38.8\%, 61.2\%)$	13.059	149.306	7.099	12.302	4.7%	0.574	1.74 s
Mean-CVaR_{95}	$(35.8\%, 64.2\%)$	13.209	153.939	7.286	12.546	4.7%	**57.5%**	53.5 s
Least Risky	$(100\%, 0\%)$	9.983	99.86	6.481	10.626	4.0%	0.375	-
Most Risky	$(0\%, 100\%)$	15.007	225.267	9.78	16.131	4.0%	0.558	-

Table 2.1: *Table comparing the portfolios obtained by optimizing the composition of two normally distributed returns of the two assets, using different methods: the Equal Weight, The Mean-Var, the Mean-$VaR_{0.95}$ and the Mean-$CVaR_{0.95}$. The table reports the weights and the results of the risk measures and ratios computed* ***numerically****.*

Method	Weights	Mean	Var	$\text{VaR}_{0.95}$	$\text{CVaR}_{0.95}$	Sharpe	STARR
Equal Weight	$(50\%, 50\%)$	12.5	133.75	6.582	11.154	4.8%	0.583
Mean-Var	$(59.91\%, 40.09\%)$	12.004	122.486	6.257	10.632	**4.9%**	0.565
Mean-$\text{VaR}_{0.95}$	$(38.77\%, 61.22\%)$	13.061	149.229	7.095	11.924	4.7%	0.592
Mean-$\text{CVaR}_{0.95}$	$(35.78\%, 64.22\%)$	13.211	153.849	7.255	12.158	4.7%	**59.3%**
Least Risky	$(100\%, 0\%)$	10	100	6.5	10.453	4.0%	0.383
Most Risky	$(0\%, 100\%)$	15	225	9.75	15.679	4.0%	0.574

Table 2.2: *Table comparing the portfolios obtained by optimizing the composition of two normally distributed asset's returns, using different methods: the EW, The Mean-Var, the Mean-$VaR_{0.95}$ and the Mean-$CVaR_{0.95}$. The table reports the weights and the results of the risk measures and ratios computed* ***analytically****.*

The results shown in Table 2.1, correspond to the expectations. We can observe that according to Sharpe Ratio, the best portfolio is obtained with Markovitz' method, while the best STARR is achieved by the portfolio obtained from the Mean-CVaR optimisation. The ratios for all methods are quite close to each other, indicating how the methods do not have big differences among their results and that they can be considered rather equivalent. We can however notice how the Mean-Variance portfolio

is significantly faster than the other methods, suggesting that for cases like this, we should prefer a more straightforward risk measure. This result is coherent with the previous discussion, emphasising that with simple distributions, such as normally distributed data, more simplistic approaches tend to outperform others.

On another note, if we compare the two tables, we can notice how the analytical values are not perfectly matching the ones obtained from the simulated data, but they are rather close. The discrepancy is rather small, and we can therefore conclude that $n = 1000$ are enough observations to guarantee a good estimate of the analytical result. In addition, we can see how the numerical results are able to maintain the risk hierarchy, and exhibit consistent relationships with VaR and CVaR across all considered measures. We can therefore conclude that we are satisfied with the results obtained. To improve the robustness of our analysis, if data availability allows it, we will try to increase the dataset size when working with the actual model.

Lastly, to have a better overview of how the weights influence the risk measures and the ratios, we computed these metrics for each possible combination of weights, ranging from the less risky $\mathbf{w} = (100\%, 0\%)$, to the most risky $\mathbf{w} = (0\%, 100\%)$. The risk measures and ratios were calculated as functions of weight w_1 (the weight of the first variable), and produced the plots shown in Figure B.1. We can notice how all three measures exhibit a decreasing trend as the weight assigned to the first variable increases, in accordance with the mean, which has a linear behaviour. Variance, VaR and CVaR have a quadratic behaviour. Additionally, we studied the evolution of the relative Sharpe and STARR Ratio: they have a concave shape, where we can clearly identify a maximum, corresponding to the optimal portfolio.

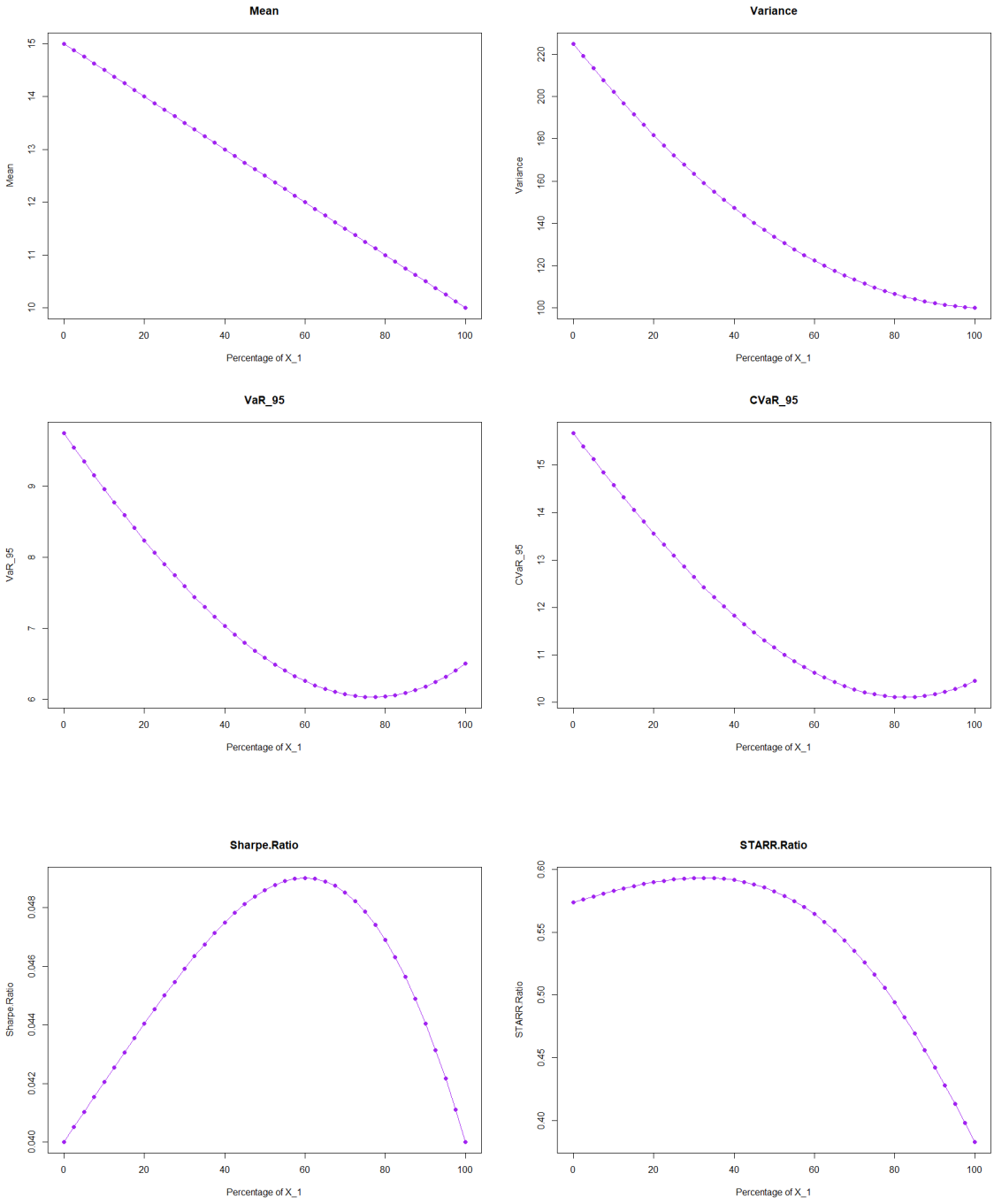


Figure 2.9: *Plots showcasing the evolution of the risk measures considered, and the relative ratios in relation to the change of weights. In particular, in the X-axis we find the percentage of the first asset (less risky).*

2.6.5 The Exponential Case

Normally distributed variables offer a series of advantages that are considered ideal. As discussed earlier, the analytical formulation is readily available for all three risk measures under consideration. This is the main reason why methods such the Delta Method, which considers the assumption that the underlings follow a normal distribution, is often used.

However, the normality assumption is not a characteristic typically observed in financial time series,

as we will see in in Chapter 8. Financial data often exhibit asymmetries and high-tail distribution, phenomena that are inadequately represented by a normal distribution. Consequently, the desirable properties associated with normality cannot be fully deployed. Methods such as Markowitz's Mean-Var prove to be too simplistic, failing to fully capture all information carried by the time series. More sophisticated models and risk measures, like Conditional VaR, become necessary to effectively address the optimisation problem under consideration.

In light of these considerations, we re-propose the same simulation, but with a different underlying distribution. We simulate n=1000 data points from two exponential distributions with parameters:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{10} \\ \frac{1}{15} \end{bmatrix},$$

hence by generating two returns that have the same mean and variance of the assets considered on the previous experiment (the mean of an exponentially distributed variable, $X \sim Exp(\lambda)$, is given by $\mathbb{E}[X] = \frac{1}{\lambda}$, while $\sigma^2 = \frac{1}{\lambda^2}$). Since, exponential data are only positive, to make sure the simulation study had financial sense, we translated the data by 2 units on the left, allowing for negative values. The resulting data are show in in the following figure 2.10.

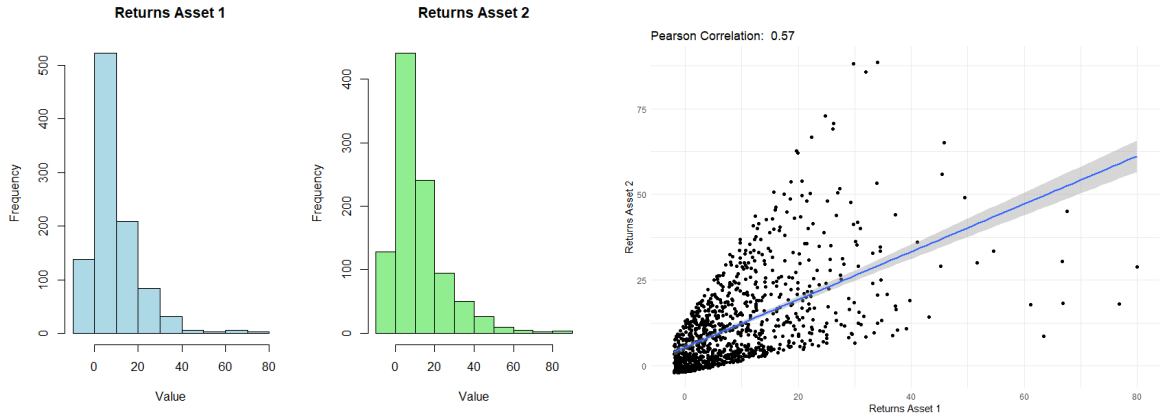


Figure 2.10: *Distribution of the simulated data, drawn from two Exponential distributions with parameters presented above. A scatterplot of the data, highlighting their linear correlation is also provided (see section 3.2).*

In Section 2.6.5 of the Appendix we provide the results for the exponential case, which will be discussed here. In particular, in Figure B.3 we have the three graphs obtained from the three optimisation methods, while in Figure B.2 we provide the graphs showcasing the evolution of the risk measures and the ratios in relation with the weight w_1, equivalently to what we discussed for the Gaussian case.

In particular, for each optimisation technique we obtain the following results.

2.6.5.1 Mean - Variance

The optimal portfolio chosen from the Mean - Variance frontier, by comparing the Sharpe ratios of each point, has the following values:

$$\begin{cases} \mu_{tp} = 10.473 \\ \sigma^2_{tp} = 124.634, \end{cases}$$

the portfolio is the tangent portfolio in blue, and is characterized by low exposure to risk. We can notice how the chosen portfolio has almost the same values as the EW portfolio, represented as a red dot, which is also located on the frontier. The result is coherent with Figure B.2, where the maximum Sharpe ratio is located in correspondence to $w_1 = 50\%$.

2.6.5.2 Mean - Value at Risk

The optimal portfolio for the Mean- VaR frontier has the following values.

$$\begin{cases} \mu_{tp} = 10.942 \\ VaR_{0.95} = 139.114. \end{cases}$$

The selected portfolio has higher risk than the ones selected with an EW approach, or the Mean-Var one, as it is noticeable with comparing the point with the EW (red dot). However, this method still is not characterized by a high exposure to risk, preferring a rather low return with minimal risk.

2.6.5.3 Mean - Conditional Value at Risk

Lastly, the Mean - $CVaR$ optimal portfolio is characterized by the following values:

$$\begin{cases} \mu_{tp} = 11.725 \\ CVaR_{0.95} = 177.932. \end{cases}$$

This specific portfolio has weights $\mathbf{w} = (7.68\%, 92.32\%)$, a result that is significantly higher in risk than the ones from other methods considered. This choice is better explained in Figure B.2, where we can see that the STARR curve reaches its peak almost at the beginning, hence where the weight w_1 is low, and then starts a constant decrease, while for the Sharpe ratio the peak is reached around the EW.

2.6.5.4 Result Comparison

Here we present the table summarizing the results presented for the exponential case, and comparing the four methods. Differently from the Gaussian case presented in Table 2.2, we do not have an analytical computation for the different measures. We therefore computed the values on the table from the simulated data using only numerical methods.

Method	Weights	Mean	Var	VaR	CVaR	Sharpe	STARR	Time
Equal Weight	$(50\%, 50\%)$	10.499	125.265	0.912	1.384	**3.59**%	3.251	-
Mean-Var	$(50.90\%, 49.10\%)$	10.473	124.634	0.912	1.383	3.58%	3.235	0.18 s
Mean-VaR_{95}	$(34.69\%, 65.31\%)$	10.942	139.114	0.965	1.423	3.55%	3.472	0.72 s
Mean-CVaR_{95}	$(7.68\%, 92.32\%)$	11.725	177.932	1.193	1.581	3.21%	**3.62**	27.8 s
Least Risky	$(100\%, 0\%)$	9.051	121.12	1.421	1.715	2.51%	1.779	-
Most Risky	$(0\%, 100\%)$	11.947	192.333	1.286	1.649	3.09%	3.607	-

Table 2.3: *Table comparing the portfolios obtained by optimizing the composition of two exponentially distributed returns, using different methods: the EW, The Mean-Var, the Mean-$VaR_{0.95}$ and the Mean-$CVaR_{0.95}$. The table reports the weights and the results of the risk measures and ratios computed* ***numerically****.*

It is immediate to notice that, the Equal Weight portfolio outperforms the Mean-Variance portfolio in terms of achieving the best Sharpe ratio. Nevertheless, the discrepancy between the two portfolios is minimal, as we have already discussed. Contrarily, the optimal STARR ratio aligns with the portfolio

selected by the Mean-CVaR method, which is consistent with our expectations.

By comparing the computational times, we can observe how Markovitz' method is the fastest, as we expected, but all methods reach a result in a few seconds, making the time not a relevant factor in the method's choice for such amount of observations.

In addition, as for the normal scenario, since we are in two dimensions, we can study the evolution of the variables according to the weights. The results are available in the Appendix, in Figure B.2. The mean and variance plot are similar to the Gaussian case. However, the VaR and CVaR preserve the quadratic shape, but are not only decreasing anymore. This implies that a lower return is associated with a higher risk: the frontier, in this case, is non-bijective, resulting in some portfolios to be inefficient: this is the case of the "Least Risky" portfolio, where the value of the VaR and CVaR is higher than all other methods, even if the expected value is lower. The Sharpe ratio maintains the shape discussed in the previous section, and we can identify the best portfolio as the peak of the curve.

From this example it is still challenging to draw conclusions on which method is best. While the Mean-Variance method stands out for its computational efficiency and simplicity, the difference is still marginal. A small indication of potential limitations in Markowitz's method arises as it is slightly outperformed by the EW portfolio. However, it is important to keep in mind that this analysis is conducted in a bidimensional setting, which results in all portfolios to lie on the same frontier, as each point can be written as a function of one weight w_i (the other one is simply given by $1 - w_i$). This is clear from the Figures in B.2.

For this reasons, we decided to expand this simulation study to three dimensions, introducing a controlled yet more complex environment for the optimisation procedure. In particular, we will discuss a 3-dimensional portfolio with Gaussian distributed returns, and one further advancement introducing copulas. While the optimisation of VaR and CVaR in this expanded setting may result in longer computational times, it offers an opportunity to study potential advantages of these two methods.

2.6.6 Gaussian 3-dimensional case

For the three-dimensional case we simulate three normal variables with the following parameters:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 12 \\ 15 \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} 10^2 & 0.7 * 10 * 12 & 0.6 * 10 * 15 \\ 0.7 * 10 * 12 & 12^2 & 0.8 * 12 * 15 \\ 0.6 * 10 * 15 & 0.8 * 12 * 15 & 15^2 \end{bmatrix}.$$

The simulated variables, and their Pearson correlations are shown in the following Figures 2.11.

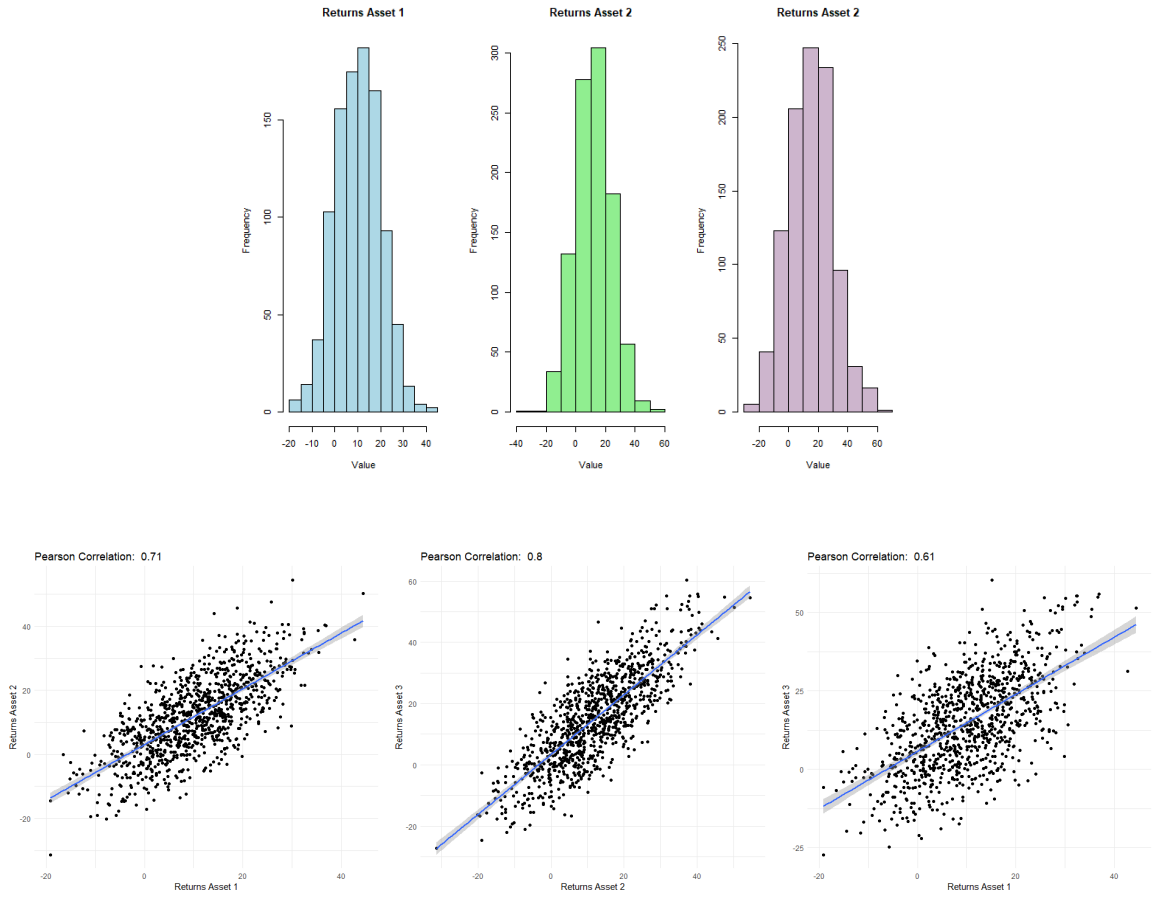


Figure 2.11: *Distribution of the simulated data, drawn from two Exponential distributions with parameters presented above. Three scatterplots of the data, highlighting their linear correlation is also provided (see section 3.2).*

Again we derive the three optimal portfolios using the three risk measures, and compare the results with the EW portfolio, the most risky portfolio and the least risky one. The results are shown in Table 2.4, and in Figure B.4 of the Appendix.

Method	Weights	Mean	Var	VaR	CVaR	Sharpe	STARR	Time
EW	$(33.3\%, 33.3\%, 33.3\%)$	12.22	123.088	5.923	10.74	5.05%	57.9%	-
Mean-Var	$(51.7\%, 16.1\%, 32.2\%)$	11.82	112.872	5.574	10.123	**5.15**%	57.4%	0.17 s
Mean-VaR_{95}	$(41.4\%, 12.7\%, 45.9\%)$	12.436	126.835	5.981	10.836	5.07%	59.3%	0.52 s
Mean-$CVaR_{95}$	$(39.9\%, 4.9\%, 55.2\%)$	12.742	135.574	6.33	11.308	4.97%	**59.6**%	42.1 s
Least Risky	$(100\%, 0\%, 0\%)$	9.903	100.658	6.612	10.753	3.87%	36.2%	-
Most Risky	$(0\%, 0\%, 100\%)$	14.867	224.796	9.743	16.036	3.94%	55.2%	-

Table 2.4: *Table comparing the portfolios obtained by optimizing the composition of three normally distributed returns, using different methods: the Equal Weight, The Mean-Var, the Mean-$VaR_{0.95}$ and the Mean-$CVaR_{0.95}$. The table reports the weights and the results of the risk measures and ratios computed* ***numerically****.*

Observing the results in the Table, we can highlight how the Mean-Variance method tends to favor a less risky portfolio, while using the Conditional value at risk results in the choice of a more risky portfolio. Despite this difference, the Sharpe ratio is still maximised by Markovitz' portfolio, while the best STARR ratio is reserved to the Mean-CVaR one. We can see how all portfolio exhibit very similar ratio values, suggesting that the portfolios might be equivalent.

In addition, also for this simulation, we have that the computational time required to optimize the CVaR portfolio is higher. However, it remains manageable, taking less than a minute, suggesting that we cannot select our preferred method solely based on computational efficiency.

What becomes interesting is the observation of the plots in Figure B.4. Here the CVaR method stands out as the only one to outperform the Equal Weight (EW) portfolio, as the red dot does not lie on the frontier, but in the inefficient surface. On the contrary, we can see how for the Mean-Variance case, the EW lies almost outside, outperforming the method.

To better compare the different methods, we plotted all optimal portfolios with to the optimal Mean-CVaR frontier in the following Figure 2.12. This visualization highlights how the EW and the Mean-VaR portfolios are outperformed by the points on the efficient frontier. However, it is still not possible to determine which of the other two methods is better as they lie both on the frontier.

It is important to remember that we are in a Gaussian scenario. To further explore the dynamics and complexities, we introduce copulas.

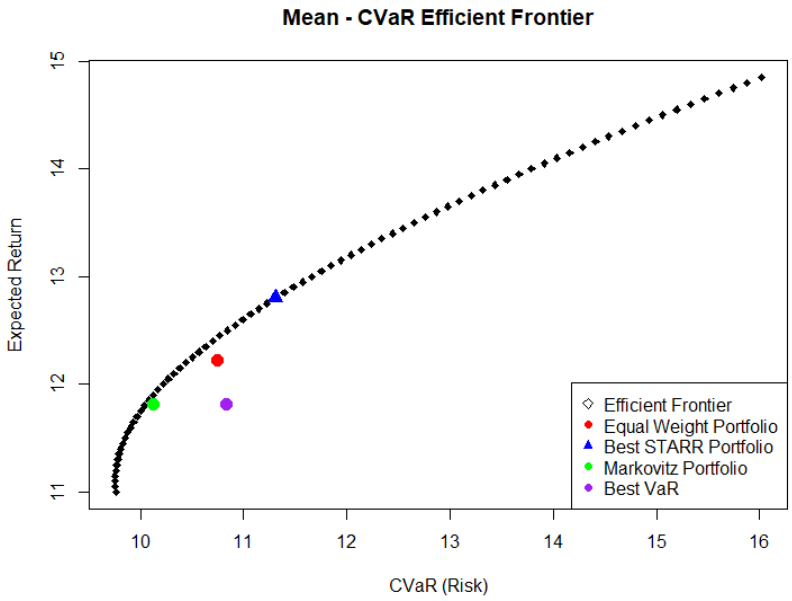


Figure 2.12: *Mean-CVaR efficient frontier, computed optimising a portfolio of three normally distributed returns. In blue we find the best STARR ratio portfolio, in red the EW portfolio, while in purple the best Mean-VaR, and in green the best portfolio according to Markovitz.*

2.6.7 Copula case

Lastly, to add some complexity and to introduce the topic that will be discussed throughout this thesis, we analyze a further simulation that involves the use of copulas, which will be introduced in the following Chapter 3.

We simulated a three dimensional Gaussian copula with correlation matrix (see 3.3.2):

$$R = \begin{bmatrix} 1 & 0.7 & 0.6 \\ 0.7 & 1 & 0.8 \\ 0.6 & 0.8 & 1 \end{bmatrix},$$

which guarantees that the simulated variables have the same correlation as the previous example, as we can notice from the Pearson correlations shown in Figure 2.13. Subsequently, since copulas are characterized by data that have a uniform distribution, we transform each simulated data by applying the inverse Cumulative Distribution of an Exponential distribution of parameter λ_i. This procedure will be explained in Section 3.1, and it results in a three dimensional dataset that is exponentially distributed with the following parameters:

$$\boldsymbol{\mu} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 1/10 \\ 1/12 \\ 1/15 \end{bmatrix}.$$

As for the bi-dimensional case presented in 2.6.5, we translate the data by two units, to guarantee that the dataset has negative values, hence that it has financial sense. The resulting dataset is shown in Figure 2.11.

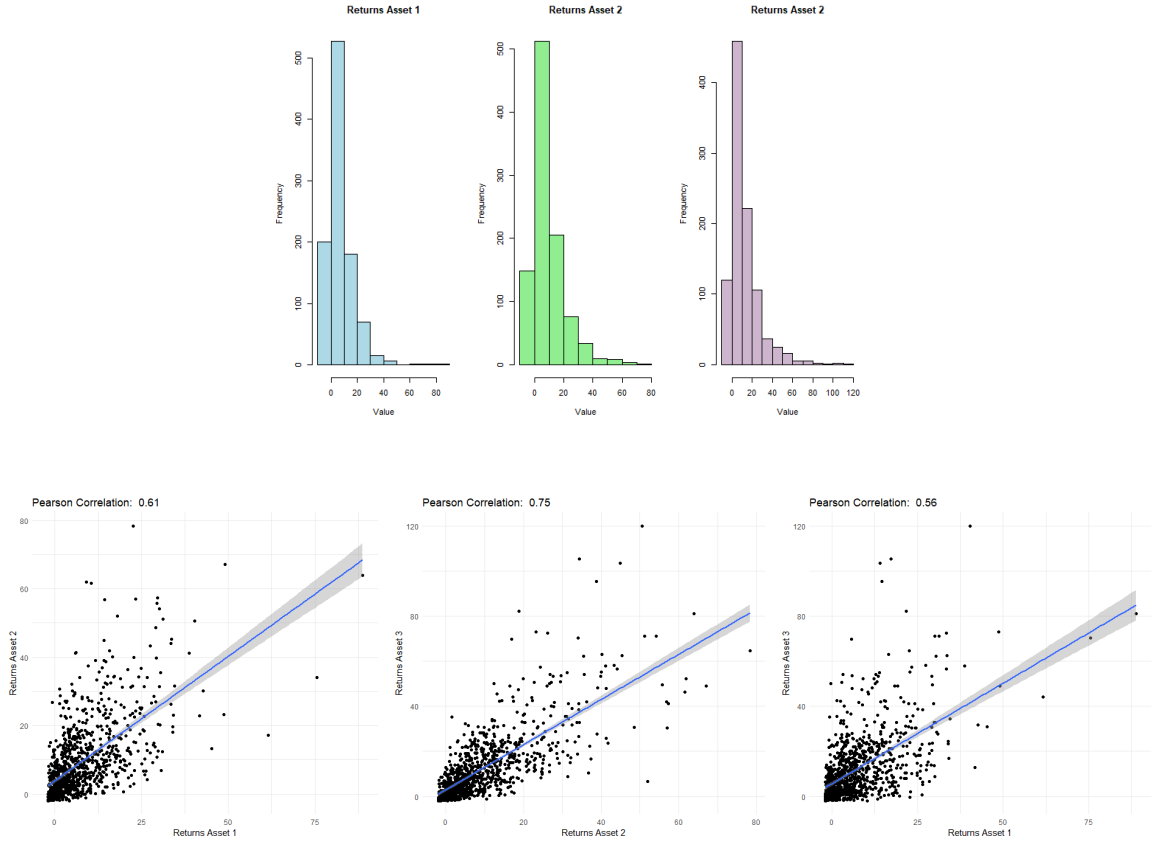


Figure 2.13: *Distribution of the simulated data, drawn from three Exponential distributions with parameters presented above, and obtained from a three dimensional Gaussian copula. Three scatterplots of the data, highlighting their linear correlation is also provided (see section 3.2).*

We proceeded by deriving the three optimal portfolios using the three risk measures, and comparing the results with the EW portfolio, the most risky portfolio and the least risky one. The results are shown in Table 7.1, and in Figure B.5 of the Appendix.

Method	Weights	Mean	Var	VaR	CVaR	Sharpe	STARR	Time
EW	$(33.3\%, 33.3\%, 33.3\%)$	10.364	119.745	0.853	1.357	4.47%	3.955	-
Mean-Var	$(44.8\%, 17.5\%, 37.6\%)$	10.26	116.003	0.841	1.35	**4.53**%	3.897	0.05s
Mean-VaR_{95}	$(18.4\%, 10.1\%, 71.5\%)$	11.826	166.261	0.909	1.396	4.10%	4.89	0.53s
Mean-CVaR_{95}	$(7.9\%, 0.8\%, 91.3\%)$	12.642	204.436	1.071	1.504	3.73%	**5.082**	55.5s
Least Risky	$(100\%, 0\%, 0\%)$	7.993	99.912	1.496	1.749	2.99%	1.711	-
Most Risky	$(0\%, 0\%, 100\%)$	13.068	227.446	1.23	1.618	3.54%	4.987	-

Table 2.5: *Table comparing the portfolios obtained by optimizing the weight of a Gaussian copula, transformed by applying Exponential distributed marginals, using different methods: the Equal Weight, The Mean-Var, the Mean-$VaR_{0.95}$ and the Mean-$CVaR_{0.95}$. The table reports the weights and the results of the risk measures and ratios computed* ***numerically****.*

The considerations made from Table 7.1 are similar to the previous simulations. Mean-Variance tends to prefer a less risky portfolio, while the other two methods present similar results, tending toward a more risky approach. The computational time is rather small for all methods considered, with Markovitz' method detaining the best time. The best Sharpe ratio is held by the Mean-Var portfolio, while the best STARR by the Mean-CVaR one. However, while the values for the Sharpe are quite close to each other, the CVaR has a significantly better value than the other portfolios considered, with the exception of the Value at Risk one.

Conversely, by analyzing Figure B.5 of the Appendix, we can observe how, while for the Mean-Variance case the EW portfolio lies on the efficient frontier, both other risk measures outperform the benchmark. Furthermore, a detailed comparison of results is presented in Figure 2.14, where optimal portfolios are plotted against the Mean-CVaR optimal frontier. In this case, all portfolios lie on the frontier (with the exception of the EW), yet Markowitz's portfolio is positioned on the inefficient side. This indicates that, for equivalent risk, we can find a portfolio that has higher expected value. This is a first example of how the simplicity of Markovitz' approach can fail to identify an optimal portfolio for more complex data.

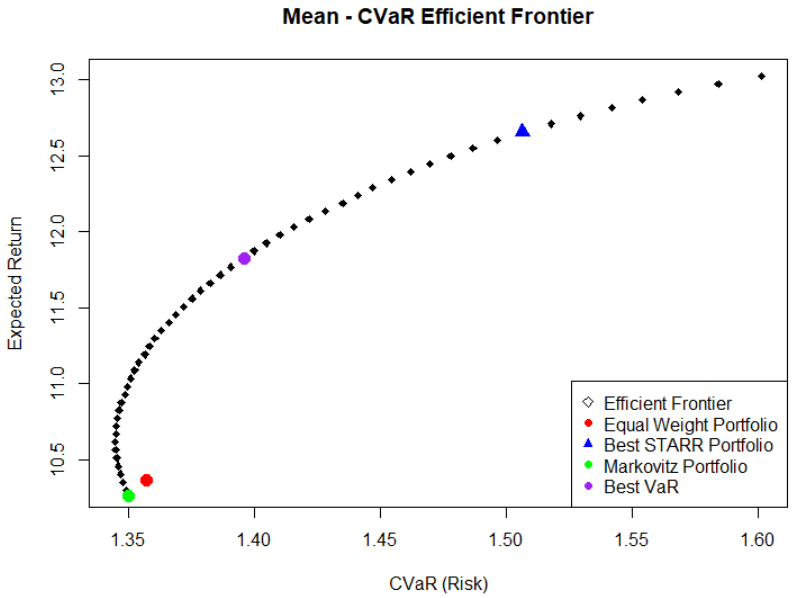


Figure 2.14: *Mean-CVaR efficient frontier, computed optimising a portfolio of three set of returns, derived from a Gaussian copula, and transformed using an exponential. In blue we find the best STARR ratio portfolio, in red the EW portfolio, while in purple the best Mean-VaR, and in green the best portfolio according to Markovitz.*

2.7 Conclusions

In light of the findings discussed in the previous sections, we will adopt $CVaR_\beta$ as the chosen risk measure. In fact, it has demonstrated favorable properties, including subadditivity, ensuring the construction of a diversified portfolio, and a consideration of asymmetries and tail behaviour, which are characteristics that become necessary in more complex distributions, such as those observed in financial scenarios. In addition, from Section 2.6.7, we have seen how adopting more complex risk measures becomes favorable in more complex settings. Lastly, while its formulation can seem more complex, the equivalent problem discussed in (2.29) allows us to deal with a much simpler problem, which can be solved with a linear programming optimisation.

As a consequence, to select the optimal portfolio from the frontier we will use the STARR Ratio (see 2.32). On the other hand, we will also run some analysis and comparison of the selected portfolio with other benchmarks. Those comparisons will be ran using Sharpe ratio ((2.31)), which is more simple and straightforward to compute.

More in general, within portfolio optimisation, many methods have been developed and explored over time, with each offering a unique set of advantages and drawbacks. While Markowitz's approach retains its significance for its simplicity, the field has evolved considerably since its formulation. [9] provides a wide overview of the state-of-the-art methodologies available, describing and comparing earlier versions with the latest advancements. In this thesis in particular, we will discuss the introduction of copula models in Chapter 3, and the one of time series models in Chapter 5. Moreover, we will combine the two models in Chapter 6, and explain the final procedure adopted to construct an efficient frontier in Chapter 7.

3

Copulas

This Chapter aims at providing all the knowledge necessary to understand copula models, which are useful tools for modelling dependencies, and particularly relevant in risk and tail modelling. Copulas have gained relevancy in financial applications, supporting modelling assets and risk management [26].

Copulas were first introduced in 1959 by Sklar in [4]. The main idea behind copulas is to study multivariate distributions, by projecting each variable to the unit hypercube $[0,1]^d$. This is carried out by applying the correct marginal distribution function to each variable of the multivariate dataset. This ensures to eliminate the information contained in the marginals, and to focus on exploring the relationships between variables, and their dependencies.

More in detail, given d random variables $X_1, \ldots, X_d$, we can define their marginal distributions $F_1, \ldots, F_d$ and their joint distribution F. However, the marginal distributions are not sufficient to characterize the joint distribution, because they do not describe the relationship between the different components. The missing link is the copula, which is defined by [4] as follows:

Definition 3 (Copula)**.** *A d-dimensional Copula $C(u_1, ..., u_d)$ is a multivariate distribution function on the hypercube $[0,1]^d$, where the marginal distributions are uniform.*

Starting from Sklar's theorem, this chapter will discuss the theoretical background behind *copulas*, and the corresponding statistical methodology following the textbooks [27], [28]. A few usual copula families will be introduced, as well as details on how to estimate their parameters. The chapter will conclude with a simulation study, aimed at illustrating the model selection procedure described in Section 3.4, and to give some insight on the application of copula models.

3.1 Sklar's Theorem

In this section, we will discuss an important theorem that illustrates how to construct a copula, and vice-versa how to derive the joint distribution from a copula. This theorem is Sklar's theorem, which details the relationship between copulas and the CDFs of a multivariate random vector.

Theorem 4 (Sklar's Theorem)**.** *Given a d-dimensional random vector* $\boldsymbol{X} = (X_1, ..., X_d)$ *that has a joint distribution* F*, and marginal distributions* $F_1, ..., F_d$*, then there exists a copula* C*, such that* $\forall \boldsymbol{x} = (x_1, ..., x_d) \in \mathbb{R}^d$*:*

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)). \tag{3.1}$$

If the density of $\boldsymbol{X}$ *exists, hence if* F *is differentiable with respect to all of its arguments, we can also express this relationship using the copula density* c*,*

$$f(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d))f_1(x_1)...f_d(x_d). \tag{3.2}$$

If the marginal distributions $F_1, ..., F_d$ *are absolutely continuous, then the copula* C *is unique.*

Moreover, given a copula C *associated to a multivariate distribution* F*, with marginals* $F_1, ..., F_d$*, we have that* $\forall u \in [0, 1]^d$ *the copula can be expressed as:*

$$C(u_1, ..., u_d) = F(F_1^{[-1]}(u_1), ..., F_d^{[-1]}(u_d)), \tag{3.3}$$

where $F_i^{[-1]}(s) = \inf\{t | F_i(t) \geq s\}$*,* $s \in [0, 1]$*, represents the pseudo inverse of the marginal CDF of the random variable* X_i*.*

From Sklar's Theorem we learn how copulas are used to represent the relationship between the joint distribution of a random vector, and the marginal distributions of its components. Copulas are therefore a useful tool used to model the dependence among the components of a random vector, and ultimately to make predictions.

3.2 Dependence Measures

Dependence is a rather complex mathematical concept, which can be mathematically formulated via the concept of copulas, by eliminating the effect of marginal distributions. In general, dependence, especially in the two dimensional case, is tightly related to correlation, a metric that incorporates the information about the relationship between two variables within the [-1, 1] interval.

The most commonly used dependence measure is Pearson correlation, which quantifies linear dependence. However, many alternatives are available, each with distinct properties: it is therefore important to chose the correct measure, depending on the characteristic relevant for a specific task. Copulas are typically associated with two rank-based measures: Kendall's Tau and Spearman's Correlation. In this section we will present these two dependence measures, which offer the advantage of being well-defined and invariant to changes in scale. We will define them, illustrate how they are estimated, and their relationship with copulas. Additionally, we will discuss the concept of tail dependence coefficients, which are particularly useful when studying extreme events.

3.2.1 Kendall's Tau

Kendall's Tau is a rank-based dependence measure, that assesses the degree of dependence between two random variables based on the ranks of their observations, rather than the actual values. Along with Spearman's ρ_s, it can be expressed only in terms of the underlying copula. It is defined as follows:

Definition 4 (Kendall's tau)**.** *If we have two random variables* X *and* Y*, where each pair* (X_1, Y_1) *and* (X_2, Y_2) *are independent and identically distributed as* (X, Y)*, we can define Kendall's tau as the*

difference of the following probabilities:

$$\tau(X,Y) = \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0). \tag{3.4}$$

Kendall's Tau is defined as the probability that the two random variables are concordant, minus the probability of discordance [27]. More precisely, "Two random variables X_1 and X_2 are concordant when large values of X_1 go with large values of X_2." [29].

3.2.2 Spearman's Correlation

Spearman's Correlation is also a rank-based measure, defined as:

Definition 5 (Spearman's Correlation)**.** *If we have two random variables X_1 and X_2, with the corresponding marginal distributions F_1 and F_2, Spearman's Correlation is defined as:*

$$\rho_s := \rho_s(X_1, X_2) = \mathrm{Cor}\left(F_1(X_1), F_2(X_2)\right). \tag{3.5}$$

Spearman's Correlation is therefore computed as the Pearson correlation $\mathrm{Cor}(\cdot,\cdot)$, between two random variables, which have been transformed to unit scale by applying their marginal distribution.

3.2.3 Tail Dependence

Tail dependence expresses the probability of joint occurrence of extreme events, which can either be extremely small or large. It is defined as follows:

Definition 6 (Upper and lower tail dependence coefficient)**.** *The upper tail dependence coefficient of a bivariate distribution with copula C of the random variables X_1, X_2, with cdf F_1, F_2, is defined as:*

$$\lambda^{upper} = \lim_{t \to 1^-} \mathbb{P}(X_2 > F_2^{-1}(t) | X_1 > F_1^{-1}(t)) = \lim_{t \to 1^-} \frac{1 - 2t + C(t,t)}{1 - t},$$

while the lower tail dependence coefficient is:

$$\lambda^{lower} = \lim_{t \to 0^+} \mathbb{P}(X_2 \leq F_2^{-1}(t) | X_1 \leq F_1^{-1}(t)) = \lim_{t \to 0^+} \frac{C(t,t)}{t}.$$

The tail dependence coefficients provide valuable insight into the occurrence of joint extreme events. A high tail dependence coefficient indicates that the probability of a similarly extreme event in variable X_2 is high when an extreme event occurs in variable X_1.

However, it is important to note that accurate estimation of the tail dependence coefficient requires a large number of observations to ensure stability. It is also worth mentioning that Gaussian copulas (which will be discussed in Definition 3.3.2) exhibit a tail dependence coefficient of zero [27], making them unsuitable for financial applications: it has, in fact, been proven that in periods of crisis, hence during extreme events, markets exhibit a high correlation [30].

3.2.4 Dependence Measures Estimation

After having defined Kendall's tau and Spearman's ρ_s, we can proceed to discuss how these dependence measures are estimated empirically from an observed dataset of n pairs, (x_{i1}, x_{i2}), for $i = 1, \ldots, n$.

From a given dataset, we can estimate Kendall's tau:

$$\hat{\tau}_n := \frac{N_c - N_d}{\sqrt{N_c + N_d + N_1} \times \sqrt{N_c + N_d + N_2}}, \tag{3.6}$$

where:

- N_c is the number of *concordant* pairs. Two pairs (x_1, y_1) and (x_2, y_2) are concordant if $x_1 < x_2$ and $y_1 < y_2$, or vice versa $x_1 > x_2$ and $y_1 > y_2$,
- N_d is the number of *discordant* pairs: the ones where the x_i have the opposite relationship than the y_i,
- N_1 is the number of pairs where $x_i = x_j$,
- N_2 is the number of pairs where $y_i = y_j$.

Subsequently, we can estimate Spearman's ρ_s.If we define r_{hj}, $j \in \{1, 2\}$, as the h-th ranked data, this is simply obtained by assigning a numerical rank to each x_{ij}, for $i = 1, \ldots, n$, based on their magnitude. The estimator is the following:

$$\hat{\rho}_s = \frac{\sum_{i=1}^{n}(r_{i1} - \bar{r_1})(r_{i2} - \bar{r_2})}{\sqrt{\sum_{i=1}^{n}(r_{i1} - \bar{r_1})^2}\sqrt{\sum_{i=1}^{n}(r_{i2} - \bar{r_2})^2}},$$

where $\bar{r_j} = \frac{1}{n}\sum_{i=1}^{n} r_{ij}$.

3.2.5 Relationship with Copulas

Moreover, from Definition 6, we have seen how tail dependence can be expressed with copulas. Both other Kendall's Tau and Spearman's Correlationn can also be expressed in terms of copulas. The following theorems illustrate the copula formulations for these measures.

Theorem 5. *If we have two random variables X and Y,*

1. *Kendall's τ is:*
$$\tau = 4 \int_{[0,1]^2} C(u_1, u_2)\, dC(u_1, u_2) - 1. \tag{3.7}$$
2. *Spearman's ρ is:*
$$\rho_s = 12 \int_{[0,1]^2} u_1 u_2 dC(u_1, u_2) - 3. \tag{3.8}$$

Both theorems are taken from [27], and the corresponding proofs, which are quite straightforward, are discussed in Sections A.2.1 and A.2.2 of the Appendix.

3.3 Copula Families

In the previous section we have addressed how to quantify dependence, with the use of different dependence measures. However, these measures gather all the information in a single number, which might not be sufficient to explore the complex relationships that are established among the variables,

especially in higher dimensions. Copulas are a valid alternative able to fully express these relationships.

In the following subsections, the most common Copula families will be introduced. We will present Gaussian and t-Student copulas, which are popular multivariate elliptical copulas. Additionally, we discuss the Fréchet-Hoeffding Bounds, exploring their significance in characterizing copula functions. Furthermore, a general overview on Archimedean copulas, along with the relative properties, is addressed.

3.3.1 Fréchet-Hoeffding Bounds

It is first useful to introduce bounds for every copula, no matter what family they belong to. Given a d-dimensional copula C, we have that:

$$W_d(\mathbf{u}) \leq C(\mathbf{u}) \leq M_d(\mathbf{u}),\ \forall \mathbf{u} \in [0,1]^d, \tag{3.9}$$

where $W_d(\mathbf{u})$ and $M_d(\mathbf{u})$ are defined as follows [31]:

Definition 7 (Fréchet-Hoeffding Bounds)**.** *Given $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$, the Fréchet-Hoeffding Bounds are defined as:*

$$\begin{cases} W_d(\boldsymbol{u}) = \max\left\{\sum_{i=1}^{d} u_i - d + 1, 0\right\} \\ M_d(\boldsymbol{u}) = \min\{u_1, ..., u_d\} \end{cases}$$

The discussion can be extended to the more broad concept of concordance ordering, which gives a way of ordering copulas [26]. The definition is the following:

Definition 8 (Concordance Ordering)**.** *We say that a copula C_1 is smaller than the copula C_2 (or C_2 is larger than C_1), and write $C_1 \prec C_2$ if $\forall \boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$:*

$$C_1(u_1, \ldots, u_d) \leq C_2(u_1, \ldots, u_d).$$

From this definition, we can rewrite (3.9) as:

$$W_d \prec C \prec M_d. \tag{3.10}$$

The importance of the Fréchet-Hoeffding bounds lies in their capability of delineating the range of copula. Hence, they are useful to represent extreme dependence scenarios, and can be applied for risk management and extreme value theory. In particular, they offer insights on the behaviour of copula's tails under extreme conditions.

It is important to mention that, for dimensions greater than 2, the Fréchet-Hoeffding lower bound $W_d(\mathbf{u})$ is not a copula anymore [27], as its image does not lie in the interval $[0,1]$ anymore. This highlights a possible limitation of these bounds, indicating a need to find alternative approaches when extending copula models to higher dimensions.

3.3.2 Gaussian Copulas

One of the most known copula models is a model that belongs to the family of Elliptical copulas: the Gaussian copula.

Given the cumulative distribution of a univariate standard normal Φ, and its inverse Φ^{-1}, we can define a multivariate Gaussian copula with mean zero and correlation matrix $R \in [-1,1]^{d\times d}$, for each $\mathbf{u} = (u_1, .., u_d) \in [0,1]^d$:

$$C(\mathbf{u}; R) = \Phi_R(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)), \tag{3.11}$$

where Φ_R is the multivariate CDF of a standard normal with mean zero, unit variances, and R as the symmetric positive definite correlation matrix. We can also derive the copula density as:

$$c(u; R) = |R|^{-1/2} exp\{\frac{x^T(I_d - R^{-1})x}{2}\}, \tag{3.12}$$

where $\mathbf{x} = (x_1, ..., x_d)^T \in \mathbb{R}^d$, $x_i = \Phi^{-1}(u_i)$, $u_i \in [0,1]$, for $i = 1, ..., d$, and for I_d the identity matrix of dimension d [27].

More in detail, we can construct the bivariate Gaussian copula applying Sklar's theorem to the bivariate normal distribution $\Phi_2(\cdot)$ with mean zero, unit variance and correlation ρ: $C(u_1, u_2; \rho) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)$. The corresponding copula density with standard Gaussian margins and correlation ρ, will be [27]:

$$c(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}} \exp\{-\frac{\rho^2(x_1^2 + x_2^2) - 2\rho x_1 x_2}{2(1-\rho^2)}\}.$$

Two examples of bivariate Gaussian copulas can be observed in Figure 3.1.

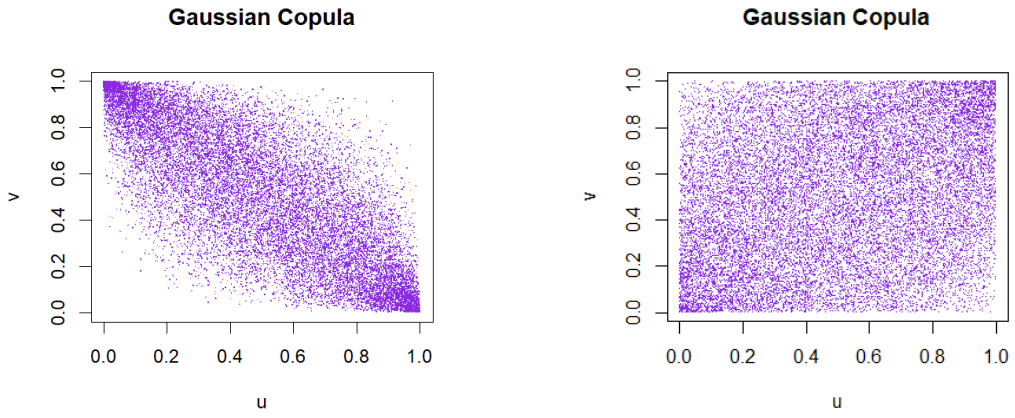


Figure 3.1: *Two examples of bi-dimensional Gaussian copulas. We provide two scatterplots of 20000 sampled data from a Gaussian copula with, respectively, a correlation ρ of -0.8 and 0.3. The Kendall's τ are -0.59 and 0.19.*

A Gaussian copula is characterized by radial symmetry, hence the copula $C(U_1, \ldots, U_d)$ has the

same distribution of its survival copula $\hat{C}(1-U_1,\ldots,1-U_d)$, where a survival copula is defined by [27] as:

Definition 9 (Survival Copula)**.** *Given a d-dimensional random vector* $\boldsymbol{X} = (X_1, ..., X_d)$*, with marginal distributions* $F_1, ..., F_d$*, the Survival Copula* $\hat{C}$*, is defined as:*

$$\hat{C}(1 - F_1(x_1), ..., 1 - F_d(x_d)) = \mathbf{P}(X_1 > x_1, \ldots, X_d > x_d). \tag{3.13}$$

In particular, when $d = 2$ *the following relationship holds:*

$$\hat{C}(u_1, u_2) = u_1 + u_2 + C(1 - u_1, 1 - u_2) - 1.$$

Moreover, a Gaussian copula is invariant under the permutation of its arguments u_i, a property known as "*exchangeability*": for example, in the bivariate case, we have $C(u_i, u_j) = C(u_j, u_i)\ \forall i, j$, which means that in a scatterplot its points are symmetrically located around the diagonal. In addition, as already discussed in Section 3.2.3, Gaussian copulas exhibit zero tail dependence [32].

Lastly, bidimensional Gaussian copulas are closely related to the Fréchet-Hoeffding bounds defined in 7. In fact, the bounds in two dimensions can be obtained by setting the parameter ρ of a Gaussian copula respectively equal to -1 and $+1$ [26]. The concordance ordering of (3.10) can be therefore represented as:

$$W_d = C_{\rho=-1} \prec C \prec C_{\rho=-1} \prec M_d.$$

3.3.3 t-Student Copulas

Another multivariate copula family of Elliptical copulas is the t-Student copula. As the name suggests, it is derived from a t-Student distribution, and generally, it is used for datasets with high tail dependence, which allows this family to be more suitable for financial data. A t-student copula is characterized by its correlation matrix R, and its degrees of freedom ν. An example of a bivariate copula with a t-Student distribution can be found in Figure 3.2.

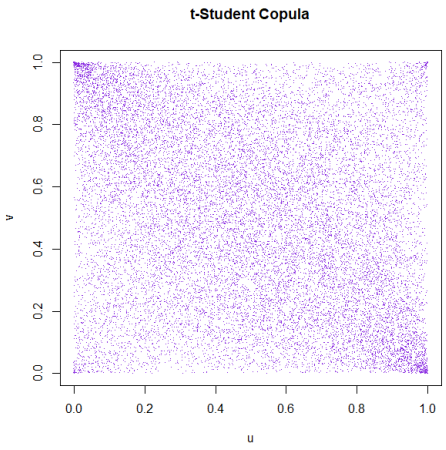


Figure 3.2: *20000 generated points of a t-Student copula, with correlation set to* $\rho = -0.3$ *and* $\nu = 3$ *degrees of freedom. Kendall's* $\tau = -0.19$*.*

The multivariate t-student copula with correlation matrix $R \in [-1,1]^{d\times d}$, and $\nu > 0$ degrees of freedom, is computed with the following formula:

$$C(\mathbf{u}; R, \nu) = T_{R,\nu}(T_\nu^{-1}(u_1), \ldots, T_\nu^{-1}(u_d)), \tag{3.14}$$

where $T_{R,\nu}$ is the multivariate Student's t distribution, and T_ν^{-1} is the inverse of the univariate standard t-Student with ν degrees of freedom.

The bivariate t-student copula, for example, can be derived by applying Sklaar's theorem to the standard student's t distribution, with mean $\mu = 0$, and correlations ρ, which has density given by:

$$t(x_1, x_2; \nu, \rho) = \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sqrt{1-\rho^2}} \left\{1 + \frac{x_1^2 - 2x_1x_2\rho + x_2^2}{\nu(1-\rho^2)}\right\}^{-\frac{\nu+2}{2}}.$$

The copula can be therefore computed by integrating over the density, and its distribution is given by:

$$C(u_1, u_2; \nu, \rho) = \int_{-\infty}^{b_1}\int_{-\infty}^{b_2} t(x_1, x_2; \nu, \rho)dx_1dx_2,$$

where $b_1 := T_\nu^{-1}(u_1)$ and $b_2 := T_\nu^{-1}(u_2)$.

T-Student copulas for dimension $d = 2$ have symmetric tail dependence, hence we have that $\lambda^{upper} = \lambda^{lower}$. Lastly, like Gaussian copulas, the bivariate t-Student copulas are radially symmetric and exchangeable [32].

3.3.4 Archimedean Copulas

Lastly, we introduce Archimedean copulas, which include a great variety of copula families, all identified by the general formula reported here [26].

Definition 10 (Archimedean Copulas)**.** *Let Ω be the set of all continuous functions $\varphi \in C^2$, with $\varphi(1) = 0$, $\varphi'(u) < 0$, and $\varphi''(u) > 0$, $\forall u \in [0,1]$. Let $\varphi \in \Omega$, then:*

$$C(\boldsymbol{u}) = \begin{cases} \varphi^{[-1]}(\varphi(u_1) + \ldots + \varphi(u_d)), & \text{if } \sum_{i=1}^d \varphi(u_i) \leq \varphi(0) \\ 0, & \text{otherwise.} \end{cases} \tag{3.15}$$

is a copula, defined as Archimedean Copula. $\varphi(x)$ is defined as the Archimedean generator function. The density of a bivariate Archimedean copula can also defined as:

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} = \frac{\varphi''(C(u_1, u_2))\,\varphi'(u_1)\,\varphi'(u_2)}{[\varphi'(C(u_1, u_2))]^3}.$$

A sufficient condition for Archimedean generators to generate a proper copula, is that they satisfy the "d-monotoniticy", hence they are functions that are all functions φ, with $\varphi(0) = 1$ that are infinitely often differentiable in their domain $(0, \infty)$ [32].

The distinction between Archimedean copulas lies in the choice of the generating function $\varphi(x)$. An example could be the Gumbel copula, which is characterized by $\varphi(x) = (-\log(x))^\theta$, $\theta \geq 1$, and Frank's

copula which has $\varphi(x) = -log(\frac{e^{-\theta x}-1}{e^{-\theta}-1})$. The parameter for Frank's copula take values in the interval $\theta \in [-\infty, \infty]\backslash\{0\}$.

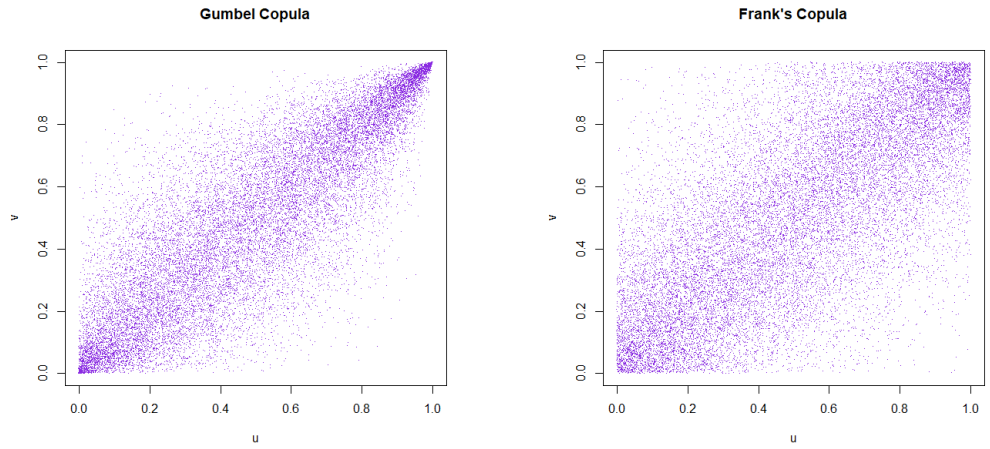


Figure 3.3: *In this picture we have two examples of Archimedean copulas: 20000 points sampled from a Gumbel copula with parameter $\theta = 3$, and 20000 from a Frank copula with θ set to 6. The corresponding Kendall's τ are* 0.67 *and* 0.51.

The importance of Archimedean copulas can be identified in several properties, in fact they are symmetric, associative and have a simplified way of computing Kendall's tau [26], which is given by:

$$\tau = 1 + 4 \int_0^3 1 \frac{\varphi(u)}{\varphi'(u)} du.$$

In addition, like Gaussian copulas, Frank copulas are characterized by tail independence. On the other hand, Gumbel copulas present only upper tail dependence λ^{upper}. Moreover, from their definition we obtain that Archimedean copulas are exchangeable even for dimensions greater than 2, which means that they are invariant under the permutation of their arguments [32].

3.3.5 Other Families

The set of available copula families extends beyond those mentioned in the previous sections, particularly when considering the bivariate case. Worth mentioning examples include Claytons's copulas, Joe's copulas, and BB copulas. For readers interested in further deepening the knowledge of copula families, [27] presents a wide coverage of the topic.

3.4 Copula Models

A parametric statistical model is a collection of probability distributions on a given sample space characterized by a set of parameters Θ (see Def. 37 of the Appendix). This concept extends to copulas, where a d-dimensional copula can be associated with a set of distributions on the sample space $\Omega = [0, 1]^d$. Starting from the definition of parametric model, we can derive the definition of a Copula model:

Definition 11 (Copula Model)**.** *A copula model $\mathcal{C}$ is a family of probabilities distributions, characterized by a finite number of parameters:*

$$\mathcal{C} = \{\mathbb{P}_\theta : \theta \in \Theta\}.$$

The set of parameters Θ that identify each copula depends on the underlying family. An example of copula's parameters can be Kendall's tau τ, or Spearman's Correlation ρ_s.

From definition 11 we can conclude that, when performing model selection in the context of copulas, we will have to face two main choices. We will need to select the family type (several families are available, some examples were introduced in Section 3.3), and estimate the values of the parameters associated with the chosen copula family.

Then one question arises: what is the process for estimating a copula model?
From a d-dimensional set of observations $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_d)$, where each $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,n})$, $u_{i,j} \in [0, 1]$, we can fit a distribution, with the objective of finding a one able to correctly mimic the behaviour of the considered dataset. The initial step involves the selection of a set of family distributions. Then, for each family, we estimate the values of the associated parameters, using different estimation methods, detailed in Section 3.4.1. The next step consists in comparing different criteria to select the best-fitted copula from the set of candidates. Lastly, the quality of the chosen model will be assessed with some Goodness of Fit tests, discussed in Section 3.4.3.

3.4.1 Parameter Estimation

To estimate the parameter θ of a copula with a specific family, we can use the Maximum Likelihood estimator, which is the the value $\hat{\theta}^{ML}$ that maximises the likelihood function $L(\theta)$. More details on both the likelihood function, and the estimator can be found in the Subsection in Definition F.1.1 of the Appendix. However, an approach that uses the Canonical Maximum Likelihood, or CML, instead of the regular ML, is generally preferred for variables of the dataset are not uniformly distributed, as it does not require optimization to compute the margins.

The method requires two steps. Firstly, we transform the variables to a uniform scale, by applying their empirical distribution. Hence, for a given observation sample $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_d)$, where each $\mathbf{x}_i$ is a vector of N observations, we find the empirical cumulative distribution $\hat{F}_i$ of each component as described in Section F.4.1. With the empirical CDF we can then transform the observations to their uniform variants $\hat{u}_{i,j} = \hat{F}_i(x_{i,j})$, where $x_{i,j}$ is the j-th component of the vector $\mathbf{x}_i$.

Once the first step is completed, we then proceed by estimating the parameters using the following estimator [33]:

$$\hat{\theta}^{CMLE} := \arg\max_{\Theta} \sum_{j=1}^{N} \log c_{\Theta}(\hat{u}_{1,j}, \dots, \hat{u}_{d,j}; \Theta), \tag{3.16}$$

where c_Θ is the density of the chosen copula family with unknown set of parameters Θ. $\hat{\mathbf{u}} = (\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_d) = (\hat{F}_1(\mathbf{x}_1), ..., \hat{F}_d(\mathbf{x}_d))$ is the d-dimensional vector of pseudo-observations, transformed to uniform scale by applying the empirical CDFs.

3.4.2 Model Selection

Once we have the different distributions, along with their estimated parameters, a comparison can be made using metrics such as Likelihood, AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion) to determine the most suitable model (refer to Def. 41 of the Appendix): we will choose the model with higher Likelihood, or lower value of AIC and BIC.

3.4.3 Hypothesis Testing

Once a model has been selected, it is left to evaluate whether the model is a good fit to the original dataset. Hence, to analyse if the chosen theoretical distribution appropriately describes the distribution of the sample data, typically through a Goodness of Fit (GoF) test. The choice of GoF test depends on the specific objectives of the model fitting process: in this Section we will discuss the White Test for copulas. Additionally, we will discuss the Vuong Test, a hypothesis test used to compare different distributions.

3.4.3.1 White Test for Copulas

In our analysis, we will mainly be working with the test implemented in the function "*BiCopGofTest*", from the VineCopula package in R [34], which assess the Goodness of Fit of bivariate copulas. The same test can be applied to multivariate copula models using different functions, such as the function 'gofWhite', implemented in Package [35].

This test is designed to assess the goodness of fit of copulas using White's information matrix equality [36], as discussed in [37]. It is based on the assumption that, under the correct model specifications, the Fisher Information matrix $\mathbf{C}(\theta)$ can be computed as the negative of the expected Hessian matrix $\mathbf{H}(\theta)$, which is the matrix of the second order partial derivatives of the likelihood function. Consequently, the null hypothesis is formulated as:

$$H_0 : \mathbf{H}(\theta) + \mathbf{C}(\theta) = 0,$$

while the alternative is that the sum is not equal to zero.

The test statistic is computed using the MLE estimator $\hat{\theta}$ (refer to F.1.1), and is equal to $d\bar{D}_{\hat{\theta}}^T V_{\theta_0}^{-1} \bar{D}_{\hat{\theta}}$, where d is the number of variables considered, $V_{\theta_0}^{-1}$ the inverse of the asymptotic variance, and $\bar{D}_{\hat{\theta}} = \frac{1}{d}\sum_{t=1}^{d} \hat{d}_t(\theta)$. Here, $\hat{d}_t(\theta) = vech(\mathbb{C}_t(\theta) \hat{+} \mathbb{H}_t(\theta))$, with $vech$ denoting the vertical vectorization of the lower triangular of a matrix. The test statistic follows asymptotically a χ^2 distribution with $\frac{p(p+1)}{2}$ degrees of freedom, where p represents the dimension of θ [37].

3.4.3.2 Vuong Test

Another test that could be used would be the Vuong test, a hypothesis test designed to verify if a chosen distribution is significantly better than the alternative ones by comparing their Kullback–Leibler Information criterion.

In order to do so, we will need to introduce the Kullback–Leibler Information criterion (KLIC), which measures the "distance" between an unknown density f^*, and the approximated parametric density $f(\cdot|\theta)$ of parameter θ. It is used to compare statistical models.

Definition 12 (Kullback–Leibler Information criterion (KLIC))**.** *Given the true density of a random vector $\boldsymbol{X}$, f^*, and the parametric density $f(\cdot|\theta)$, we define the KLIC as:*

$$KLIC(f^*, f, \theta) := \int \ln \frac{f^*(\boldsymbol{x})}{f(\boldsymbol{x}|\theta)} f^*(\boldsymbol{x}) d\boldsymbol{x} = \mathbb{E}^*[\ln f^*(\boldsymbol{X})] - \mathbb{E}^*[\ln f(\boldsymbol{X}|\theta)], \tag{3.17}$$

where $\mathbb{E}^*[\boldsymbol{X}] = \int f^*(\boldsymbol{x}) d\boldsymbol{x}$ is the expected value of a random variable or random vector $\boldsymbol{X}$ with density f^*.

Once we have defined the KLIC, we can finally introduce the Vuong hypothesis test, which is used to check if, given two models $\mathcal{M}_1$ and $\mathcal{M}_2$ with parametric densities $\{f_j(\cdot|\theta)\forall\theta \in \Theta, j = 1, 2\}$, one model has to be preferred over the other. The Null hypothesis of the Voung is that the two models have the same KLIC, hence, the H_0 hypothesis is:

$$H_0 : \mathbb{E}_0[\ln f_1(\mathbf{X}|\theta_1)] = \mathbb{E}_0[\ln f_2(\mathbf{X}|\theta_2)].$$

The alternative hypothesis is that they have a different KLIC:

$$H_1 : \mathbb{E}_0[\ln f_1(\mathbf{X}|\theta_1)] \neq \mathbb{E}_0[\ln f_2(\mathbf{X}|\theta_2)].$$

For a given significance level α, we reject H_0 if $|\nu(x_1, ..., x_d)| > \Phi^{-1}(1 - \frac{\alpha}{2})$, where ν is defined as the Likelihood ratio statistic $LR_d(\hat{\theta}^1, \hat{\theta}^2)(x_1, ..., x_d)$ for the two models, divided by the square root of the sample size times the sample variance of LR_d (the detailed formula can be found in [27]), and Φ^{-1} is the inverse of the CDF of a standard normal. For non-nested models, the test statistic asymptotically follows a Gaussian distribution [27].

3.5 Simulation Study to illustrate Model Selection

In order to provide a more comprehensive understanding of the steps discussed in the previous Section regarding the construction of a Copula model, a small simulation study has been conducted. This should help the reader to understand the main steps, by presenting a concise example that establishes a general framework that will serve as the basis for analyzing real-world data in the subsequent chapters.

The simulation study is conducted on a two-dimensional copula for simplicity, and focuses on two main objective: evaluate the accuracy of parameter estimation for small datasets, and visualize the convergence of parameter estimators as the number of observations increases, by computing the mean squared error (MSE). The simulation is then extended to a 4-dimensional framework, and could be easily extended to higher dimensions.

3.5.1 Bidimensional Case

We started by generating five different bivariate copulas, one for each copula family introduced in Section 3.3 and an additional one using a Clayton copula. Since we want to test the efficiency of the estimators, we work with a relatively small dataset to analyze the estimation performance. we therefore simulate $N = 100$ point from each copula using the following parameters. We simulate from a Gaussian with correlation $\rho = -0.8$ ($\tau = -0.590$), a t-Student with $\rho = -0.3$, and $\nu = 3$ degrees of freedom ($\tau = -0.194$), then we proceeded to simulate a Clayton with $\theta = 5$ ($\tau = 0.714$), a Gumbel copula with parameter $\theta = 3$ ($\tau = 0.667$), and lastly a Frank copula with $\theta = 6$, $\tau = 0.514$. The true parameters are summarized in Table 3.2.

To each simulated copula, we assumed that the original copula distribution was unknown, proceeding to fit different copula families to determine the most suitable one, by comparing the the Bayesian Information Criterion (BIC) values. From the chosen family, hence the family that presented the lower BIC value, we then saved the estimated the parameters, which are reported in Table 3.2. The results presented in Table 3.1, indicate that the algorithm successfully identified the original copula family for each tested family despite working with a small dataset.

True Distribution	Gaussian	Student-t	Clayton	Gumbel	Frank
Gaussian	**-84**	-79	5	5	-79
t-Student	-2	**-19**	5	4	-5
Clayton	-111	-112	**-160**	-83	-117
Gumbel	-137	-134	-94	**-143**	-132
Frank	-42	-41	-34	-35	**-59**

Table 3.1: *BIC values obtained by fitting a bivariate copula simulated using the distribution presented in the first column. The results are presented after fitting a Gaussian, a Student-t, a Clayton, a Gumbel, and a Frank copula to the generated copula. We highlight in bold the lowest BIC value for each row.*

Furthermore, we compare the parameters estimated with an MLE estimator from the chosen copula family in Table 3.2, with the ones corresponding to the original copula: as inferred from observing the previous table the parameters are estimated from the same family as the original data-set. We also compare the values of the original Kendall's τ, with the estimated ones.

The estimated parameters appear rather close to the initial ones, in particular for the t-Student case. The estimated Kendall Taus present values that are extremely close to the original ones.

True Dist	True Θ	True τ	Estimated $\hat{\Theta}$	Estimated $\hat{\tau}$
Gaussian	-0.8	-0.590	-0.763	-0.553
Student-t	-0.3, 3	-0.194	-0.306, 2.106	-0.198
Clayton	5	0.714	4.395	0.687
Gumbel	3	0.667	3.316	0.698
Frank	6	0.514	5.904	0.508

Table 3.2: *Estimated parameters and Kendall's τ of each copula, compared to the original ones. The parameters are estimated from the copula's family selected in Table 3.1, in all cases it correspond to the original one.*

Since we are working with only $N = 100$ initial data points, we can consider the results rather satisfactory. To further validate this observation in Table 3.3 we report the p-values of White's Goodness of Fit test, discussed in Section 3.4.3.1. We observe that all p-values provided are larger than the target 0.05, hence we cannot reject the null hypothesis, and we can conclude that all samples considered come from the distribution used to generate the original data.

True Dist	White's p-value
Gaussian	0.83
Student-t	0.93
Clayton	0.83
Gumbel	0.48
Frank	0.94

Table 3.3: *P-values of White's Goodness of Fit test for each bivariate copula.*

However, one question arises: what is the impact of the number of datapoints, and how much can we improve the estimation if we increase the number of observations? We will address this in the following Subsection 3.5.2.

3.5.2 Parameter Estimator's Convergence of 2-Dimensional Copulas

In this section we expand on the observations of the preceding subsection, studying the rate of convergence of the estimators of the parameters of the two-dimensional previously copulas considered. For this

simulation we assume the copula family known, and estimate the parameters by varying the number of observations N, set to $N = \{50, 100, 200, 500, 1000, 2000\}$. We replicated the experiment 100 times for each copula family considered, and computed the mean squared error (MSE) from the true values of each estimated parameter, for each value of N. The results are shown in Figure 3.4.

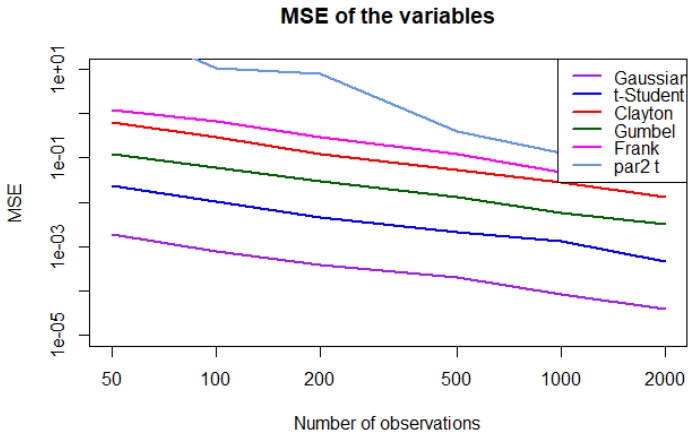


Figure 3.4: *Logarithmic scaled plot studying the convergence of the fitted parameters of the considered copulas. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error values.*

Figure 3.4 represents the rate of convergence of the MSE in a logarithmic scale of the estimated parameters in relationship with the size of the simulated dataset. The graph shows that, for all considered copula families, the convergence rate is rather steep and reaches satisfactory results even with a relatively small amount of data. Already for $N = 200$, the MSE for all estimated parameters is approximately 10^{-1} or lower. The only exception is for the estimator of the second parameter of the t-student copula, ν, which has an initial MSE larger than the other estimated parameters considered.

In addition to the plot, in the Appendix we have included a boxplot showcasing the distribution of the estimated parameters for each copula family: the results are shown in Figure B.6. By observing the plots, it becomes evident that even with a minimal amount of simulated data, such as $N = 50$, each simulated variable exhibits a median that correspond to the original value used to simulate the different copulas. Clearly by increasing N, we observe a reduction in the variability of the data, with observations becoming more concentrated around the true value. Based on these observations, we could conclude that $N = 500$ should sufficient to guarantee a satisfactory estimation for all families considered.

The same observations that were made regarding the parameter estimators can be extended to the plots analyzing the convergence of estimators for Kendall's τ, which are available in the Appendix in Figures B.7 and B.8. With the only specification that the MSE values are even smaller in the case of Kendall's τ, providing an additional layer of confidence in the precision of the estimations.

3.5.3 Extension to 4-dimensions

Building upon the simulation study introduced in the previous Subsections, we expand the simulation to 4-dimensional copulas. We maintain the same number of simulated data ($N = 100$), and the same families previously considered. The parameters used to simulate the copulas are presented in Table 3.5,

alongside the Kendall's τ values. In the same table we compare these parameters with the corresponding estimated values.

As for the two dimensional case, from Table 3.4 we can conclude that for each family considered the BIC value indicates that the model fitted using the original copula family is the one to consider. In Table 3.5 we compare the original parameters to the ones fitted using the copula with the best BIC. Once again, we observe that, even with a small dataset, the estimation proves to be successful. Both the parameters and Kendall's tau are rather close to their original values in most cases for all copulas considered. Similarly to the bidimensional case, we will proceed by studying the evolution of the estimation with the change of the sample size in the following Subsection 3.5.4.

True Distribution	Gaussian	Student-t	Clayton	Gumbel	Frank
Gaussian	**-270**	-268	-124	-137	-125
t	-286	**-305**	-80	-60	-55
Clayton	-376	-374	**-491**	-392	-333
Gumbel	-283	-280	-236	**-328**	-280
Frank	-1070	-1094	-899	-1024	**-1124**

Table 3.4: *BIC values obtained by fitting a 4d copula generated as described using the distribution presented in the first column. The results are presented after fitting a Gaussian, a Student-t, a Clayton, a Gumbel, and a Frank copula to the simulated copula. We highlight in bold the lowest BIC value for each row.*

True Dist	True Θ	True τ	Estimated $\hat{\Theta}$	Estimated $\hat{\tau}$
Gaussian	(0.5, 0.6, 0.72, 0.4, 0.1, 0.8)	(0.33, 0.42, 0.52, 0.27, 0.07, 0.6)	(0.44, 0.68, 0.82, 0.44, 0.20, 0.84)	(0.30, 0.49, 0.67, 0.28, 0.18, 0.66)
Student-t	(0.3, 0.4, 0.8, 0.72, 0.4, 0.1), $\nu = 3$	(0.19, 0.27, 0.60, 0.53, 0.24, 0.12)	(0.16, 0.28, 0.82, 0.73, 0.28, 0.05) , $\nu = 3.51$	(0.10, 0.25, 0.62, 0.51, 0.25, 0.05)
Clayton	4	0.67	3.55	0.66
Gumbel	8	0.60	7.48	0.62
Frank	10	0.90	9.01	0.89

Table 3.5: *Estimated parameters and Kendall's τ of each copula, compared to the original ones. The parameters are the ones of the family that resulted in the lowest BIC value (for all cases it coincides with the original one).*

Lastly, in Table 3.6 we provide the p-values of the White Goodness of Fit test, used to evaluate the models, as discussed in section 3.4.3.1. All copula families have p-values that are greater than the significance level 0.05, hence we cannot reject the null hypothesis. The data generated from the fitted model come from the same distribution of the original datasets.

True Dist	White's p-value
Gaussian	0.67
Student-t	0.81
Clayton	0.99
Gumbel	0.48
Frank	0.24

Table 3.6: *P-values of White's Goodness of Fit test.*

3.5.4 Convergence of 4-Dimensional Copulas

Also for the four dimensional case we proceeded to study the estimators' convergence. For simplicity we only run the simulations for the Clayton, the Gumbel and the Frank copula using the parameters presented in Table 3.5: these copulas are characterized by a single parameters even in higher dimensions. The results are available in the appendix, in Section B.2.2.

Clearly, fitting a copula in higher dimensions results in additional challenges, which are reflected in the MSE values of Figure 3.5. The MSE, in fact, are higher if compared to the ones of the bidimensional case. Nevertheless, an increase in number of observations clearly improves of the estimations in terms of MSE. From B.9 we can observe that, with the exception of Frank's case, the median of the boxplots is aligned with the true value already at $N = 100$, and the dispersion of the estimations decreases with the growth of N.

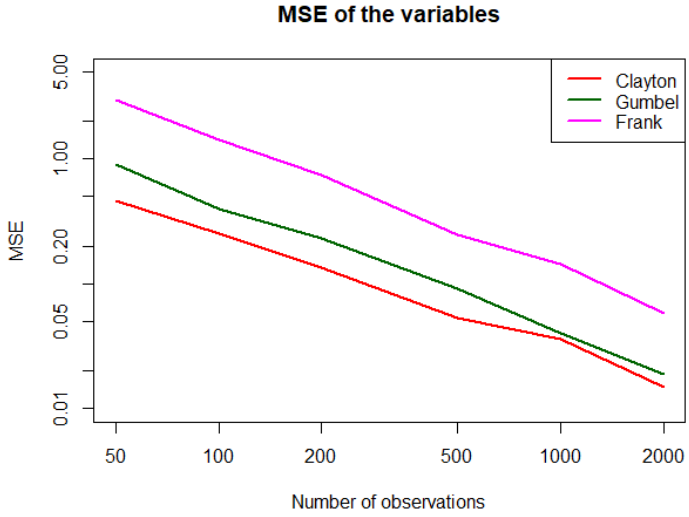


Figure 3.5: *Logarithmic scaled plot studying the convergence of the fitted parameters of the considered copulas. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error values.*

Concluding the analysis, we evaluate the performance of the estimators for Kendall's τ in Figures B.10 and B.11. Similar to the bidimesnional case, these results are much more accurate that the estimators of the parameters. Moreover, we have already good estimations for $N = 50$.

3.6 Conditional Copulas and the Simplifying Assumption

Particularly relevant for the introduction of vine copulas, that will be address in the following chapter 4, is the discussion about *conditional copulas*. Representing an extension of traditional copulas, conditional copulas are constructed from conditional distributions. They can be extremely complex, especially when we are trying to perform a model inference. The *simplifying assumption* can help to reduce their complexity since it implies that the copula does not depend directly on the conditioning variables. In this section, we will introduce conditional copulas, discuss the importance of the simplifying assumption, and some methods that can be used to assess whether this assumption is verified for a given copula.

3.6.1 Conditional Copulas

If we have a random vector $\mathbf{X} \in \mathbb{R}^d$, we can divide it in two sub-vectors $\mathbf{X}_I = (X_1, ..., X_p)$ and $\mathbf{X}_J = (X_{p+1}, ..., X_d)$, such that $\mathbf{X} = (\mathbf{X}_I, \mathbf{X}_J)$, with $I \cup J = \{1, ..., d\}$ and $I \cap J = \emptyset$. From here we can derive the conditional copula of the sub-vector $\mathbf{X}_I$, given the information of $\mathbf{X}_J$.

Definition 13 (Conditional Copula)**.** *For any $\boldsymbol{u} = (u_1, ..., u_p) \in [0,1]^p$, a conditional copula on the event $(\boldsymbol{X}_J = \boldsymbol{x})$, with $\boldsymbol{x}_J \in \mathbb{R}^{n-p}$ is defined by Sklar's theorem as:*

$$C_{I|J}(\boldsymbol{u}|\boldsymbol{X}_J = \boldsymbol{x}_J) = F_{I|J}(F^{[-1]}_{1|J}(u_1|\boldsymbol{X}_J = \boldsymbol{x}_J), ..., F^{[-1]}_{p|J}(u_p|\boldsymbol{X}_J = \boldsymbol{x}_J)|\boldsymbol{X}_J = \boldsymbol{x}_J), \tag{3.18}$$

where $F_{I|J}(\boldsymbol{x}_I|\boldsymbol{X}_J = \boldsymbol{x}_J) := \mathbb{P}(\boldsymbol{X}_I \leq \boldsymbol{x}_I|\boldsymbol{X}_J = \boldsymbol{x})$ is the conditional CDF of $\boldsymbol{X}_I$ given $\boldsymbol{X}_J$, and where $F^{[-1]}_{1|J}$ is the Pseudo-inverse of this CDF (as previously defined in Theorem 4).

3.6.2 The simplifying assumption

The simplifying assumption is verified when a conditional copula is constant with respect to the value its conditional variables. The definition is the following:

Definition 14 (Simplifying assumption)**.** *Suppose we are given two random sub-vectors $\boldsymbol{X}_I$ and $\boldsymbol{X}_J$, and a conditional copula $C_{I|J}(\cdot|\boldsymbol{X}_J = \boldsymbol{x}_J)$. The conditional copula satisfies the simplifying assumption if the value of the copula does not depend on $\boldsymbol{x}_J = (x_{p+1}, ..., x_n)$ for every $\boldsymbol{u} = (u_1, ..., u_p) \in [0,1]^p$. Hence, the function $\boldsymbol{x}_J \in \mathbb{R}^{n-p} \to C_{I|J}(\boldsymbol{u}|\boldsymbol{X}_J = \boldsymbol{x}_J)$ is a constant function that depends only on $\boldsymbol{u}$.*

If the simplifying assumption is verified, we are then able to have simpler copula set to:

$$C_{I|J}(\mathbf{u}|\mathbf{X}_J = \mathbf{x}_J) := C_{I,s|J}(\mathbf{u}).$$

This last definition is possible since the influence of the conditional vector is passed only through the conditional marginal distributions: the effect is directly on the values of the vector $\mathbf{u}$, and can be therefore removed from the conditional copula itself.

3.6.3 How to test the assumption

Here we will discuss some hypothesis testing used to verify if a copula satisfies the simplifying assumption. The results are taken from [38], where additional tests can be found.

The main objective is to perform a hypothesis testing where the Null hypothesis is:

$$H_0 : C_{I|J}(\cdot|\mathbf{X}_J = \mathbf{x}_J) \text{ does not depend on the value of } \mathbf{X}_J,$$

while the alternative hypothesis H_1 is that this condition is not verified. Derumigny and Fermanian discuss various ways of testing if the above hypothesis is verified, providing a wide overview of the possible methodologies that can be used to test the simplifying assumption. A particular use of the simplifying assumption is discussed in Section 6.6, where it will be used to asses the time dependence of the underlying copula structure.

The simplifying assumption, when confirmed, significantly reduces model complexity, and is particularly helpful in high-dimensional scenarios. The assumption will not be the scope of this thesis, however we provide further details on how to test it in Appendix G. Particularly interesting is the discussion on

the advantages, and possible implications, of adopting the assumption even if we are not able to verify it. [39], [40], and [41] explore contrasting opinions on this matter.

3.6.4 Simulation Study on the Simplifying Assumption

To illustrate the idea behind the simplifying assumption, and to illustrate how it is tested we propose a small simulation study. We will generate data from two copulas: one where the simplifying assumption holds, and another where it does not. Statistical tests will then be applied to assess the assumption on the simulated data.

3.6.4.1 Case 1: The Assumption is Not Verified

We start by constructing a copula that does not satisfy the simplifying assumption. We simulate $N = 500$ observations from a standard normal distribution, which will serve as the conditional variable U_J. The bivariate copula is then simulated as follows:

- From a t-Student copula with parameters $\rho = -0.3$ and $\nu = 3$ if $U_J > 0.7$.
- From a Clayton copula with parameter 5 if $U_J \in [0.3, 0.7]$.
- From a Gumbel copula with parameter 8 if $U_J < 0.3$.

The simplifying assumption is then tested on the simulated data using CondCopulas package's function bCond.simpA.CKT, which identifies a set of relevant partitions on the dataset using a binary trees that compares the Kendall's taus [42]. The function also provides a p-value to indicate if the simplifying assumption is verified. More details on this function will be discussed in Section 6.6.

The identified decision tree is illustrated in Figure 3.7, where three different partitions are identified:

- $u_J < 0.332$;
- $0.332 < u_J < 0.716$;
- $u_J > 0.716$.

These partitions are very close to the original partitions used to simulate the dataset. Additionally, the p-value of the test is equal to 0, indicating that the simplifying assumption is not verified.

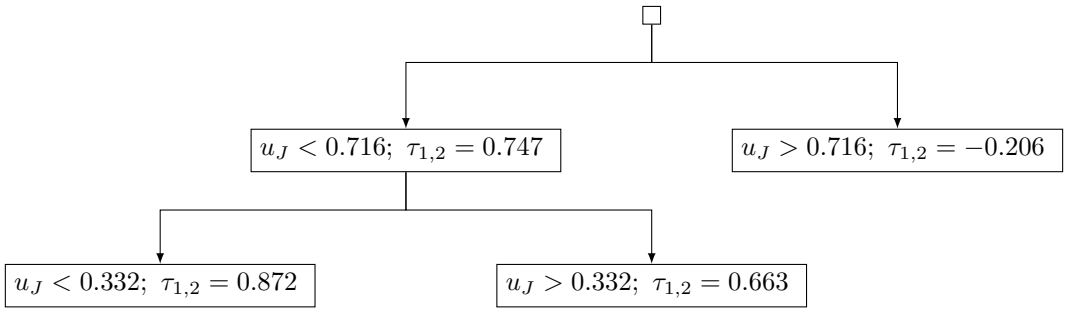


Figure 3.6: *Decision Tree of the partitions identified on the simulated dataset of $N = 500$ samples. The tree indicates the value of the conditional variable U_J, and the Kendall's tau of the bivariate copula in the identified partition, $\tau_{1,2}$.*

3.6.4.2 Case 2: The Assumption is Verified

Subsequently, a bivariate copula of $N = 500$ points is simulated from a t-Students with parameters $\rho = -0.3$ and $\nu = 3$. Here, the simplifying assumption is verified, as the model does not depend on any

conditional variable.

We again test the simplifying assumption, using the procedure described in the previous example, conditioning on the same simulated data, u_J. The resulting tree is reported in Figure 3.7, where four different partitions are identified.

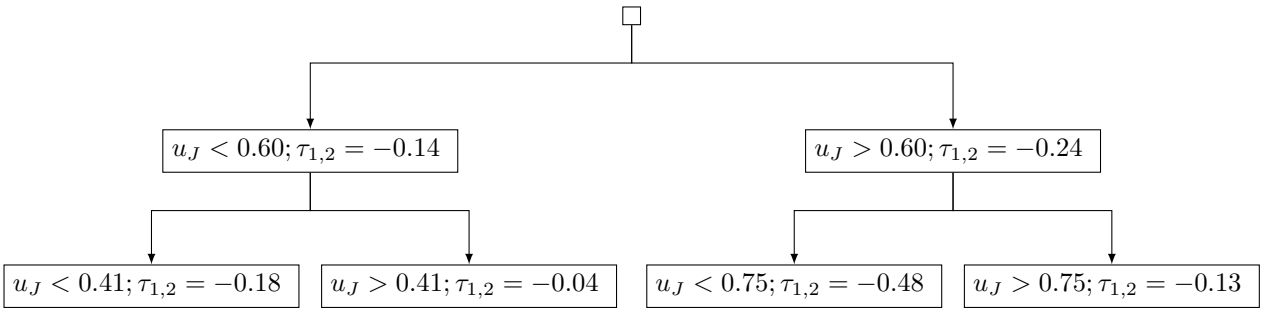


Figure 3.7: *Decision Tree of the partitions identified on the simulated dataset of $N = 500$ samples. The tree indicates the value of the conditional variable U_J, and the Kendall's tau of the bivariate copula in each identified partition, $\tau_{1,2}$.*

However, the p-value of the test is equal to 1, indicating that the simplifying assumption is verified, as we do not have enough evidence to reject it.

4

Vine Copula Models

When working in higher dimensions, simple copula models can result to be "restrictive with regard to symmetry and tail dependence properties" [43]. In fact, each copula family is characterized by specific properties that might not be shared by all variables of a dataset we aim to model. For instance, Gaussian copulas exhibit tail independence, while t-Student copulas display high values of tail dependence. A dataset may exhibit both features, depending on the variable pair under consideration. Given that there are not many families available in dimensions higher than 2, simple copulas might not be able to incorporate all the difference of information that characterize a multivariate dataset, particularly in terms of tail dependence and symmetry. We are therefore in need of a more flexible tool, able to capture specific behaviours of high dimensional datasets. In 2002 Bedford and Cooke proposed a new concept that resolved this limitations by developing a model that uses only bivariate copulas: the vine models [5].

In this section we will discuss the definition of Vine models, their construction, and the estimation of the parameters as presented in [39] and [31]. We will focus on Regular vines, which are generally referred to as R-vines.

4.1 Decomposition in three dimensions

Vine models are obtained by decomposing multivariate distributions into bi-variate copulas, achieved through conditioning. We illustrate the intuition behind their construction, by decomposing a three-dimensional density. The computations can be generalized to higher dimensions.

In Section 3.1 we saw how, using Sklar's theorem, we can compute the density of a vector of d-dimensional random variables $\mathbf{X}$ in terms of the corresponding copula. The formulation obtained was presented in equation 3.2. Moreover, the conditional density of a random variable can be rewritten as the product of a copula density, and the marginal distribution of the conditional variable [27]:

$$f_{1|2}(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2))f_2(x_2). \tag{4.1}$$

By using these two equations, we are able to derive the density $f(x_1, ..., x_d)$ using recursive factorization. The vine structure and decomposition, however, are not unique. For example, in the case of

R-vines, there are $d! \times 2^{\frac{(d-2)(d-3)}{2}-1}$ possible combinations: finding the correct one is one of the main challenges of dealing with vine models.

We can, for example, factorize the density of a 3-dimensional dataset as:

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1),$$

where $f_{i|J}$ is the conditional density of X_i given the random vector $\mathbf{X}_J$, and f_i is the marginal density if the i-th variable. By applying Eq.(3.2), we can rewrite the following conditional density as:

$$f_{13|2}(x_1, x_3|x_2) = c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|F_2(x_2)) f_{1|2}(x_1|x_2) f_{3|2}(x_3|x_2).$$

From Bayes' theorem, we have that:

$$f_{3|12}(x_3|x_1, x_2) = f_{13|2}(x_1, x_3|x_2) f_{3|2}(x_3|x_2).$$

Lastly, from equation 4.1:

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2),$$

$$f_{3|2}(x_3|x_2) = c_{32}(F_3(x_3), F_2(x_2)) f_3(x_3).$$

Collecting all the previous computations, we obtain that the original density can be expressed as the combination of three bivariate copulas, one of which is a conditional copula:

$$\begin{aligned} f(x_1, x_2, x_3) =& c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2) c_{12}(F_1(x_1), F_2(x_2)) \\ & \times c_{32}(F_3(x_3), F_2(x_2)) f_1(x_1) f_2(x_2) f_3(x_3). \end{aligned} \tag{4.2}$$

The set of bivariate copula that composes the above equation can be represented with a tree structure, where each bivariate copula represents an edge. This is only one of the possible three combinations of vine model that we can formulate for a three-dimensional copula: we can, however, consider this a generalisation since we can always reassign the variables by renaming them. Also, let's note that under the Simplifying Assumption, discussed in Definition 14, we could rewrite the conditional copula as

$$c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2) = c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)),$$

resulting in the final formulation of the simplified vine:

$$\begin{aligned} f(x_1, x_2, x_3) =& c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) c_{12}(F_1(x_1), F_2(x_2)) \\ & \times c_{32}(F_3(x_3), F_2(x_2)) f_1(x_1) f_2(x_2) f_3(x_3). \end{aligned} \tag{4.3}$$

4.2 Definition

In [5] Bedford and Cooke derived a new class of graphical models, obtained from a nested set of Markov trees, by weakening their conditional independence property. The new models are called *vine models*, and they aim to study multivariate distributions by proposing a more flexible model that deploys only bivariate conditional copulas.

A Markov tree distribution of a random vector is a tree where, two nodes that are not connected

by an edge, will be independent conditionally on the variables located in the path between them. This refers to the conditional independence property mentioned above. On the other hand, vine is defined as a sequence of trees, where the edges of each tree T_i are the nodes of T_{i+1}. This is translated in the *proximity condition*, which states that an edge between the nodes of T_i is only possible if the corresponding edges in T_{i-1} share a node [43]. By remembering that a graph is defined to be *acyclic* if any two vertices are connected by exactly one path, then we can state the definition of *tree*:

Definition 15 (Tree)**.** *A tree $T = \{N, E\}$ is an undirected acyclic graph, where N is the set of nodes, and E is the set of edges (unordered pairs of nodes).*

As previously discussed, the decomposition of Equation 4.2, and the corresponding higher dimensionsional version, allow for different conditioning orders, and hence for different structures. Three main type of vines can be identified from different tree construction ordering, and specific graphical rules: the regular vines, or R-vines, the canonical vines (or C-vines), and the drawable vines (or D-vines).

R-vines, which were first introduced in [44], are the most common type of vines, as they represent the most general case. C and D-vines represent special cases of vine structures. In fact, C-vines are characterized by trees with a star-like structure around a single node, called the *root*, that connect all other nodes. D-vines have path structures for each tree, meaning that every node has a maximum of two neighbours, and that the structure of the first tree determines all the following ones [43]. In this thesis we will focus on R-vines, which will be presented more carefully in the following Subsection. For a more in depth analysis of C- and D-vines the reader can refer to section 4.2 of [27].

4.2.1 R-vines

A regular vine, or an R-vine, is characterized by a sequence of trees where each set of edges becomes the nodes for the following tree. They are defined as follows:

Definition 16 (Regular Vine)**.** *The set of graphs $\mathcal{V} = (T_1, \dots, T_{d-1})$ is a regular vine tree sequence on d elements if:*

1. *T_1 is a tree with node set $N_1 = 1, ..., d$ and edge set E_1.*
2. *For $j \geq 2$, T_j is a tree with node set $N_j = E_{j-1}$ and edge set E_j .*
3. *For $j = 2, ..., d-1$ if two nodes in T_j are joined by an edge, the nodes must share a common node in T_{j-1} (Proximity condition).*

We also introduce an alternative notation, which allows for more compact formulations. For any edge $e_i \in E_i$ we define:

$$A_e = \{j \in N_1 | \exists e_1 \in E_1, ..., e_{i-1} \in E_{i-1}, s.t., j \in e_1 \in ... \in e_{i-1} \in e_i\}, \tag{4.4}$$

the complete union. The conditioning set of an edge $e = \{a, b\}$ is then defined as $D_e := A_a \cap A_b$, and the conditioned sets $C_{e,a}$ and $C_{e,b}$ are defined as:

$$C_{e,a} = A_a \backslash D_e \text{ and } C_{e,b} = A_b \backslash D_e. \tag{4.5}$$

R-vines, being the most general type of vines, exhibit an elevated number of possible combinations, reaching super-exponential growth for the number of variables d. This might pose a challenge in higher dimensions; to address this issue, and to ensure that a valid factorization is produced, different tools have been developed.

4.3 Vine Models

In this Section we take a further step forward, providing the definition of an R-vine model, and by discussing how to visualize its components in a more compact way.

Definition 17 (Vine Model)**.** *A vine model $\mathcal{M}_V$ of a d-dimensional random vector $\boldsymbol{X}$ is characterized by the triplet $(\mathcal{F}, \mathcal{V}, \mathcal{B})$. Where $\mathcal{F}$ is the vector of the marginal distributions of each random variable of $\boldsymbol{X}$, $\mathcal{V}$ is the R-vine sequence of tree structures of a total of d elements, and $\mathcal{B} = \{C_e \mid e \in E_i; i = 1, \ldots, d-1\}$ is the set of copula C_e, where E_i represents the edge set of the tree T_i.*

Considering a simplified R-vine model, to which the simplifying assumption has been applied, [5] provides a theorem that guarantees the existence of the regular vines:

Theorem 6 (Existence of R-vine distribution)**.** *Assume that $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ satisfy the three properties that define a regular vine (Def. 16), then there exists a d-dimensional distribution F with:*

$$f_{1,\ldots,d}(x_1, \ldots x_d) = f_1(x_1) \times \cdots \times f_d(x_d) \times \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{\mathcal{C}_{e,a}\mathcal{C}_{e,b};D_e}\left(F_{\mathcal{C}_{e,a}|D_e}\left(x_{\mathcal{C}_{e,a}} \mid \boldsymbol{x}_{D_e}\right), F_{\mathcal{C}_{e,b}|D_e}\left(x_{\mathcal{C}_{e,b}} \mid \boldsymbol{x}_{D_e}\right)\right),$$

such that for each $e \in E_i, i = 1, \ldots, d-1$, with $e = \{a, b\}$ we have for the distribution function of $X_{C_{e,a}}$ and $X_{\mathcal{C}_{e,b}}$ given $\boldsymbol{X}_{D_e}$

$$F_{\mathcal{C}_{e,a}\mathcal{C}_{e,b}|D_e}\left(x_{\mathcal{C}_{e,a}}, x_{\mathcal{C}_{e,b}} \mid \boldsymbol{x}_{D_e}\right) = C_e\left(F_{\mathcal{C}_{e,a}|D_e}\left(x_{\mathcal{C}_{e,a}} \mid \boldsymbol{x}_{D_e}\right), F_{\mathcal{C}_{e,b}|D_e}\left(x_{\mathcal{C}_{e,b}} \mid \boldsymbol{x}_{D_e}\right)\right).$$

The one-dimensional margins of F are given by $F_i(x_i), i = 1, ..., d$.

4.3.1 Compact Visualization

Generally, a vine model is presented by showing the set of trees graphically, where each edge represents a specific pair copula. In addition, [45] developed a compact representation of the main components of the model that made possible to represent the tree structure $\mathcal{V}$ using a $d \times d$ triangular matrix V. The label of the j-th edge of tree T_t is obtained as:

$$(V_{d-j+1,j}, V_{t,j} | V_{t-1,j}, ..., V_{1,j}).$$

Moreover, as for copulas, we can store the set of families of all bivariate copula that compose the vine (each family has a corresponding integer value) in an upper $d \times d$ triangular matrix, which has entries zero on the diagonal. Lastly, an upper triangular matrix Θ can be used to collect the parameter of the bivariate copulas that compose the vine model: $\theta_{t,j}$ is the parameter of the j-th edge in the t-th tree.

To better understand the construction of matrix $\mathbf{V}$, which can be challenging, we discuss its construction for the 4-dimensional vine structure with the following decomposition:

$$\begin{aligned} & f(x_1, x_2, x_3, x_4) \\ =& c_{14;23}\left(F_{1|23}(x_1 \mid x_2, x_3), F_{4|23}(x_4 \mid x_2, x_3); x_2, x_3\right) \\ & \times c_{13;2}\left(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2); x_2\right) \times c_{24;3}\left(F_{2|3}(x_2 \mid x_3), F_{4|3}(x_4 \mid x_3); x_3\right) \\ & \times c_{34}(F_3(x_3), F_4(x_4)) \times c_{23}(F_2(x_2), F_3(x_3)) \times c_{12}(F_1(x_1), F_2(x_2)) \\ & \times f_4(x_4) f_3(x_3) f_2(x_2) f_1(x_1). \end{aligned} \tag{4.6}$$

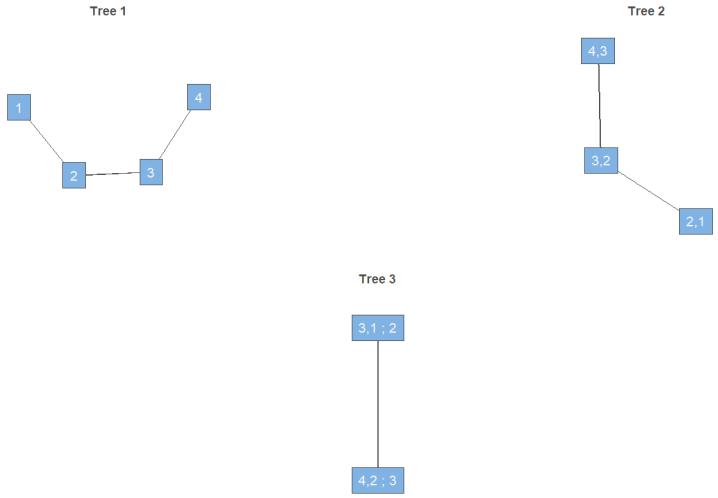


Figure 4.1: *Structure of the tree presented in Equation* (4.6).

The set of trees that correspond to equation (4.6) is presented in Figure 4.1, while the equivalent matrix representation is given by:

$$V = \begin{bmatrix} 2 & 3 & 4 & 4 \\ 3 & 4 & 3 & 0 \\ 4 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

How to interpret matrix V in practice? From the elements of the matrix we are able to compile the edges of each tree. The j-th edge of tree T_t is obtained by the pair $((V)_{d-j+1,j}, (V)_{t,j})$, conditioned to the elements $(V_{t-1,j}, ..., V_{1,j})$.

In more simple words, the goal is to construct a set of trees trees T_t, $t = \{1, \ldots, d-1\}$, by using the elements on rows i, and columns j. In particular, to label element j of the t-th tree, we start from the element on the counter-diagonal of column j, hence $(V)_{d-j+1,j}$, and pair it with the element in row t and column j, hence pair it with element $(V)_{t,j}$. Then we condition on all the elements of the same column j, located above element $(V)_{t,j}$ [43]. Let's analyze some the steps required to construct Figure 4.1, from V.

We illustrate how to construct the elements of three T_2 $(t = 2)$, which will be characterized by two pair copulas. The first copula, corresponding to $j = 1$, is composed starting from the element located on the first column of the counter-diagonal, hence from $(V)_{d-j+1,j} = (V)_{4,1} = 1$. The element is paired with $(V)_{t,j} = (V)_{2,1} = 3$. We then condition on the only element above $(V)_{2,1}$, which is $(V)_{1,1} = 2$. We therefore obtained the first edge of tree T_2: the pair copula $c_{13|2}$. Similarly we can determine the element of the second edge, which will be $c_{24|3}$.

4.3.2 Estimation and Selection of the Model

To summarize, each vine model is constructed of three main parts: the tree structure V, the set families that characterize each bivariate copula, and the corresponding set of parameters. When we want to select a proper vine model, we need to evaluate and estimate these three .components. However, not all

problems are equally complex, in fact, as described in [43], estimating a vine copula model can imply solving three different problems, presented here with increasing complexity:

- Given the vine tree sequence and a set of pair copula families, estimate the copula parameters.
- Given the vine tree sequence and a catalogue of pair copula families, select the best family and estimate the corresponding parameters for each edge in the vine.
- Select the vine tree structure, the pair copula families and estimate the corresponding parameters for each edge.

The last problem, particularly complex and potentially infeasible in higher dimensions, requires some prior assumptions about the family used, and about the structure. The general approach would involve evaluating all the possible combinations and choosing the one with the best Log, AIC, or BIC value (for a more complete overview of the selection criteria adopted, check Def. 41 of the Appendix), however we will discuss some simplifications that ensures a more feasible process.

Firstly, to estimate the parameters, we tend to prefer a step-wise approach, rather than directly using the Maximum Likelihood estimator, which will be extensively discussed in the following Subsections. Another simplification can derive from the *simplifying assumption*, introduced in Definition 14. However, checking for the simplifying assumption in the context of vine models implies testing the assumption for every pair-copula. This elevates the computational costs, and the complexity of the problem. Sometimes, it is therefore simply assumed to hold true even without checking if it truly holds.

4.4 Structure and Family Selections

Assuming that we are in the most complex scenario, hence that we need to select the vine tree structure and the pair copula families, while estimating their corresponding parameters for each edge, the following steps can be followed. Typically, the parameters' estimation, the tree structure selection,e and the pair copula selection, are conducted simultaneously. The results are then compared, and the model with a lower AIC or BIC value is chosen. To simplify the explanation, we will discuss each procedure separately.

In particular, in Subsection 4.4.1 we will explain the Sequential Method, used to select an appropriate vine structure by sequentially determining the edges of the tree structure. Subsection 4.4.2 will discuss how to choose among different combinations of pair copulas' families. The parameter estimation process will be presented in Section 4.5.

4.4.1 The Sequential Method for Structure Selection

Assuming that we have no prior assumption on the vine model that we need to select, the first thing to do is to select a proper vine structure. For R-vines the number of possible structures of a d-dimensional vine model is $\frac{d!}{2} \times 2^{\binom{d-2}{2}}$. Since this number increases rapidly with the dimension d, trying all possible combinations or evaluating manually the best one can become extremely challenging and time consuming, if not infeasible. To overcome this problem, [46] developed the *sequential method*, an algorithm that selects each trees sequentially, by linking the pairs that show the strongest dependence.

The method uses Kendall's τ (see Eq. (3.6)) as a measure of dependence, and applies a maximum spanning tree (MTS) algorithm to find the tree structure that maximizes the sum of the empirical

Kendall's taus $\hat{\tau}_n$ for each tree. At each step we need to maximize the following optimisation problem:

$$\max \sum_{e_{ij} \text{ in spanning tree}} |\hat{\tau}_{ij}|. \tag{4.7}$$

As the name suggests, the algorithm is sequential, we therefore start from the first tree T_1, and choose the complete graph that optimize the objective function (4.7). For all following trees T_i, we then fit a bivariate copula to each pair of edges of the preceding spanning tree that satisfies the proximity condition, in order to evaluate the conditional distributions. This last step allow to estimate the empirical Kendall's taus of the conditioned variables $\hat{\tau}_{j,k|D}$. In the following steps, we proceed to find the structure that maximizes the sum of the estimated τ of the conditional copulas:

$$\max \sum_{e_{ij|D} \text{ in spanning tree}} |\hat{\tau}_{ij|D}|.$$

The output of the algorithms is an R-vine copula specification $(\mathcal{V}, \mathcal{B})$. The method is not guaranteed to converge, but provides some advantages. Firstly, it allows to obtain a sequence in a reasonable amount of time, without computing all possible combinations. Moreover, choosing the pairs with higher dependence first, allows to minimize the influence of rounding errors in later trees, since this type of error is mostly attributable to strong dependence between the variables.

4.4.2 Family Selection

For a given structure of the vine model, we are required to find an appropriate family for each bivariate copula on the vine trees. The selection can be carried out by comparing the value of the Likelihood, AIC, or BIC of each family tested, and by choosing the family that produces the best value. The Bayesian Information Criterion (Def. 41) is usually preferred, as it penalizes a higher number of parameters compared to the other two criteria: we will adopt the BIC to select the set of families in our model.

4.5 Parameter Estimation

When estimating the set of parameters Θ of a vine Model, or in general of pair-copula constructions (or PCC), many estimators that can be chosen. Certainly, the most popular one is the Maximum Likelihood Estimator (see Def. 40) which is asymptotically efficient, but when the number of parameters increases it becomes challenging to use. We therefore tend to prefer methods that treat the margins separately (such as two-step estimators), or methods that estimate the parameters at each level by conditioning on the parameters estimated on the previous levels. One example of the latter type of method would be the Stepwise semiparametric estimator, or SSP, presented in the following subsection. The section is a summary of [47], which focuses on the case of D-vines, and extend to C-vines (briefly introduced in Subsection 4.2). However, the results can be applied to Regular vines as well.

4.5.1 Step-wise Semiparametric Estimator

The Stepwise Semiparametric Estimator, or SSE, is used when the number of parameters that have to be estimated makes the computation too costly to be done simultaneously. The SSE, in fact, is computed by treating the margins separately, and by estimating all other parameters level by level, conditioning on the ones of the preceding levels of the structure [47].

Assuming we want to estimate the parameters of a d-dimensional vector $\mathbf{X}$, from the set of observations $\mathbf{X}_1, \ldots, \mathbf{X}_d$, will result in $d-1$ levels at which we will estimate the parameters. We start by defining for $j = 1, ..., d-1$ the log pair-copula density of the j-th level of the vine structure for a generic vector $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$. If we define the conditioning set as $D_{i,j} = (X_k)_{k=i+1,\ldots,i+j-1}$, we obtain:

$$\begin{aligned}\psi_j(u_1, \ldots, u_d; \theta_1, \ldots, \theta_j) = \\ = \sum_{i=1}^{d-j} \log(c_{i,i+j|D_{ij}}(F_{i|D_{ij}}(F_i^{-1}(u_i)|F_{i+1}^{-1}(u_{i+1}), \ldots, F_{i+j-1}^{-1}(u_{i+j-1}); \theta_{i\to i+j-1}), \\ F_{i+j|D_{ij}}(F_{i+j}^{-1}(u_{i+j})|F_{i+1}^{-1}(u_{i+1}), \ldots, F_{i+j-1}^{-1}(u_{i+j-1}); \theta_{i+1\to i+j}); \theta_{i+j|D_{ij}})).\end{aligned} \tag{4.8}$$

$c_{i,j+1|D_{ij}}$ is the conditional copula density of the pair of variables (X_i, X_{i+j}), given D_{ij}; the set of parameter $\theta_{i\to j} = \{\theta_{s,s+t|D_{st}} : (s, s+t) \in \{i, ..., i+j\}\}$. F_i are the marginal CDFs, and F_i^{-1} represent the inverse of the same CDF, for each variable X_i. (4.8) refers to the particular case of D-vines copulas, and is formulated under the simplifying assumption since, without this simplification, inference on the models would not be feasible in practice [47].

At each level j of the vine model we estimate the parameter $\hat{\theta}_j^{SSP}$, by conditioning on the parameters estimated in the preceding level, and by maximising the pseudo-likelihood of the j-th level $\hat{l}_j$, with respect to θ_j. Where $\hat{l}_j$ is the sum of the log pair-copula densities, computed by applying the empirical marginal distributions $\hat{F}_i$ to the corresponding observations $\mathbf{x}_i = (x_{i1}, \ldots, x_{iN})$:

$$\hat{l}_j(\theta_1, ..., \theta_j; x) = \sum_{k=1}^{N} \sum_{l=1}^{j} \psi_l(\hat{F}_1(x_{1k}), ..., \hat{F}_d(x_{dk}); \theta_1, ..., \theta_l)). \tag{4.9}$$

Moreover, [47] discusses the theoretical convergence and the asymptotic properties of the estimator. In fact, under certain conditions which are listed by Tsukahara in [48] , $\hat{\theta}^{SSP}$ results to be consistent for the set of parameters θ and asymptotically normal. According to [47], $\hat{\theta}^{SSP}$ exhibits a rate of convergence of $\frac{1}{\sqrt{n}}$, which we will address in the following Section through a simulation study.

4.5.2 Goodness of Fit

Lastly, as previously discussed for copulas in Section 3.4.3, once we have selected a models we will be left to evaluate it, using one of the many Goodness of Fit tests available.

Similarly to the copula case, one of the test that we will use to asses the goodness of the fitted model is an hypothesis test based on White's equality [36], which was already discussed in Section 3.4.3.1. The test's version for Vine models is implemented in the R function "RVineGofTest" of the VineCopula package [34].

Another test would be the Vuong test (Appendix 3.4.3.2), an hypothesis test used to compare two chosen distributions, by analyzing their Kullback–Leibler Information criterion. A practical application of this test will be illustrated in the following simulation study. For a further discussion on the hypothesis tests available, please refer to Appendix F.4.

4.6 Simulation Study

To investigate the properties of the estimator $\hat{\theta}^{SSP}$ outlined in Subsection 4.5.1, and to offer a more practical visualization of a simple vine model, we conducted a small simulation study.

The experiment has two main objectives, each carried out in separate stages. The first is to visualize the convergence of parameter estimators by increasing the number of observations, and computing the mean squared error (MSE). The second goal is to assess the consistency and accuracy of these estimators for a fixed number of observations N, by estimating the parameters of a 3-dimensional Gaussian vine copula. Different vine structures, which not necessarily corresponded to the original structure, are considered and compared. The experiment is then extended to a t-Student and a Gumbel vine copula.

For all simulations, the original data is simulated from a 3-dimensional Gaussian vine copula, characterized by the same structure shown in Figure 4.2, which we later refer to as *structure* V_1. The corresponding matrix is (refer to section 4.3.1 for an explanation of how to read the matrix):

$$V_1 = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 1 & 0 \\ 2 & 0 & 0 \end{bmatrix}.$$

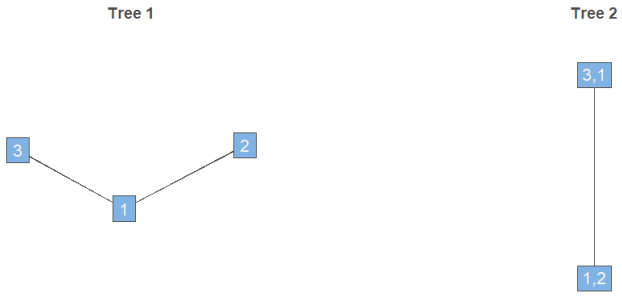


Figure 4.2: *Structure of the tree chosen for the original model.*

In part two of the simulation we will also use the two alternative structures allowed by a 3d vine model, which will be referred to as V_2, and V_3. The corresponding matrix will be:

$$V_2 = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 0 \\ 2 & 0 & 0 \end{bmatrix},$$

and

$$V_3 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 3 & 0 & 0 \end{bmatrix}.$$

4.6.1 Part I: Estimators convergence

We started by performing an analysis of the rate of convergence of the estimators of the three parameters of a 3-dimensional Gaussian vine model of known structure. As already anticipated, the structure considered is the one presented in Figure 4.2. We know that the theoretical rate of convergence should be $\frac{1}{\sqrt{N}}$, with N being the number of observations.

We assigned the following values to the vine's parameters: $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (-0.8, -0.1, 0.4)$, which correspond to the values of Kendall's Tau $(\tau_{12}, \tau_{31}, \tau_{32|1}) = (-0.59, -0.06, 0.26)$. We assumed the structure and the family (Gaussian) known, estimating only the three parameters. We considered different number of observations N, which were set equal to $N = \{50, 100, 200, 500, 1000, 2000\}$, and for each N, we estimated the set of parameters. For each N, the experiment was replicated 100 times, and computed the mean squared error (MSE) of each parameter from the true values. The results are shown in Figure 4.3 of the Appendix, where we plotted the rate of convergence of the estimated variables in logarithmic scale. In Figure B.12 of the Appendix, we provided the corresponding boxplots of the estimated values.

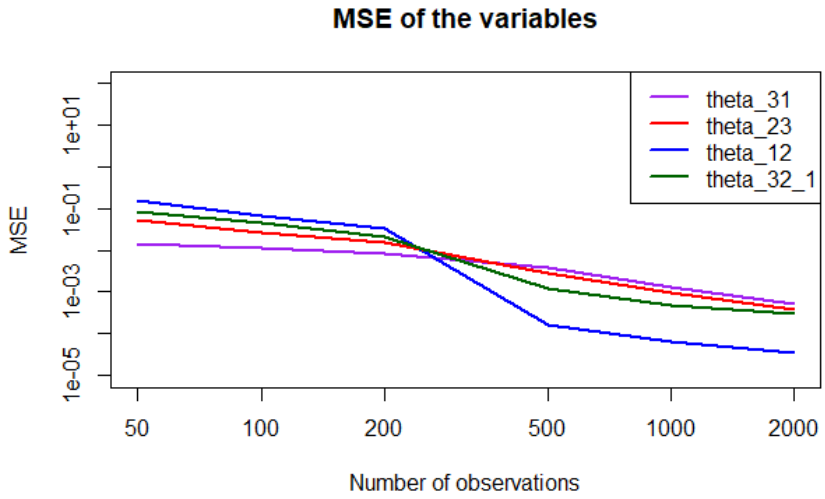


Figure 4.3: *Logarithmic scaled plot of the convergence of the three estimated variables, and the conditional parameter. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error values.*

From Figure 4.3, we can indeed conclude that all estimated parameters have a MSE that converges to zero, with a speed that is coherent with the theoretical rate of $\frac{1}{\sqrt{N}}$. We also included the correlation parameter θ_{23}, which was computed manipulating Equation (4.10), a property of the correlation of Gaussian distributions [47]:

$$\theta_{ij|k} = \frac{\theta_{ij} - \theta_{ik}\theta_{jk}}{\sqrt{(1-\theta_{ik}^2)(1-\theta_{jk}^2)}}. \tag{4.10}$$

The trend and speed of convergence are similar for all parameters, except for θ_{12}, which shows a steeper convergence, reaching 0 slightly faster than the other estimators. The result is coherently expressed by the boxplots of Figure B.12, where only θ_{12} has sufficiently good results already for a

sample size of $N = 50$, while remaining estimators only begin to align their medians with the true values, and to exhibit reduced variability only around $N = 500$. This value is higher than the result obtained for copulas in Section 3.5.4, where we performed a similar experiment. For the copulas case, in fact, we concluded that $N = 200$ was a sufficient amount of data to reach satisfactory estimations. This necessity for a larger dataset to attain accurate results, can be attributed to the higher complexity of vine models, which involve more estimation steps, leaving more chances for errors in the estimations.

Additionally, for $N = 50$ and $N = 100$, the estimated values for θ_{31} shown in the boxplot of B.12 are concentrated around zero: this is a sign that the number of observations was not sufficient, as 0 is generally returned by the algorithm when an estimate is failed to be produced. Lastly, trying to explain the better performance of θ_{12}, we notice that it is the parameter with the highest absolute value: a higher correlation is more easily identifiable from the data, which would explain the steeper convergence.

To validate this last observation that a higher absolute value of correlation leads to faster convergence of the estimator, we added a further step to the experiment. We fixed the parameters $\theta_{31} = -0.5$ and $\theta_{23|1} = 0.4$, and replicated the previous experiment for the following values of $\theta_{12} = \{-0.9, -0.7, -0.5, -0.3, -0.1, 0.1, 0.3, 0.5, 0.7, 0.9\}$, to see if different values would lead to different convergence rates. The experiment can be generalized by simply changing the variables' name. For each value of the parameter θ_{12}, we performed a nonlinear regression using the Least Squares method to estimate the parameters α and β of the convergence rate curve: $\log MSE(N) = \alpha \log N + \beta$. Where α, is the rate of convergence, while β is the intercept of the fitted linear relationship.

Based on the results obtained, we were able to draw the conclusion that indeed the observation holds true. If fact, we were able to observe how the value of the coefficient α became more negative if we increased the absolute value of the parameter θ_{12}. Moreover, for all values considered, the convergence rate was always equal, or below, the theoretical $\alpha = -\frac{1}{2}$, aligning our results with the theoretical rate of convergence discussed in section 4.5 (in fact, $\alpha = -\frac{1}{2}$ corresponds to a linear rate of convergence of $\frac{1}{\sqrt{N}}$). We conducted additional iterations of the experiment using different values of θ_{32}, and the results were consistent and comparable.

4.6.2 Part II: Vine structures Equivalency

In the second part of the simulation study, we generated a three-dimensional Gaussian vine model with $N = 1000$ observations, following the structure presented in Figure 4.2. The choice of 1000 observations followed from the first part of our study, as it was found to guarantee estimates with reasonable accuracy. This observation held true for all parameters and copula families under consideration.

To conduct the simulation, we vary the combinations of the parameters $(\theta_{12}, \theta_{31}, \theta_{32|1})$ (as listed in Table 4.1). For each parameter set, we replicate the experiment 100 times. Subsequently, we fit a vine model to the resulting data frame, assuming it was generated from Gaussian bivariate distributions with unknown parameters. We perform this process for all three possible vine structures allowed by a 3D vine model, hence for structures V_1, V_2, and V_3.

Figure B.13 illustrates some of the results obtained. For each parameters considered, we plotted the boxplot displaying the 100 different estimated values, divided by the underlying vine structure used to fit the model. This enables to study the influence of different structures on the estimation, and to

compare the results with the true value, represented by a red dashed line. Additionally, we compared, two by two, the models with different underlying vine structures, using Vuong test, to assess if they were equivalent. We repeated the test for all 100 iterations, and in Table 4.1 we provided the percentages of times the test did not reject the null hypothesis: hence we provide the percentage of how many times two models resulted equivalent fits to the original one.

$(\theta_{12}, \theta_{31}, \theta_{23\|1})$	$(\tau_{12}, \tau_{31}, \tau_{32\|1})$	Vuong V_3 vs V_1	Vuong V_2 vs V_1	Vuong V_3 vs V_2
(0.4, 0.3, 0.8)	(0.26, 0.19, 0.59)	96%	93%	93%
(-0.99, 0.4, 0.6)	(-0.90, 0.26, 0.41)	98%	97%	99%
(0.4, 0.86, 0.1)	(0.26, 0.65, 0.06)	96%	96%	95%
(-0.2, -0.5, -0.9)	(-0.12, -0.33, -0.71)	95%	92%	88%
(-0.4,0.86,0.1)	(-0.26, 0.65, 0.06)	100%	98%	98%
(0.2, -0.5, 0.9)	(0.12, -0.33, 0.71)	97%	95%	91%

Table 4.1: *Percentages of P-values of the Vuong test greater than 0.05, obtained by comparing the results from fitting the different vine structures, compared two by two. The results are reported for all six combinations of parameters that we analyzed. We also provide the corresponding values of Kendall's Tau for each copula analyzed.*

For Gaussian vine copulas the structures should be equivalent. The results obtained from this simulation confirm the theory: the underlying structure does not influence the estimation of the parameters, and the three structures are equivalent. This is evident by the high percentages shown in Table 4.1, and from the boxplots B.13 that display very similar distributions of the estimated parameters. Additionally, we can observe that the estimation is generally successful, with the exception of a few outliers (mainly with a value $\theta = 0$, which indicates that the algorithm encountered an error). The estimated parameters have a mean that coincides with the true one for all models.

Lastly, we extended the previous experiment by relaxing some restrictions. In the new experiment, we simulate a Gaussian vine model with $N = 1000$ observations, for each parameter set provided in Table 4.1. But, instead of imposing a predefined set of families for the bivariate copulas in the vine structure, we let the algorithm determine the best fitting set of families, based on the comparison of BIC values. Although this approach results in slower computations, we expect more satisfactory outcomes.

In Table 4.2, we calculate the frequency at which the bivariate copulas composing the fitted vines, out of the total number of replications ($N = 100$), had the same set of families as the original vine (i.e. were Gaussian bivariate copulas). Due to the variability across iterations, we do not provide boxplots as the different families obtained may have incomparable parameters. Finally, in Table 4.3, we present the frequency of acceptance for the null hypothesis of the Vuong test, computed using the newly fitted vine models.

After analyzing both Tables 4.2 and 4.3, we can once again draw the conclusion that the specific structure of a Gaussian vine model does not significantly impact the results. In fact, from Table 4.2, the occurrence frequency of Gaussian bivariate copulas within the vine structure is consistently high, with values exceeding 50% for the majority of cases. Moreover, there is a general accordance among the different structures chosen, hence the frequency tends to be consistent for all three structures considered. Table 4.3 also suggests equivalency the three models, as the frequencies are coherent and similar to the ones observed in Table 4.1.

$(\theta_{12}, \theta_{31}, \theta_{32\|1})$	Structure V_3 $(c_{12}, c_{31}, c_{32\|1})$	Structure V_1 (True) $(c_{12}, c_{31}, c_{32\|1})$	Structure V_2 $(c_{12}, c_{31}, c_{32\|1})$
(0.4, 0.3, 0.8)	(78%, 96%, 10%)	(78%, 60%, 98%)	(60%, 96%, 54%)
(-0.99, 0.4, 0.6)	(96%, 78%, 92%)	(96%, 78%, 92%)	(78%, 78%, 88%)
(0.4, 0.86, 0.1)	(78%, 68%, 96%)	(78%, 100%, 10%)	(100%, 68%, 26%)
(-0.2, -0.5, -0.9)	(50%, 86%, 94%)	(50%, 82%, 90%)	(82%, 86%, 88%)
(-0.4, 0.86, 0.1)	(78%, 68%, 94%)	(78%, 92%, 22%)	(92%, 68%, 54%)
(0.2, -0.5, 0.9)	(36%, 90%, 90%)	(36%, 76%, 92%)	(76%, 90%, 92%)

Table 4.2: *Frequency that indicates how many times, for each fixed vine structure, the fitted set of families of the bivariate copulas coincide with the original set (gaussian). Only family coincidence is considered, without regard to parameters.*

$(\theta_{12}, \theta_{31}, \theta_{32\|1})$	Vuong V_3 vs V_1	Vuong V_2 vs V_1	Vuong V_3 vs V_2
(0.4, 0.3, 0.8)	96%	92%	100%
(-0.99, 0.4, 0.6)	98%	84%	84%
(0.4, 0.86, 0.1)	100%	100%	94%
(-0.2, -0.5, -0.9)	90%	84%	90%
(-0.4, 0.86, 0.1)	94%	100%	94%
(0.2, -0.5, 0.9)	98%	80%	92%

Table 4.3: *Frequency of P-values of the Vuong test greater than 0.05, obtained by comparing the results from fitting the different vine structures, compared two by two. The results are reported for all six combinations of parameters that we analyzed.*

4.6.3 t-Student Vine Copula

Subsequently, we conducted two additional experiments by modifying the vine model family to a t-Student copula at each level. In the first experiment, we fixed the parameters to be $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (-0.9, -0.1, 0.4)$, and $(\nu_{12}, \nu_{31}, \nu_{32|1}) = (3, 4, 5)$. For the second experiment, we varied the parameter values according to the specifications provided in Table 4.4, and set the degrees of freedom to $\nu = 3$ for all bivariate pairs. The results obtained from these experiments are presented in the following Figures B.14, and B.15 of the Appendix, and in Tables 4.4, and 4.5.

The convergence plot (Figure 4.4) appears again faster than the theoretical rate, similarly to the results obtained for the Gaussian case. Specifically, the parameter $\theta_{12} = -0.9$ has a faster convergence compared to the other parameters. This behavior can be attributed to θ_{12} having the highest absolute value, which aligns with the observations made in the Gaussian case. We also plotted the convergence of the estimators for the degrees freedom ν, which appear to have much slower convergence compared to the parameter θ. This discrepancy might be attributed to the broader range of potential values for ν, as it can assume values in the interval $[1, \infty]$.

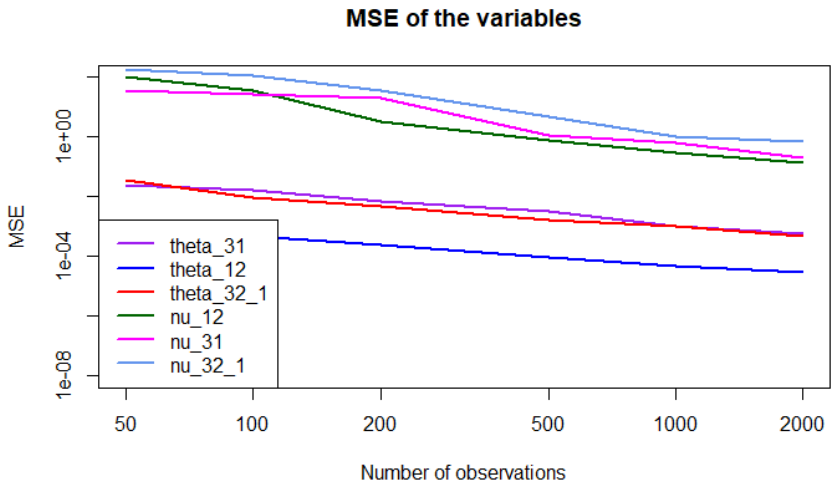


Figure 4.4: *Plot of the evolution of the convergence of the three variables, and the conditional parameter. On the x-axis we have the number of observations on a logarithmic scale, and on the y-axis the corresponding Mean Squared Error values.*

For the second part of the experiment, we simulated $N = 1000$ points from t-student vine copulas with tree structure of 4.2, the same used for the Gaussian case, and using the parameters listed in Table 4.4.

$(\theta_{12}, \theta_{31}, \theta_{23\|1})$	$(\tau_{12}, \tau_{31}, \tau_{32\|1})$	Vuong V_3 vs V_1	Vuong V_2 vs V_1	Vuong V_3 vs V_2
(0.4, 0.3, 0.8)	(0.26, 0.19, 0.59)	95%	99%	100%
(-0.99, 0.4, 0.6)	(-0.90, 0.26, 0.40)	94%	33%	43%
(0.4, 0.86, 0.1)	(0.26, 0.65, 0.06)	67%	79%	95%
(-0.2, -0.5, -0.9)	(-0.12, -0.33, -0.71)	81%	93%	84%
(-0.4,0.86,0.1)	(-0.26, 0.65, 0.06)	55%	67%	92%
(0.2, -0.5, 0.9)	(0.12, -0.33, 0.71)	77%	90%	86%

Table 4.4: *Percentage of p-value of the Vuong test was greater than 0.05. The test was performed by comparing the three different vine structures, two by two. Each vine structure considered is composed of t-Student bivariate copulas, with original parameters indicated in column $(\theta_{12}, \theta_{31}, \theta_{32|1})$. We also provide the corresponding values of Kendall's Tau for each copula analyzed.*

In Figure B.15 we report the results of the vine generated from a t-Student with parameters $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (-0.4, 0.86, 0.1)$, and hence $(\tau_{12}, \tau_{31}, \tau_{32|1}) = (-0.26, 0.65, 0.06)$. From the Figures, we can observe how having the correct underlying structure can positively impact the estimation of the parameters. This is confirmed by the results provided in Table 4.4, where the Vuong test suggests that the three models are not equivalent anymore. For example, in the case $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (-0.4, 0.86, 0.1)$ we have that, while the two alternative structures seem to be equivalent according to the Vuong tests performed (in 92% of cases the two structures result equivalent), if they are compared to the original structure the Vuong test rejects the equivalency in almost half of the cases (55% and 67% the times they result equivalent). Not all set of parameters considered apply to this observation (e.g. $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (0.4, 0.3, 0.8)$, $(\tau_{12}, \tau_{31}, \tau_{32|1}) = (0.26, 0.19, 0.59)$), but in general we can conclude that the choice of a correct structure matters to guarantee a good convergence of the parameters.

Moreover, Table 4.5 presents extremely high frequencies, meaning that, independently of the vine structure chosen, the model selection results almost in all cases in a vine structure model composed by bivariate t-Student copulas.

$(\theta_{12}, \theta_{31}, \theta_{32\|1})$	Structure V_3 $(c_{12}, c_{31}, c_{32\|1})$	Structure V_1 (True) $(c_{12}, c_{31}, c_{32\|1})$	Structure V_2 $(c_{12}, c_{31}, c_{32\|1})$
(0.4, 0.3, 0.8)	(100%, 100%, 100%)	(100%, 98%, 90%)	(98%, 100%, 94%)
(-0.99, 0.4, 0.6)	(98%, 100%, 96%)	(98%, 98%, 96%)	(98%, 100%, 98%)
(0.4, 0.86, 0.1)	(98%, 100%, 70%)	(98%, 92%, 98%)	(92%, 100%, 100%)
(-0.2, -0.5, -0.9)	(100%, 92%, 100%)	(100%, 90%, 96%)	(90%, 92%, 94%)
(-0.4,0.86,0.1)	(100%, 100%, 62%)	(100%, 90%, 100%)	(90%, 100%, 100%)
(0.2, -0.5, 0.9)	(98%, 92%, 82%)	(98%, 100%, 98%)	(100%, 92%, 84%)

Table 4.5: *Frequency that indicates how many times, for each fixed vine structure, the fitted set of families of the bivariate copulas coincide with the original set (t-Student). Only family coincidence is considered, without regard to parameters.*

4.6.4 Gumbel's Copula

Lastly, we replicated the two experiments, setting the vine model family to a Gumbel copula at each level. For the first experiment, we set the values of the parameters equal to $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (10, 4, 8)$, corresponding to $(\tau_{12}, \tau_{31}, \tau_{32|1}) = (0.9, 0.75, 0.88)$. For the second experiment, the values used are available in Table 4.6. For vine models composed of Gumbel's bivariate copulas, the conclusions drawn from the previous experiments do not hold, as we find that selecting the appropriate vine structure becomes crucial for parameter estimation.

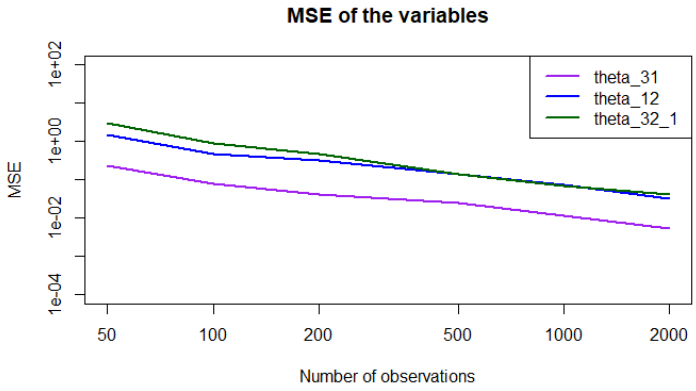


Figure 4.5: *Plot of the evolution of the convergence of the three variables, and the conditional parameter. On the x-axis we have the number of observations on a logarithmic scale, on the y-axis the corresponding Mean Squared Error values.*

Based on Figures 4.5 and B.16, we observe that all three variables exhibit a similar rate of convergence. However, Figure B.17 highlights how, for most sets of parameters, only the choice of the original vine structure Model V_1 guarantees that the boxplots have average values closely aligned to the true ones. This observation is further supported by Table 4.6, where the provided percentages indicate that the three models are no longer equivalent. The most evident cases are $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (15, 14, 16)$ and $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (10, 17, 8)$, where the Null hypothesis of the Vuong test is always rejected, indicating

that the models fitted using different structures are indeed significantly different.

$(\theta_{12}, \theta_{31}, \theta_{23\|1})$	$(\tau_{12}, \tau_{31}, \tau_{32\|1})$	Vuong V_3 vs V_1	Vuong V_2 vs V_1	Vuong V_3 vs V_2
(3, 5, 1)	(0.66, 0.8, 0)	13%	73%	63%
(10, 5, 3)	(0.9, 0.8, 0.66)	36%	0%	1%
(10, 17, 8)	(0.9,0 0.94, 0.87)	**0%**	**0%**	**0%**
(3,4,8)	(0.66, 0.75, 0.87)	1%	6%	1%
(15, 14,16)	(0.93, 0.92, 0.93)	**0%**	**0%**	**0%**

Table 4.6: *Gumbel: Percentage of p-values of the Vuong test are greater than 0.05, by comparing the results from fitting the different vine structures, compared two by two. The results are reported for all six combinations of parameters considered. We also provide the corresponding values of Kendall's Tau.*

Lastly, from Table 4.7, we can observe from the high percentages that for the bivariate copulas of the first tree (c_{12} and c_{31}), there is a general tendency to choose a Gumbel copula. While for the conditional copula $c_{23|1}$ of the second tree, Gumbel's family is almost never the best fit, particularly if we consider different structures from the original one. In fact, only for structure 213 (the original one) we can observe a high frequency of Gumbel bivariate copulas in the second tree. This consideration is coherent with the original observation that for Gumbel's vine models, the underlying structure influences the outcome. Moreover, the importance of choosing the correct structure is again confirmed by 4.8. Here, with the exception of the first two parameter combinations, the Vuong test suggests that the three structures produce significantly different results: case $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (15, 14, 16)$ is the most evident.

$(\theta_{12}, \theta_{31}, \theta_{32\|1})$	Structure V_3 $(c_{12}, c_{31}, c_{32\|1})$	Structure V_1 (True) $(c_{12}, c_{31}, c_{32\|1})$	Structure V_2 $(c_{12}, c_{31}, c_{32\|1})$
(3, 5, 1)	(80%, 76%, 0%)	(80%, 72%, 0.00%)	(72%, 76%, 0%)
(10, 5, 3)	(94%, 70%, 0%)	(94%, 86%, 80%)	(86%, 70%, 0%)
(10, 17, 8)	(92%, 0%, 0%)	(92%, 96%, 70%)	(96%, 0%, 0%)
(3, 4, 8)	(86%, 50%, 0%)	(86%, 88%, 86%)	(88%, 50%, 0%)
(15, 14, 16)	(98%, 0%, 0%)	(98%, 96%, 96%)	(96%, 0%, 0%)

Table 4.7: *Frequency that indicates how many times, for each fixed vine structure, the fitted set of families of the bivariate copulas coincide with the original set (Gumbel). Only family coincidence is considered, without regard to parameters.*

$(\theta_{12}, \theta_{31}, \theta_{32\|1})$	Vuong V_3 vs V_1	Vuong V_2 vs V_1	Vuong V_3 vs V_2
(3, 5, 1)	72%	92%	96%
(10, 5, 3)	70%	10%	28%
(10, 17, 8)	4%	8%	2%
(3, 4, 8)	8%	20%	32%
(15, 14, 16)	0%	0%	32%

Table 4.8: *Frequency of P-values of the Vuong test greater than 0.05, obtained by comparing the results from fitting the different vine structures, compared two by two. The results are reported for all six combinations of parameters that we analyzed.*

5

Time Series

Time series are defined as a collection of random variables $X_1, X_2, \ldots$ that are ordered in time [49]. They are used to develop mathematical models to describe sampled data, and are particularly relevant to financial applications. These variables enable us to understand patterns, trends, and dependencies in data, facilitating forecasting, risk assessment, and investment decision-making.

In this chapter, we will introduce the key definitions and tools necessary for analyzing financial time series. We will discuss the most common time series models used in finance, and how parameter estimation and model selection are conducted.

5.1 Time Series Models

Before jumping to the definitions, it is important to understand the concepts of *stationarity* and *White Noise*, or WN, which will be crucial to discuss the various time series models [49].

A desirable property for time series is *stationarity*, characterized by two distinct definitions: strict stationarity, and weak stationarity [49]. We will further discuss this property in Section 5.3.1.

Definition 18 (Stationarity)**.** *A time series is **strictly stationary** if the joint distribution of every collection of variables $\{X_{t_1}, \ldots, X_{t_k}\}$ is identical to that of the time shifted set $\{X_{t_1+h}, \ldots, X_{t_k+h}\}$, hence we have that $\forall k = 1, 2, \ldots$, $\forall t_1, ..., t_k$, $\forall x_1, ..., x_k \in \mathbb{R}$, and for all time shifts $h = 0, \pm 1, \pm 2, \ldots$:*

$$\mathbb{P}(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq x_k) = \mathbb{P}(X_{t_1+h} \leq x_1, \ldots, X_{t_k+h} \leq x_k).$$

*A **weakly stationary** time series, X_t, is a finite variance process that has:*

- *a constant mean value function $\mu_{Xt} = \mathbb{E}(X_t)$, independent of time t, and*
- *the autocovariance function $\gamma_X(t, s) = \mathbb{E}(X_s - \mu_s)(X_t - \mu_t)$, defined as the second-moment product for all time points t and s of the same series, that depends on s and t only through their difference $|s - t|$.*

Moreover, a White Noise is defined as follows.

Definition 19 (White Noise)**.** *A weakly stationary time series w_t (see Def 18), with mean $\mu_w = 0$, and autocovariance*

$$\gamma_w(h) = \begin{cases} \sigma_w^2 \text{ if } h = 0 \\ 0 \text{ else,} \end{cases}$$

is a ***White Noise*** *series $w_t \sim WN\left(0, \sigma_w^2\right)$. The autocovariance is defined as $\gamma_X(h) = cov(X_{t+h}, X_t)$.*

5.1.1 AR, MA, and ARMA

Autoregressive Models are models based on the intuition that the value of the time series X_t, at time t, can be explained as a function of the past values $X_{t-1}, \ldots, X_{t-p}$, and hence they are obtained by regressing on the past [49].

Definition 20 (Autoregressive Model (AR))**.** *An Autoregressive model of order p,* AR(p)*, is of the form*

$$X_t = \alpha + \sum_{i=1}^{p} \phi_i X_{t-i} + w_t, \tag{5.1}$$

where X_t is weakly stationary (see Def. 5.3.1), $w_t \sim WN\left(0, \sigma_w^2\right)$, and $\alpha, \phi_1, \phi_2 \ldots, \phi_p$ are constants ($\phi_p \neq 0$).

An example of Autoregressive Model is shown in Figure 5.1. The plot displays $N = 150$ points, obtained by simulating from an AR(1) of parameters $\phi_1 = 0.9$, $\sigma = 1$, and with initial observation $X_1 = 20$. Due to the high value of ϕ, the series is characterized by high memory, where each observation is strongly influenced by the preceding one. This high autocorrelation results in a plot that appears smooth, with persistence of values causing extended periods of gradual increases or decreases. A clear example of gradual decay can be noticeable in the first period, where the high initial value is slowly brought toward the mean 0.

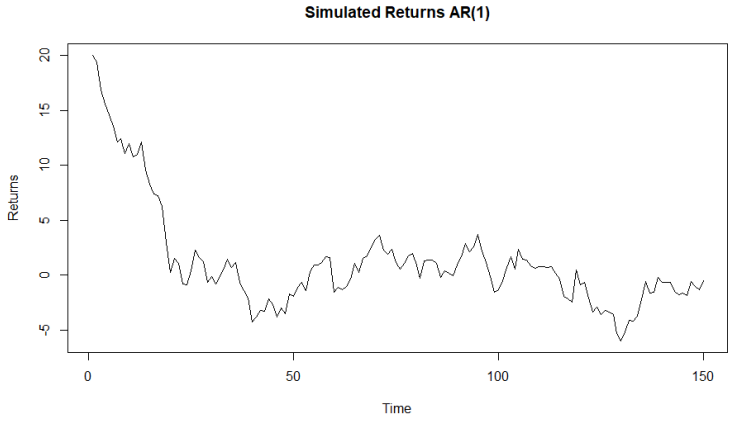


Figure 5.1: *Example of an AR(1) time series of 150 points, with parameters $\alpha = 0, \phi_1 = 0.9$. $w_t \sim N(0, 1)$.*

Moving Average models are derived as a function of the white noise, observed at different times [49]. They are defined as:

Definition 21 (Moving Average Model (MA))**.** *A moving average model of order q, or* **MA**(q) *model, is defined as*

$$X_t = w_t + \sum_{j=1}^{q} \theta_j w_{t-j}, \tag{5.2}$$

where $w_t \sim WN\left(0, \sigma_w^2\right)$*, and* $\theta_1, \theta_2, \ldots, \theta_q\,(\theta_q \neq 0)$ *are the model's parameters.*

Figure 5.2 shows $N = 150$ data simulated from a MA(1) time series, with parameters $\theta_1 = 0.9$, and $w_t \sim N(0, 1)$. The series is characterized by short term dependence, where a shock in the data affects the following observation, but not further ones. The series appears less smooth than the AR(1) plot, with higher variance. Additionally, MA time series are generally characterized by mean reversion. This means that deviations from the mean are usually followed by observations that contrast the effect, stabilizing the series around its mean.

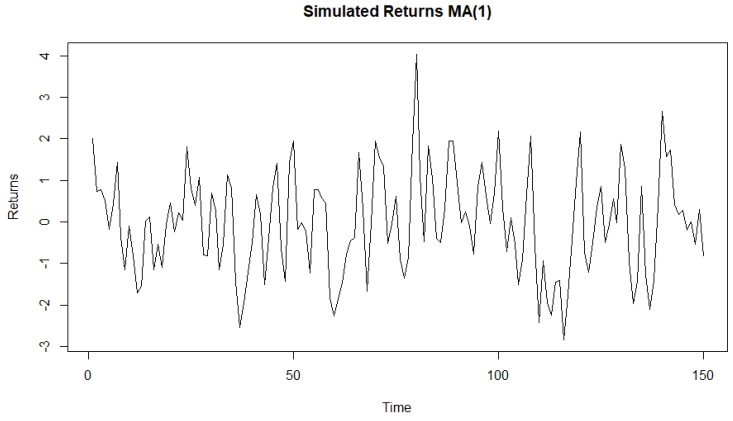


Figure 5.2: *150 sampled points from a MA(1) time series with parameters* $\theta_1 = 0.9$*, and* $w_t \sim N(0, 1)$*.*

Time series with constant volatility can also be described using ARMA(p, q) models, which are a combination of AR and MA models. Both time series imply a linear relationship with past observations of the stationary time series, and are characterized by the orders p, hence the number of autoregressive terms, and q, the number of lagged forecast errors [50]. The formal definition is the following:

Definition 22 (Autoregressive Moving Average Model (ARMA))**.** *A time series* $X_t, t = 0, \pm 1, \pm 2, \ldots$ *is ARMA(p, q) with mean* μ*, if it is stationary and*

$$X_t = \alpha + \sum_{i=1}^{p} \phi_i X_{t-i} + w_t + \sum_{j=1}^{q} \theta_i w_{t-j},$$

where $w_t \sim WN\left(0, \sigma_w^2\right)$*, and* $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$*. Moreover, we have that* $\sigma_w^2 > 0$*,* $\phi_p \neq 0$*,* $\theta_q \neq 0$*.*

In Figure 5.11 we show a simulation of an ARMA(1,1) time series with parameters $\phi_1 = 0.9, \theta_1 = 0.7$, and with $w_t \sim N(0, 1)$. The plot exhibit characteristics of both AR and MA processes. The strong influence of the past, due to the high value of the AR parameter results in a rather smooth plot. However, the series appears less predictable than a pure AR process, with isolated peaks where high volatility affects only the following observation, a characteristic associated with the MA component.

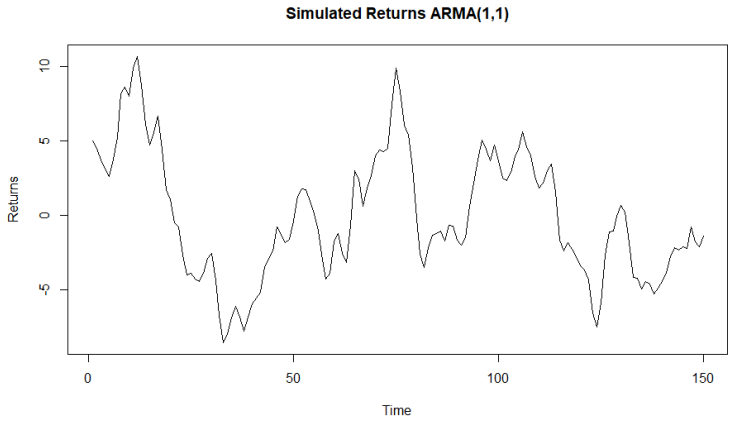


Figure 5.3: *150 sampled points from an ARMA(1,1) time series with parameters* $\phi_1 = 0.9, \theta_1 = 0.7$*, and with* $w_t \sim N(0,1)$*.*

5.1.2 ARCH, and GARCH

Financial assets, such as stock returns, often exhibit volatility clustering and non-linearity of the variance, features that are captured both by the Autoregressive Conditional Heteroscedasticity model (ARCH), and by the Generalized Autoregressive Conditional Heteroscedasticity, two of the mostly used time series models. As the name suggests, in fact, they are used to characterize heteroscedastic processes, hence processes that exhibit a non-homogeneous and time-varying variance. An ARCH process of order 1 is defined as:

$$X_t = \mu + \sigma_t \epsilon_t,$$

where:

$$\begin{cases} \epsilon_t = \sigma_t w_t, \\ \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2. \end{cases}$$

X_t is the value of the asset return at time t, μ the mean of the time series. $w_t \sim N(0,1)$ represents independently and identically distributed white noise. The conditional variance, σ_t^2, is determined by α_0 and α_1, which are parameters that capture the impact of squared past observations X_{t-1}^2, on the current conditional variance [49].

GARCH models, introduced by [51] in 1986, are widely used for option pricing and studying volatility. They serve as an extension of ARCH models, where the variance is not a linear function of past observations anymore, but it incorporates information gained by including lagged conditional variances. The definition of a GARCH of order p, q is taken from [52], and is the following:

Definition 23 (GARCH(p, q))**.** *A process* ϵ_t *is called a GARCH(p, q) process if its first two conditional moments exist and satisfy:*

1. $\mathbb{E}(\epsilon_t | \epsilon_u, u < t) = 0, \ t \in \mathbb{Z}$.
2. *There exist constants* α_0*,* α_i*,* $i = 1, \ldots, 1$*, and* β_j*,* $j = 1, \ldots, p$ *such that:*

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2.$$

An example of a simulated GARCH(1,1) time series is shown in Figure 5.4. In the plot, we can recognize some volatility clustering, a behavior typical of GARCH models. This is characterized by periods of high variance alternating with periods of lower volatility. Moreover, GARCH plots exhibit mean reversion, making the series appear rather random with no clear trends.

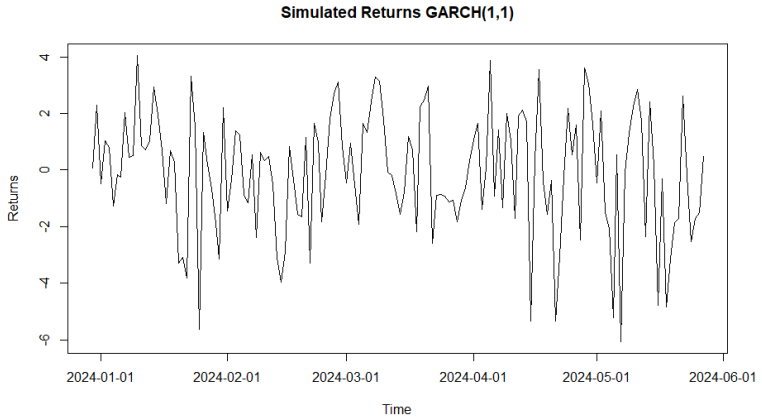


Figure 5.4: *Example of a GARCH(1,1) time series, generated sampling 150 points from a GARCH with* $\omega = 0.1, \alpha = 0.1, \beta = 0.9$.

5.1.3 ARMA-GARCH

ARMA-GARCH time seris are a combination of the previously discussed models, the ARMA component is used to model the mean of the time series, while the GARCH component is used to describe its variance [53]. In particular, we can consider these models as an extension of ARMA models, where the error ϵ_t is not a white noise, but follows a GARCH process [54].

Definition 24 (ARMA-GARCH)**.** *A time series* $X_t, t = 0, \pm 1, \pm 2, \dots$ *is ARMA(p, q)-GARCH(l,h) if it is stationary, and is characterized by the following structure:*

$$X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_i \epsilon_{t-j},$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{l} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{h} \beta_j \sigma_{t-j}^2,$$

$$\epsilon_t = \sigma_t w_t,$$

where α_i *and* β_j *are constant coefficients,* l *and* h *are orders, and* $w_t \sim WN\left(0, \sigma_w^2\right)$ *is a white noise sequence. Moreover, we have that* $\sigma_w^2 > 0$, $\phi_p \neq 0$, $\theta_q \neq 0$.

In Figure 5.5 we can find an example of simulation of an ARMA-GARCH. The plot exhibits a combination of ARMA and GARCH effects: the high ϕ value leads to long memory, resulting in a smoothed plot where future observations are influenced by past values. Additionally, the presence of volatility clustering can be observed, which is attributed to the high value of the parameter β_1 in the GARCH component.

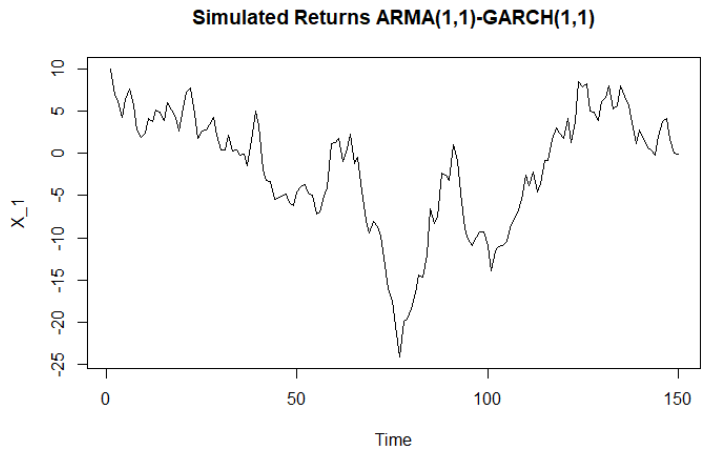


Figure 5.5: *Example of an ARMA(1,1)-GARCH(1,1) time series, generated sampling 150 points from a time series with parameters* $\phi_1 = 0.9, \theta_1 = 0.3$*,* $\alpha_0 = 1, \alpha_1 = 0.3, \beta_1 = 0.5$*.*

5.1.4 Other Models

Several other time series models can be explored, including HAR models, ARFIMA models, and EGARCH, among many others. For more comprehensive details on these models, readers can refer to the book by Shumway and Stoffer titled "Time Series Analysis and its Applications" [49].

Additionally, a natural evolution involves considering the relationship between different time series: multivariate time series analyze the interdependencies among data of a multivariate dataset. In financial data analysis, several multivariate time series models are commonly used to capture the complex dynamics and interdependencies among different financial variables. Some of the most widely used models include: Vector Autoregression (VAR) and VARMA(p, q).

5.2 Box–Jenkins Methodology

We have now presented the main time series models, but the process of identifying an appropriate fit for a given dataset, is more complex and involves several steps. One widely recognized methodology for analyzing time series, and for developing forecasting models is the Box-Jenkins Methodology [52], which has a structure that resembles the ones previously discussed for copulas and vine models. Named after its creators, George Box and Gwilym Jenkins, this iterative approach is composed of six main steps:

1. A priori identification of the trends and patterns observed in the data, which may require transformations such as differencing ($Y_t = X_t - X_{t-k}$) or a log-transformation ($Y_t = \log(X_t)$), to stationarize the data (i.e., make its statistical properties constant over time);
2. A priori identification of the orders p and q or the autoregressive (AR) and moving average (MA) components of the ARMA model are determined based on the autocorrelation and partial autocorrelation functions;
3. Estimation of the parameters of the time series;
4. Validation of the selected parameters, and diagnostic;
5. Choice of a model: if the selected model does not provide satisfactory Goodness of Fit results, we iteratively refine the model, re-fitting the parameters, and eventually changing the used model;

6. Once a satisfactory model is identified, it is deployed to generate forecasts for future observations.

To illustrate each step, we provide a practical example involving a simulated Moving Average (MA) time series, introduced in Section 5.3.4. The example will help to the steps illustrate more in depth.

5.3 Preliminary Analysis

In time series analysis, the initial phase involves some preliminary analysis of the considered dataset, aimed at identifying its main characteristics, and to determine if any preprocessing is necessary to ensure reliable analysis. Generally, this analysis begins with data visualization, which provides an initial glimpse of the presence of any seasonality or patterns. A more in-depth diagnosis then requires analyzing the stationarity of the time series, examining the ACF and PACF plots (which will be discussed in section 8.4), and testing for heteroskedasticity.

5.3.1 Stationarity

A stationary time series, which was defined in Def. 18, is one that does not show any seasonality, pattern nor trend: this allows to have a simpler modeling process and enables the application of various statistical tools and techniques. Most collections of data are not originally stationary, but the property can sometimes be achieved by applying transformations, such as differencing or a logarithmic transformation [49].

Consequently, a first approach to verify whether a time series is stationary is to conduct a visual inspection of the time series, and its residuals. The presence of clear trends, cycles, or systematic patterns suggests non-stationarity. Alternatively, a more formal approach is hypothesis testing. The Dickey-Fuller Test is commonly employed for this purpose.

5.3.1.1 The Dickey-Fuller Test

To test stationarity, many hypothesis testing have been developed. One of the most common ones is the Dickey-Fuller (DF) Test. The test is a unit root hypothesis test, which is used to assess if a specific time series process has a stochastic trend or exhibits non-stationarity.

The intuition behind a unit root can be extracted by considering an autoregressive process X_t of order $p = 1$, as defined in Def. 20. If $\phi_1 = 1$, hence if the root ϕ_1 lays on the unit circle, we have that sharp changes at one point in time strongly influence the future, making the time series not stationary. Contrarily, if $|\phi_1| < 1$, this considerations does not hold anymore: in this case the root lays inside the unit circle, and we do not have a unit root. The time series in this case can be considered stationary.

This is exactly what the Dickey-Fuller (DF) test does, which is discussed by [49] for an AR(p) process. Testing for a unit root can be translated in the following hypothesis:

$$H_0 : |\phi| = 1 \text{ vs } H_1 : |\phi| > 1,$$

where $|\phi|$ is the norm of the vector of autoregressive roots, obtained by solving the corresponding characteristic equation $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ for z. If the unit roots lie outside the complex unit circle, then the time series exhibits stationarity.

The test statistic for the case of an AR(1) is $(\hat{\phi} - 1)$ appropriately normalized, where $\hat{\phi}$ is the least squared estimator, computed as:

$$\hat{\phi} = 1 + \frac{\frac{1}{n}\sum_{t=1}^{n} w_t X_{t-1}}{\frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2}.$$

The p-value is computed by comparing the value of the test statistic to some critical values. In particular, the p-values of the test can be extracted from [55]. For more detailed information on how the regression is performed, and how to compute the test statistic refer to [56].

A more general test, is the Augmented Dickey-Fuller test, or ADF, which has been developed to detect non-stationarity in ARIMA(p,d,q) models, by applying an autoregression [56]. The test's hypothesis are the following:

$$H_0 : \text{ The time series has a unit root,}$$

$$H_1 : \text{ The time series is stationary.}$$

This is tested using the following test statistic:

$$t_n = \frac{\hat{\phi}_1 - 1}{\hat{Std}(\hat{\phi}_1)},$$

where $\hat{\phi}_1$ represents the ordinary least squares (OLS) estimator of the first autoregressive term, while $\hat{Std}(\hat{\phi}_1)$ represents the estimator of the standard deviation of $\hat{\phi}_1$. The distribution of the test statistic is non-standard, and can be found in [57]. Again, for a p-value smaller than 0.05 we are able to conclude that the analyzed time series is stationary.

Once we have established stationarity, we can explore various statistical measures to analyze and understand the dependencies within the time series. One such measure is the autocorrelation function, which is used to measure the linear dependence between two different time points of the same time series X_t and X_s. It is, therefore, a way of expressing numerically how much lack of independence there is between the points.

5.3.2 ACF and PACF

The Autocorrelation Function, or ACF, and the Partial Autocorrelation Function, or PACF, are two of the main statistical tools used in time series analysis to understand the autocorrelation and partial autocorrelation patterns in a given sequence of observations. Their importance lies in identifying stationarity, determining the order of a specific model, and gaining insight for forecasting. The subsequent sections will provide further details on these concepts.

Definition 25 (Autocorrelation Function)**.** *The autocorrelation function, or ACF, is:*

$$ACF(s,t) = \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}},$$

where $\gamma_X(s,t) = cov(X_s, X_h)$.

The ACF, which has values $ACF(s,t) \in [-1,1]$, is a measure of the linear predictability of the series

at a certain time t, having the value at time s. A stationary time series will generally have a ACF close to zero, or an ACF that decreases exponentially with the lag, which is the distance between the time points t and s.

For some time series, it is also worth computing the Partial Autocorrelation function, or PACF, which is a measure of the correlation between two random variables removing the linear effect of a third variable. For a stationary time series process the definition is the following:

Definition 26 (Partial Autocorrelation Function)**.** *The PACF of a stationary process X_t is:*

$$PACF(h) = \begin{cases} corr(X_{t+1}, X_t), & h = 1 \\ corr(X_{t+h} - \hat{X}_{t+h}, X_t - \hat{X}_t), & h \geq 2, \end{cases}$$

where $\hat{X}_t$ is obtained as $\hat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \ + \beta_{h1} X_{t+h-1}$, with coefficients β_i chosen to minimize thee MSE (Def. 39) of $\hat{X}_t$ from the true value X_t. Additionally, $corr(X_{t+h}, X_t) = \frac{\gamma_X(h)}{\gamma_X(0)}$, denotes the autocorrelation of X_t at lag h.

Both ACF and PACF can prove to be useful to identify the orders p and q of an AR, a MA or an ARMA model. In particular, from the PACF model we can identify the order p of the AR model: if the PACF plot shows a significant spike at lag p and then trails off, it suggests that the autoregressive order is p. Similarly, from the ACF plot we can identify the order of the moving average component q.

5.3.3 Heteroskedasticity

Lastly, as previously discussed for GARCH models in Section 5.1.2, a time series can exhibit time-varying variance. Testing for heteroskedasticity becomes therefore crucial to assess whether an ARCH or a GARCH model could potentially provide a good fit for the considered time. One example of a hypothesis test that can be employed in this context is the McLeod-Li Test.

5.3.3.1 McLeod-Li Test

The McLeod-Li test is an hypothesis test introduced in 1983 in [58], to detect if a model is conditionally heteroskedastic.

The test is based on the Ljung-Box Test (F.4.2), with Null and Alternative hypothesis:

$$H_0 : \text{The residual are homoscedastic.}$$

$$H_1 : \text{The residual of the time series are heteroscedastic.}$$

The test statistic is given by:

$$Q = n(n+2) \sum_{h=1}^{H} \frac{\widehat{ACF}^2(h)(\epsilon^2)}{n-h},$$

where n is the sample size, $\widehat{ACF}^2(h)$ is the sampled autocorrelation function at lag h of the squared residuals ϵ^2. Q asymptically follows a χ^2 distribution, with $(H - p - q)$ degrees of freedom, where H represents the maximum lag, while p and q the autoregressive and the moving average order respectively [52].

5.3.4 A practical Example: MA

As already anticipated in the previous section, throughout this Chapter we will support the theory by illustrating time series with a practical example. We therefore start by simulating a MA(1) times series of $N = 500$ time points, length similar to the one of the real-data application of Chapter 8. The time series is simulated using parameter $\theta_1 = 0.9$, while the white noise follows a normal distribution ($w_t \sim N(0, 1.5)$). The results are shown in the following Figure 5.6 below, where as discussed for Figure 5.2, we can notice a rather volatile time series, characterized by sharp changes and mean reversion.

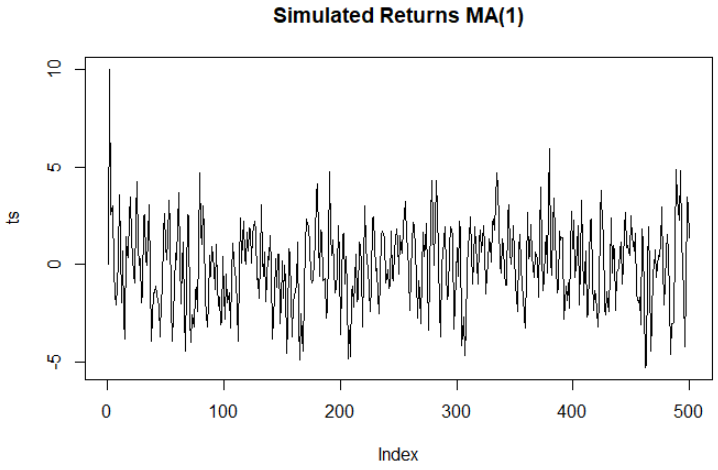


Figure 5.6: *Example of a MA(1) time series, generated sampling 500 points from a MA(1) of parameter $\theta_1 = 0.9$, and White Noise $w_t \sim N(0, 1.5)$.*

The goal of this simulation is to explore different aspects of time series analysis, including model diagnostic, model fitting and selection, residual analysis, and predictions. With these analysis on simulated data, we can gain insight about the behavior of time series models under different conditions and better understand their strengths and limitations. Additionally, the simulation study is useful to validate the methodologies for analyzing real-world time series data that will be deployed in Chapter 8. In Section 5.8, we will extend this simulation to other time series models.

5.3.4.1 Preliminary Analysis

Assuming that we do not know the original distribution used to simulate the MA(2) time series, we can start to analyse the plot of Figure 5.6, and to make some preliminary analysis following the steps discussed in this Section.

From a a first glance, the time series plot does not show any pattern nor seasonality, which might indicate stationarity; moreover, there isn't any clear sign of volatility clustering, and the overall plot seams to have a generally constant volatility, suggesting homoskedasticity. Proceeding with a more in-depth analysis, Figure 5.7 shows the ACF and PACF of the time series. In particular, the ACF has one relevant spike, suggesting that the order of the moving average (MA) of the time series should be $q = 1$. On the other hand, the PACF might indicate an AR component of order 1, or 2.

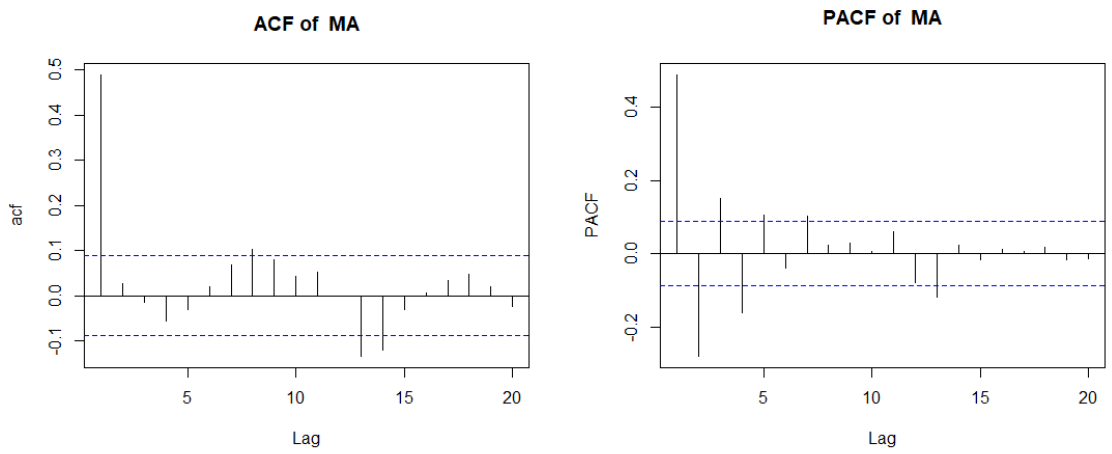


Figure 5.7: *ACF, PACF of a MA(1) of parameter* $\theta_1 = 0.9$, $w_t \sim N(0, 1.5)$.

Additionally, the Augmented Dickey-Full test for stationarity (see 5.3.1), resulted in a p-value of 0.01. Since the p-value is lower than 0.05 we can reject H_0, and conclude that the time series is indeed stationary, coherently with the observations made previously. It is also important to remark that MA time series are always stationary, since they lack a direct influence of the past observations (which is indeed the AR term). Lastly, we tested the heteroskedasticity by fitting and ARMA model on the time series, and by performing a McLeod-Li test on the residuals. The p-value of 0.13 indicates homoschedasticity.

In conclusion, based on the results obtained in this preliminary analysis, we may consider fitting an AR(2), a MA(1), or even an ARMA(2,1). Now that we have identified a set of potential time series models suitable for our dataset, two questions arise: how do we estimate the parameters of a time series model, and how do we select the best-fitting model? Answers to both questions will be provided in the following section.

5.4 Model Fitting

In this section, we will discuss how to estimate the parameters of a time series, and how to compare different models in order to choose the best fit. We will end by resuming the MA(1) example of Section 5.3.4.

5.4.1 Parameter Estimation

When estimating the parameters of simple time series models, two methodologies are generally preferred: the method of moments, and the maximum likelihood method [49]. The idea behind the method of moments is to find the estimators by matching the theoretical moments to the empirical ones. An example of a moment would be the expected value, corresponding to $k = 1$, where its empirical correspondent is the mean. On the other hand, the *Maximum likelihood estimate* (MLE) of a specific parameter θ, denoted as $\hat{\theta}^{ML}$, is obtained as the value that maximizes the Likelihood function $L(\theta)$, introduced in Definition 40. It is often derived by setting the partial derivatives of the Log-Likelihood function to zero. Further discussion on this approach can be found in Section F.1.1 of the Appendix.

For more complex models, such as GARCH models, more advanced methods are required. An example of method used would be the Quasi-Maximum Likelihood Estimator (QMLE), introduced in Chapter 7 of [52].

5.4.2 Model Selection

In the context of time series analysis, model selection is generally performed by comparing the Akaike's Information Criterion (AIC) or the Bayesian Information Criterion (BIC), and selecting the model with the lowest value, procedure that has been already discussed for copulas and vine models. BIC values are often preferred due to their penalty for model complexity, thereby mitigating the risk of overfitting. However, the results frequently agree with those obtained from comparing the AIC. Therefore, for simplicity, we will use AIC, as it is often already implemented in the functions used for this analysis in R.

For more insight into AIC and BIC, the reader can consult Definition 41 in the Appendix, where we discuss both criteria.

5.4.3 The MA Example

Coming back to the example presented in Section 5.3.4, we proceed by fitting different time series families to the dataset, and compare the relative AIC values to see if the best one aligns with the one used to simulate the dataset. Specifically, we fitted MA, ARMA, and GARCH models with orders ranging from 1 to 3, and in Table 5.1 we present the AIC values obtained for the best fit for each family. The overall best model, hence the one with lower AIC, results to be a MA(1), which is aligned with the model used to simulate the original dataset.

	AIC	Parameters
True	1881	$\theta_1 = 0.9, \sigma = 1.5$
MA(1)	**1878**	$\hat{\theta}_1 = 0.817, \hat{\sigma} = 1.574$
ARMA(1,2)	1885	$(\hat{\phi}_1, \hat{\theta}_1, \hat{\theta}_2) = (0.252, 0.543, -0.239), \hat{\sigma} = 1.573$
GARCH(1,1)	2095	$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (3.198, 0.193, 0.001)$

Table 5.1: *AIC values obtained by fitting different time series models to the dataset simulated using a MA(1) model, and the corresponding fitted parameters. We highlighted the best values in bold.*

The chosen model has fitted parameters equal to $\hat{\theta}_1 = 0.817, \hat{\sigma} = 1.574$, which are close to the original values, resulting in an absolute error of 0.083 for θ_1, and 0.074 for the standard deviation of the white noise.

5.5 Model Evaluation

In the previous Sections we have discussed how to fit the parameters of a time series model, and how to select the model based on the AIC or BIC criteria. However, the selection of a model based on these criteria doesn't guarantee its adequacy. In this section, we explore how to evaluate a chosen model to determine its suitability for the original dataset. In particular, we will discuss analysis of the innovations and some Goodness of Fit tests that can be performed.

5.5.1 Evaluation of the Innovations

One possibility of evaluating if the chosen model produces a reasonable fit, would be to examine its residuals or innovations. These residuals represent the difference between the observed values of the time series and the corresponding one-step-ahead predictions made by the fitted model. Mathematically, they are computed as $e_t = X_t - \hat{X}_t^{t-1}$, and are generally used for ARMA models. Additionally, for standardization and easier interpretation, it is also common to compute standardized residuals:

$$e_t = \frac{X_t - \hat{X}_t^{t-1}}{\sqrt{\hat{P}_t^{t-1}}}, \tag{5.3}$$

where $\hat{P}_t^{t-1}$ is the estimated one-step-ahead error variance [49].

In a good fit, residuals should be approximately independent and identically distributed (iid), with mean zero and variance equal to one. This property can be checked by observing the QQ-plot in case of normality, or by conducting a Ljung-Box test. The Ljung-Box is, in fact, an hypothesis test that has as null hypothesis the independence of the data considered, while the alternative hypothesis suggests dependence. More details are provided in the Appendix, in Section F.4.2.

Additionally, we can perform the Augmented Dickey-Fuller test and the McLeod-Li test, discussed in Section 5.3, on the residuals. In a good fit, these tests should indicate stationarity and homoscedasticity. Furthermore, we can analyze the autocorrelation function (ACF) plots of the residuals, which should not exhibit any patterns or significant values [49].

Lastly, in the case of multivariate time series we could perform a Q-test, to check if the noise is a white noise (WN): more details on this hypothesis test are reported in Subsection F.5.1 of the Appendix.

5.5.2 The MA Example

From the fitted model of the MA(1) example that was derived in Section 5.4.3, we proceed to evaluate the fit by performing some Goodness of Fit test on the obtained residuals, following the steps outlined above.

In Figure 5.12 the plot of the ACF, and the QQ-plot of the residuals can be visualized. The first exhibits low values of ACF, indicating that the chosen model could be a good fit for the original datasets. The points of the QQ-plot align with the diagonal line, indicating that the distributions of the residuals is aligned with the normal one, and hence suggesting that the fitted model is good.

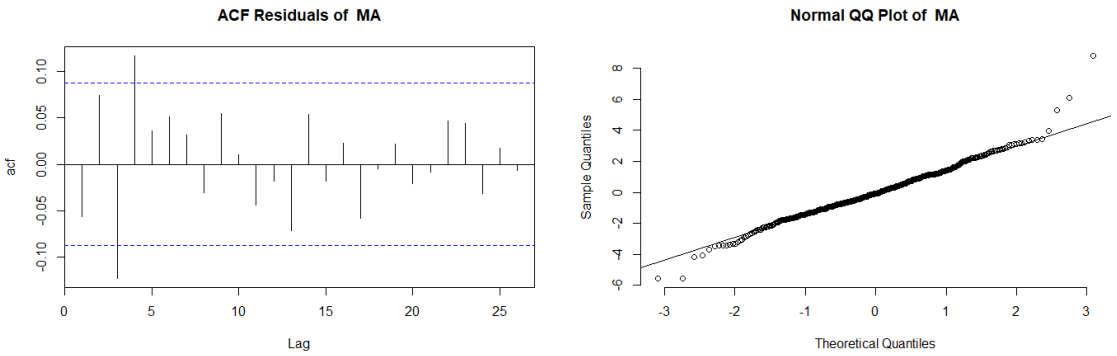


Figure 5.8: *ACF, PACF and QQ-plot of a MA(1) of parameter* $\theta_1 = 0.9$.

The Ljung-Box test yields a p-value of 0.205, indicating independence of the residuals. Additionally, the Augmented Dickey-Fuller (ADF) test returns a p-value of 0.01, suggesting stationarity. Finally, the McLeod-Li test produces a value of 0.332, indicating homoskedasticity. These results support the conclusion drawn from observing the autocorrelation function (ACF) and QQ-plot.

With the goodness of fit confirmed, we can now leverage our fitted model to mimic the behavior of the original dataset. This offers various applications, one of which would be making predictions. We will cover this particular application in the following Section.

5.6 Forecasting

When fitting a time series model to a sequence of observations $X_1, X_2, \ldots, X_t$, where t represents present time, and obtaining a set of estimated parameters $\hat{\Theta}$, we are able to leverage the information gained to predict future values of the time series $\hat{X}_t(m)$, $m = 1, 2, \ldots$. Here, t is defined as the forecast origin, and m as the lead time. The forecasted value is typically computed as the minimum mean square error, expressed as:

$$\hat{X}_t(m) = \mathbb{E}[X_{t+m}|X_1, \ldots, X_t],$$

where X_{t+m} represents the value of the times series at time $t + m$. Additionally to the predicted value, we should also provide a prediction interval that reflects the uncertainty of the forecast [59].

To evaluate the prediction we can use two different parameters. One choice is to compute the relative Root Mean Squared Error, or rRMSE, which is formulated as follows:

$$rRMSE = \sqrt{\frac{\frac{1}{n}\sum_{t=1}^{N}(X_t - \hat{X}_t)^2}{\sum_{t=1}^{N}\hat{X}_t^2}}, \tag{5.4}$$

where X_t is the value of the original dataset at time t, while $\hat{X}_i$ represents the corresponding predicted value. This is an alternative formulation of the MSE, which allows for better comparison with other methods by normalizing the data.

5.6.1 The MA Example

Since the goodness-of-fit test performed on the fitted MA(1) model presented in Section 5.3.4 indicated a satisfactory fit, we can proceed to make predictions. The model was trained using $N = 470$ data points

from the original time series, hence the training set. The remaining $N = 30$ data points are reserved as the test set, to evaluate the performance of the predictions. We make one day ahead predictions, adjusting the model each day. The result are shown in Figure 5.9, along with the prediction interval.

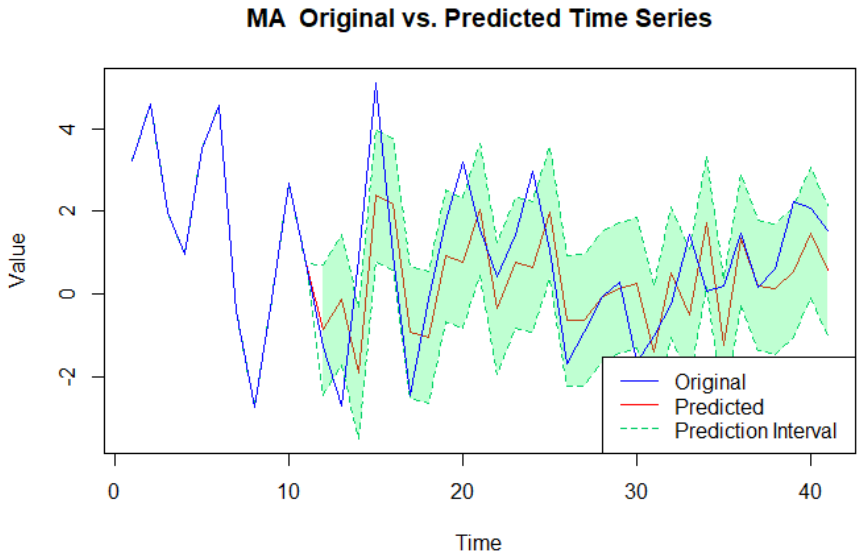


Figure 5.9: *1-day-ahead predictions of a MA(1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 95% *prediction interval at each time point, given the past.*

To evaluate the prediction, we computed the rRMSE (see (5.4), which resulted $rRMSE = 0.222$: the resulted forecasts are close to the original values. We can, in fact, also see from the plot how the prediction follow a similar trend to the one of the original dataset.

5.7 Time Windows

Lastly, time series analysis often uses different types of time windows to capture and analyze data trends, dividing the dataset into smaller partitions created in various ways. Here we illustrate the most common ones.

An expanding window approach starts with an initial subset of data and progressively includes more data points as time advances. This method ensures that all available observation at a specific time are used [60].

Alternatively, a rolling window approach can be employed. This method involves a fixed-size window that moves forward through the data, offering a consistent time frame for analysis at each step. This allows to account for structural changesand updating the model by forgetting past data, which could wrongly influence the present by holding outdated information [60].

Another approach is the sliding window. Similar to the rolling window, it can vary in size and overlap, offering flexibility in capturing short-term and long-term trends simultaneously.

5.8 Simulation Study

Following the methodology employed in the analysis of the Moving Average model discussed in the preceding sections, we extend our simulation study to include an AR(1) and a AR(1)-GARCH(1,1) model. This aims to provide a more in depth understanding of working with time series data, offering insights into various modeling techniques. By incorporating these additional models, we aim to validate the methodologies for analyzing real-world data, which will be applied in Chapter 8. Additionally, to support and test the methodology used in this simulation study, in Appendix C we provide a deeper analysis, focusing on each step separately.

5.8.1 AR(1)

We simulate $N = 500$ points from an AR(1) characterized by parameter $\phi_1 = 0.9$, and white noise $w_t \sim N(0, 1.5)$. The resulting time series is shown in Figure 5.10, where no trend, seasonality, nor volatility clustering is identifiable. The choice of such a high parameter value allow for long memory of the time series. Consequently, if something unusual happens, its effect propagates over several subsequent days. This allows for easier identification of autoregressive (AR) effects. This is, for example, visible at the beginning of the plot in Figure 5.10, where the time series, starting from a high initial value ($X_1 = 10$), takes some time to adjust around its mean, which should be zero.

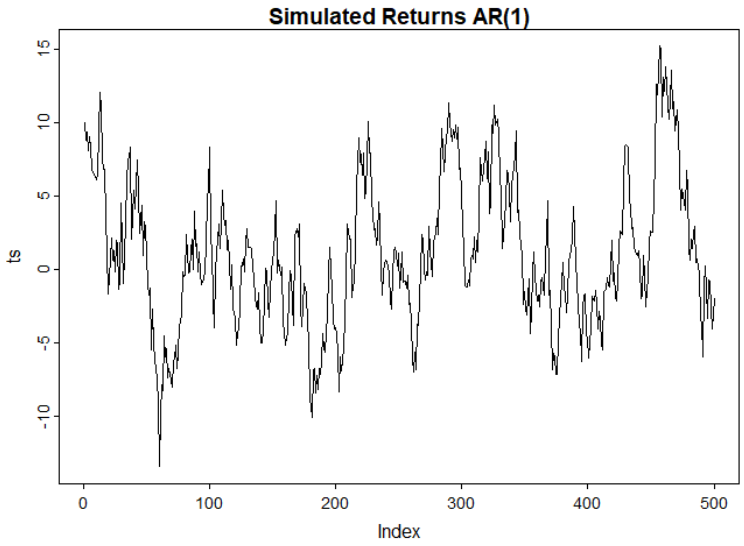


Figure 5.10: *Example of an AR(1) time series, generated sampling 500 points with parameter* $\phi_1 = 0.9$ *and* $w_t \sim N(0, 1.5)$.

5.8.1.1 Preliminary Analysis

Based on the ACF and PACF plots shown in Figure 5.11, we could assume that the time series could be modelled with an AR of order $p = 1$ (as indicated by the PAC). The ACF has significant values for many lags, which is still an indicator of stationarity, as it decays exponentially [49]. Moreover, this may suggest a Moving Average (MA) order equal to zero. However, it might be worth mentioning that there is a strong duality between AR and MA processes. They are, in fact, mathematically equivalent under inversion, property that is commonly known as "reversibility". Specifically, an $AR(p)$ process can be represented as an $MA(\infty)$ process, and vice versa [61].

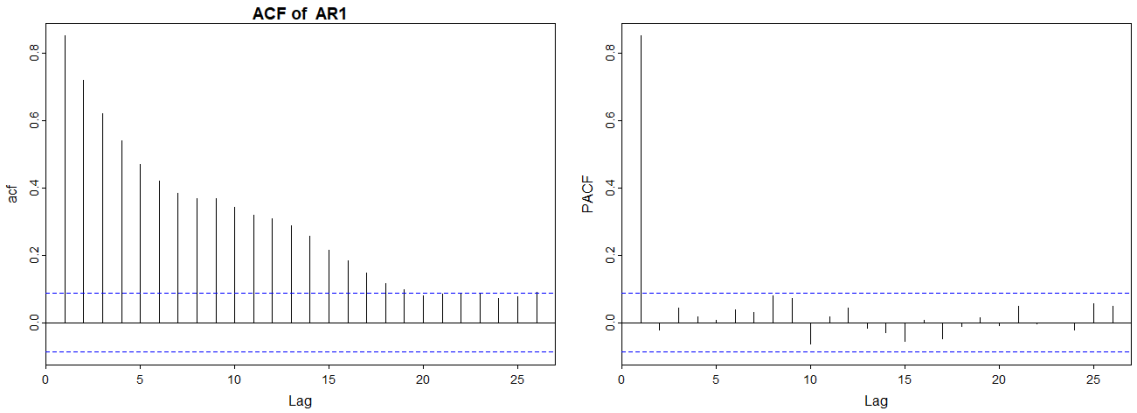


Figure 5.11: *ACF, PACF of an AR(1) time series, generated sampling 500 points with parameter* $\phi_1 = 0.9$ *and* $w_t \sim N(0, 1.5)$.

The low p-values obtained from the Augmented Dickey-Fuller test ($p - value = 0.01$), suggests that the time series should be stationary. This is coherent with our expectations. As discussed in Subsection 5.3.1.1, we can analytically determine whether a time series is stationary by finding the roots of the characteristic equation, which in this case is $\phi(z) = 1 - 0.9z = 0$. The corresponding root is $z = \frac{10}{9}$, which lies outside the unit circle, confirming that the time series is indeed stationary.

Moreover, we tested for heteroskedasticity of the time series by computing the p-value of the McLeod-Li test, which resulted equal to 0.112. Since we cannot reject H_0, the time series should be homoskedastic, which corresponds to expectations.

5.8.1.2 Model Fitting

As for the MA case, we start by fitting different time series with orders ranging from 1 to 3 to the dataset. Then, for each family considered we select the best fit by comparing the AIC values, and compare the results in Table 5.2. We then select the model with the lower AIC value, which is an AR(1), which corresponds to the original time series used to simulate the data.

	AIC	Parameters
True	1805	$\phi_1 = 0.9, \sigma = 1.5$
AR(1)	**1804**	$\hat{\phi}_1 = 0.87, \hat{\sigma} = 1.463$
MA(1)	2114	$\hat{\theta}_1 = 0.712, \hat{\sigma} = 1.994$
ARMA(1,1)	1806	$(\hat{\phi}_1, \hat{\theta}_1) = (0.859, 0.043), \hat{\sigma} = 1.461$
GARCH(1,1)	1831	$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (1.629 * 10^{-3}, 7.4 * 10^{-10}, 0.998)$
AR(1)-GARCH(1,1)	1832	$(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0.864, 2.12 * 10^{-3}, 5.88 * 10^{-10}, 0.998)$

Table 5.2: *AIC values obtained by fitting different time series models to the dataset simulated using an AR(1) model, and the corresponding fitted parameters. We highlighted the best values in bold.*

The parameters of the chosen fitted model are $\hat{\phi}_1 = 0.87, \hat{\sigma} = 1.463$. The fitted values are rather close to the original ones, with a maximum absolute difference between the original and fitted parameter of 0.03, which is a rather good result.

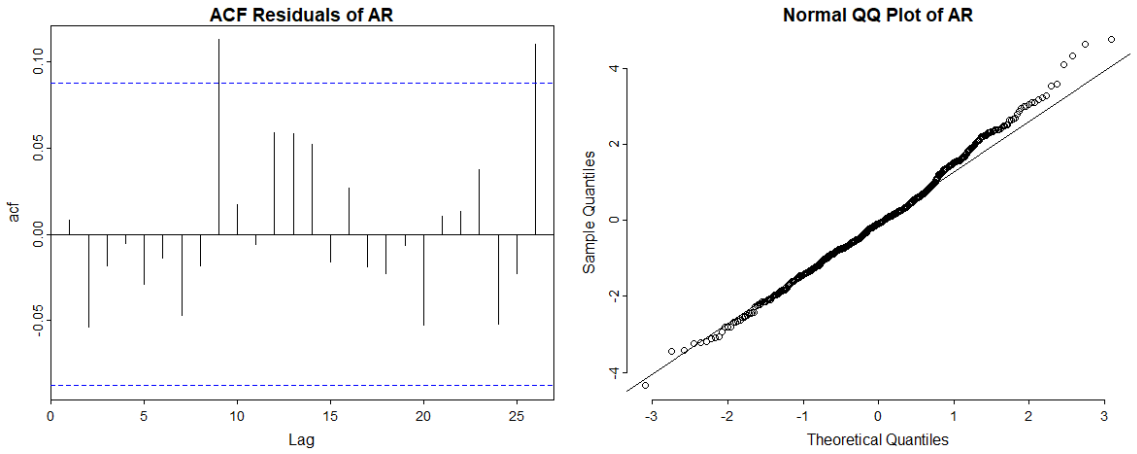


Figure 5.12: *ACF,and QQ -plot of the residuals of a fitted ARMA(2,2) time series.*

We then proceed to evaluate the goodness of the fit of the selected model. From Figure 5.12, both the extremely low values of ACF, and the aligned QQ-plot, suggest stationarity of the residuals. In addition, the high p-value (0.849) of the Ljung-Box test suggests that the residuals are independent over time, while the p-value of the McLeod-Li, equal to 0.130, indicates that they should not present any heteroskedasticity. Lastly, a p-value of 0.01 for the ADF test suggests stationarity of the residuals. All those indicators suggest that the chosen model might be a good fit for the dataset.

5.8.1.3 Predictions

In Figure 5.13, we present the predictions generated by fitting an AR(1) model to the initial $N = 470$ days, and then by using the fitted model to predict the last $N = 30$ days. The predictions are plotted against the test set. The prediction results in an $rRMSE = 0.111$ (see 5.4), which is a very good result: this might be due to the high parameter value of both the original dataset, and the fitted model, which allow to easily make 1 day ahead predictions, since they are highly influenced by the preceding time series value. We also computed the percentage of times the true time series lies between the upper and lower bound, which resulted equal to 93.3%, confirming that we have a good result.

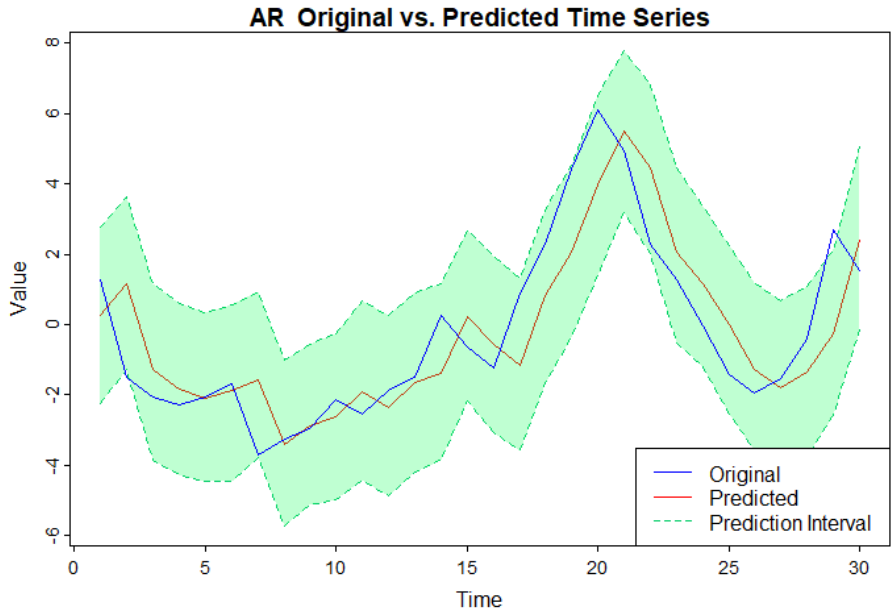


Figure 5.13: *1 day ahead prediction of a AR(1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

5.8.2 AR-GARCH

Lastly, we conclude this simulation study by repeating the same steps of the previous simulations for an AR(1) - GARCH(1,1) of parameters $(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0.9, 1, 0.3, 0.5)$, and noise $w_t \sim N(0, 1.5)$, which is shown in the following plot.

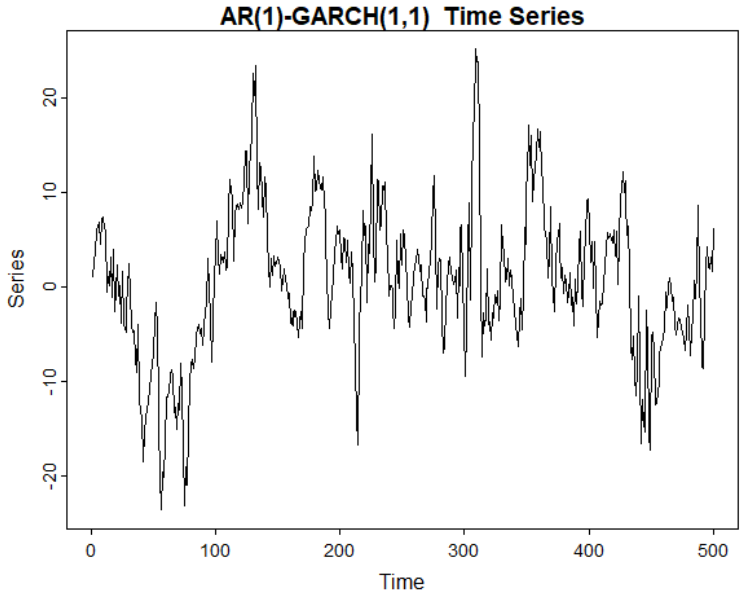


Figure 5.14: *Time series generated sampling 500 points from an AR(1) - GARCH(1,1) of parameters* $(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0.9, 1, 0.3, 0.5)$*, and noise* $w_t \sim N(0, 1.5)$.

5.8.2.1 Preliminary Analysis

As for the AR case, the autocorrelation function (ACF) and partial autocorrelation function (PACF) displayed in Figure 5.15, might indicate the presence of an AR(1) model. In fact, the PACF has only one significant lag, suggesting an order $p = 1$ of the AR component. On the other hand, the ACF has many significant lags that decay exponentially, suggesting no MA order. An alternative would be to consider different models, such as GARCH as a fit. These observations are supported the p-value of 0 of the McLeod-Li test, which suggests that the time series might be heteroskedastic. Moreover, in Figure 5.14 some volatility clustering are evident, a further indication of potential heteroskedasticity. The low p-value of the Dickey-Fuller test (equal to 0.01) suggests stationarity of the data considered.

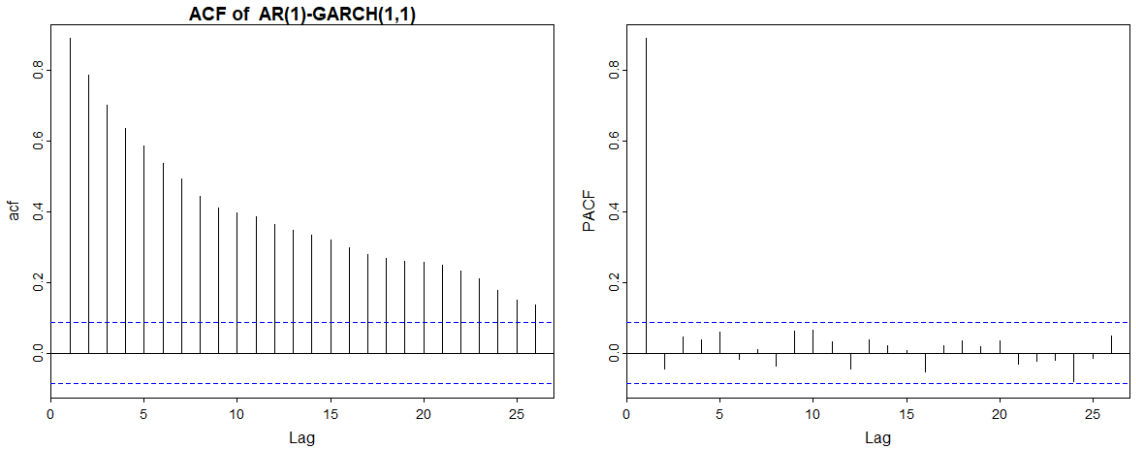


Figure 5.15: *ACF, PACF of a time series generated sampling 500 points from an AR(1) - GARCH(1,1) of parameters* $(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0.9, 1, 0.3, 0.5)$*, and noise* $w_t \sim N(0, 1.5)$.

5.8.2.2 Model Fitting

We proceed to fit various time series models with orders spanning from 1 to 3 to the dataset. Subsequently, for each model family, we identify the best fit and compare the outcome of the different models. Ultimately, we choose the model with the lowest AIC value, which happened to be an AR(1)-GARCH(1,1) model, coherently with the model used to simulate the original dataset: the results are provided in 5.3.

	AIC	Parameters
True	2586	$(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0.9, 1, 0.3, 0.5)$
AR(1)	2673	$\hat{\phi}_1 = 0.891, \hat{\sigma} = 3.486$
MA(1)	3049	$\hat{\theta}_1 = 0.767, \hat{\sigma} = 5.082$
ARMA(1,1)	2674	$(\hat{\phi}_1, \hat{\theta}_1) = (0.878, 0.059), \hat{\sigma} = 3.482$
GARCH(1,1)	2583	$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (1.846, 0.368, 0.5063)$
AR(1)-GARCH(1,1)	**2582**	$(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0.884, 1.870, 0.363, 0.508)$

Table 5.3: *AIC values obtained by fitting different time series models to the dataset simulated using an AR(1)-GARCH(1,1) model, and the corresponding fitted parameters. We highlighted the best values in bold.*

If we compare the original values of the parameters, with the ones obtained by estimating the parameters of the best fit, which are $(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0.884, 1.870, 0.363, 0.508)$, we can see how the values are rather close to the original ones, with the only exception of $\hat{\alpha}_0$, which has a absolute difference of

0.87 from the original parameter. This might be a good sign that the fitted model is a good fit for the original dataset.

Additionally, to evaluate if the selected model, we proceed to perform some Goodness-of-Fit analysis on the residuals. We start by analyzing the ACF and QQ-plot of the residuals, shown in Figure 5.16. The very low ACF values, and normality of data from the QQ-plot, are generally associated with a good fit. Moreover, we provide the p-values of the Ljung-Box test, of the ADF test, and of the McLeod-Li test. The first is equal to 0.849, suggesting that the residuals should be independent. The ADF results in a p-value of 0.01, indicating stationarity, and the McLeod-Li in a p-value of 0.130, which indicates homoskedasticity of the residuals. All indications that we have a good fit.

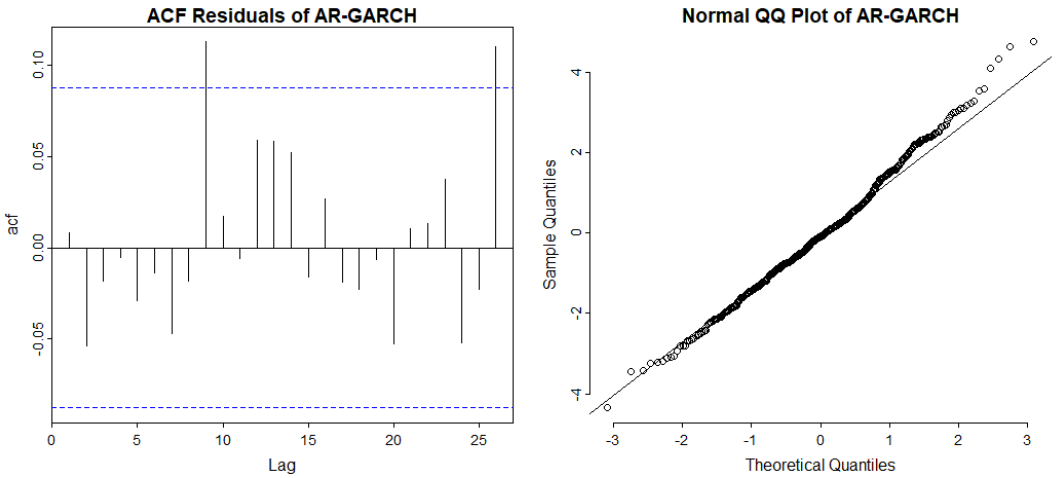


Figure 5.16: *ACF, and QQ-plot of the residuals of a fitted AR(1)-GARCH(1,1) time series.*

5.8.2.3 Predictions

In Figure 5.17, we compare the $N = 30$ predicted data points, using the fitted model analyzed in the previous Subsection to make 1 day ahead predictions. The predictions exhibit a rather reactive behaviour, that mimic the one of the original dataset. The $rRMSE = 0.145$, is a very good result. Moreover, 90% of the original dataset falls within the forecasting interval, which correponds exactly with the expectation.

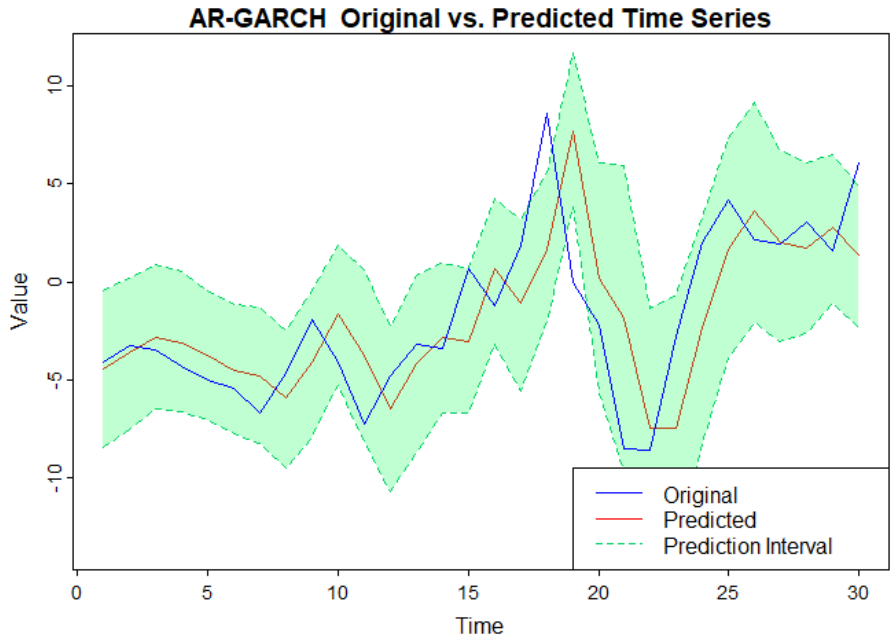


Figure 5.17: *1 day ahead days ahead prediction of an AR(1)-GARCH(1,1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

6

Copula-GARCH Models

In previous Chapters we have separately treated the concepts of copulas, vine copulas, and time series analysis. However, these models are frequently combined together to study the dependence structure of multivariate time series, which were briefly introduced in Section 5.1.4.

A widely used approach, particularly relevant in financial applications, is the one of copula-GARCH models. These models combine the strengths of copulas in studying dependence structures with the flexibility of GARCH models in capturing volatility dynamics, to accurately model datasets. They prove useful in examining cross-sectional dependence, which focuses on the relationship between different random variables at a specific time t. The main idea behind copula-GARCH models is filtering the margins with a time series, and then fitting an appropriate copula or vine model to the resulting innovations [62], [63].

6.1 Copula - GARCH Models

Copula-GARCH models are a class of statistical models used to analyze the dependence structure and volatility dynamics of a multivariate time series $(\mathbf{X}_t)_{t=1,\dots,T} = (X_{t,1}, \dots, X_{t,d})_{t=1,\dots,T}$. Their construction can be structured using three distinct steps [64].

The initial step involves applying a GARCH filter to each variable in the multivariate time series. By fitting a GARCH model (Def. 5.1.2) to each stationary variable, heteroskedasticity in the data is effectively removed. Consequently, for each variable $X_{t,i}$, $t = 1, \dots, T$, a set of independent and identically distributed innovations $\{e_{t,i}\}_{t=1,\dots,T}$ is obtained (refer to Section 5.5.1). Alternatively, we could consider filtering the time series with an ARMA-GARCH, a combination of the Autoregressive Moving Average and GARCH models, introduced in Section 5.1.3.

Subsequently, a probability distribution is fitted to the innovations of each variable, to model their marginal behavior. This step involves obtaining the cumulative distribution function (CDF) F_i for each residual, and transforming the residuals with the transformation $u_{t,i} = F_i(e_{t,i})$ for $t = 1, \dots, T$ and $i = 1, \dots, d$, to achieve a uniformly distributed multivariate time series. In case fitting the marginal distribu-

tion resulted too challenging, transforming data with the relative empirical CDF remains an option (see Def. 12). The result is a set of uniformly distributed time series $(\mathbf{U}_t)_{t=1,\dots,T} = (U_{t,1}, \dots, U_{t,d})_{t=1,\dots,T}$.

Lastly, the newly obtained time series is used to fit an appropriate copula model [65], or Vine model, following the steps discussed in Section 4.3.2, [66].

Examples of applications of copula-GARCH models can be found in [67] and [64]. A practical application of this methodology will be discussed in the simulation study of Section 6.7, or in the real data application presented in Chapter 8, where we will apply this methodology to real financial data.

6.1.1 Assumptions

When working with copula-GARCH models, we have to make sure that the following assumptions are verified at each step of the modelling process, to guarantee that the final model is able to produce a satisfactory fit for the original dataset.

- *Stationarity of Marginal Time Series:* The time series should be stationary (see Def. 18), meaning that their statistical properties do not change over time. In case of non-stationary, we can intervene by applying different transformations to make sure this property is satisfied.

- *IID, Stationary and Homoskedastic Innovations:* The residuals obtained from fitting time series models should be independent. Moreover they should be stationary, and homoskedastic. All these are a sign of goodness of fit, so if not satisfied, we might consider a different model.

- *Stable Copula Structure:* The copula structure used to model dependence should be stable over time, and hence should not be time dependant. As a consequence, the dependency parameter should be constant [65]. Alternatively, we will have to consider a more complex, time dependant model. We will discuss this extension in Section 6.6.

We will discuss how to test these assumptions in Section 6.4.1.

6.2 Other Models

In addition to copula-GARCH models, several extensions offer an alternative for modeling complex dependencies in multivariate time series data. A more advanced type of Vines, specifically Stationary Vines (or S-Vines), can be used to model simultaneously cross-sectional and temporal dependencies of a multivariate time series [66], [63]. Other possible extensions incorporate long-term memory such as the ones discussed in [68]. Others propose a model that accounts for non-stationarity, as proposed by [69]. Alternatively, [63] proposes a simplification of S-Vines, by assuming that the multivariate time series is a Markov chain, implying that most conditional pairs in higher trees will be independent. While these extensions offer interesting advantages, and opportunities to improve copula-GARCH methodology, they fall outside the scope of this thesis and will not be further explored.

6.3 Model Estimation and Selection

The process of selecting an appropriate copula-GARCH model is conducted sequentially. In this Section we will outline the main steps.

6.3.1 GARCH Selection

Initially, a GARCH model (or ARMA-GARCH) is fitted to each variable of the dataset individually. The parameters of the GARCH models are estimated using Maximum Likelihood Estimation (MLE) or Quasi-Maximum Likelihood Estimation (QMLE), both of which have been previously discussed in Section 5.4. Following the GARCH fitting, the various models considered are compared using criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion, or BIC, (refer to Def. 41 for additional information), to determine the best fit. Once the optimal GARCH model is identified, the corresponding innovations are computed.

6.3.2 Marginal Model

The next step is to determine a suitable marginal distribution for each vector of residuals to obtain the corresponding Cumulative Distribution Function (CDF) necessary for transforming the data to uniformly distributed variables. This involves parameter estimation using methods such as Maximum Likelihood Estimation (MLE). Again, the selection of the best-fitting marginal distribution is based on criteria such as log-likelihood, AIC, or BIC values. In alternative a semi-parametric procedure can be adopted, where the marginal selection is skipped, and the data are transformed by applying the correspondent empirical CDF [64].

6.3.3 Copula Selection on the Innovations

After projecting the data onto the unit hypercube, the final step involves fitting a copula model. The copula's parameters are estimated via MLE estimation (refer to Sect. F.1.1 of the Appendix), or with a Canonical MLE, as already discussed in Section 3.4.1. For more complex structures like vine models, a Stepwise MLE approach may be employed (see Sect. 4.5.1). Also the selection of the best copula models is carried out by comparing the AIC or BIC values of the different models taken into consideration.

6.4 Model Diagnostic

The diagnostic of the chosen model is conducted in different steps.

6.4.1 Assumption Testing

The initial step is to verify the hypotheses outlined in Section 6.1.1. Each verification can be performed using the following tests:

- *Stationarity of Marginal Time Series:* Stationarity can be assessed using statistical tests such as the Augmented Dickey-Fuller test, or ADF, which was introduced in Section 5.3.1.1, or by visual inspection of the time series plots.

- *IID, Stationary and Homoskedastic Innovations:* Independence can be evaluated using diagnostic tests such as the Ljung-Box test for autocorrelation in the residuals. Stationarity can again be tested using the ADF test, while heteroskedasticity tests, such as the McLeod-Li test (see 5.3.3.1) can be employed to check for constant variance in the residuals of the Copula-GARCH model.

- *Stable Copula Structure*: Stability of copula structure can be examined by partitioning the data into different time periods and comparing copula parameters or structures. Alternatively, we

could consider to test the time dependence with tests used to assess the simplifying assumption discussed in Section 3.6. In this case, the conditional variable, X_3, is time. We will further discuss this option in Section 6.6.

6.4.2 Goodness of Fit

Once the hypotheses are validated and suitable models are fitted, the goodness of fit of these models must be assessed. The procedures are the following:

- *GARCH Model:* The goodness of fit of the GARCH model can partly be associated with verifying the second hypothesis mentioned above, hence by studying the characteristics of the obtained residuals. A further test involves analyzing the Autocorrelation Function (ACF) plot of the innovations, which should not have any significant value, or decay exponentially, and conducting a Q-test to check if the noise is White Noise (WN), as discussed in Section 5.5.1.

- *Copula or Vine Model:* To evaluate the goodness of fit of the chosen copula or vine model, a Vuong test (discussed in Sect. 3.4.3.2) can be performed to determine if a selected distribution significantly outperforms alternative ones. Alternatively, a Goodness of Fit test based on the matrix equality of White (see Sect. 3.4.3) can be employed to compare the selected distribution with the distribution of the original dataset.

6.5 Predictions

Once the parameters of each variable $X_{t,i}$ of the copula-GARCH model have been estimated form historical data, and a model has been selected, future values of the time series can be predicted for a specified lead time. The lead time, denoted as m, represents the number of days-ahead at which the prediction is made: for the predictions considered in this thesis, we will always consider $m = 1$. The predicted values $\hat{X}_{t,i}(m)$ are computed from the forecast origin, t, which represents the current time point, using the estimated parameters of the model and appropriate forecasting techniques, such as Monte Carlo simulations (see F.6) [70], [71].

In particular, assuming we want to simulate the one-day-ahead value of a d-dimensional time series considered, $\hat{X}_{t,i}(1), i = 1, \dots, d$ we can use the following steps, outlined by [72].

1. Simulate Q realizations $U_{t,i}^{(j)}$ from the fitted copula model $\mathbf{U}^{(j)} \sim \mathcal{C}(\cdot), j = 1, \dots, q$ (or equivalently from the fitted vine model).
2. For each i, compute the corresponding innovation, by applying the inverse transformation $\hat{e}_{t,i}^{(j)}(1) = F_i^{[-1]}(U_{t,i}^{(j)})$.
3. From the obtained innovation $\hat{e}_{t,i}^{(j)}(1)$, use the fitted parameters of the GARCH, or ARMA-GARCH model, to compute the corresponding value $\hat{X}_{t,i}^{(j)}(1)$.
4. Compute the forecasted value as the average of the simulated values:

$$\hat{X}_{t,i}(1) = \frac{1}{Q} \sum_{j=1}^{Q} \hat{X}_{t,i}^{(j)}(1).$$

Additionally to the single point forecast, we are able to provide an interval, which is aimed at quantifying the uncertainty associated with the prediction. The prediction interval for each variable i of

level $1-\alpha \in (0,1)$ is estimated by taking the estimated quantiles of order $\frac{\alpha}{2}$, and $1-\frac{\alpha}{2}$ among the set of simulated values $\hat{X}_{t,i}^{(j)}(1), j = 1, \ldots, Q$. We will generally consider $Q = 1000$.

Lastly, to evaluate the quality and accuracy of the predictions we can use different metrics, such as the Root Mean Squared Error (RMSE), which was presented in equation (5.4).

6.6 Extension to Time Dependant Models

As discussed in Section 6.4.1, the stability assumption for copula-GARCH models can be relaxed by allowing for time-dependent models. Practically, this involves assessing the time dependence of the underlying structure and dividing the dataset into relevant partitions where the copula's structure remains stable.

6.6.1 Time Dependence Test and Partitions' Identification

To test the time dependence of the considered copula, we use an already implemented function from the R package CondCopulas, "*bCond.simpA.CKT*" [42]. This function assesses the simplifying assumption by splitting the dataset into partitions, or boxes. In our application, we substitute the conditional variables with an ordered vector $T = \{0, \ldots, 1\}$, representing time.

The function works by recursively dividing the conditioned observations. It employs a set of binary trees to identify relevant partitions by comparing different Kendall's tau values, and identifying those with the largest difference in Kendall's tau values among all pair copulas considered. Hence, at each level of the decision tree, by identifying the time t^* at which the difference $|\hat{\tau}_{i,j|X_J \in A \cap \{-\infty, t*\}} - \hat{\tau}_{i,j|X_J \in A \cap \{t*, \infty\}}|$ is maximal, for each pair of copulas i, j, where $\hat{\tau}$ is the estimated Kendall's tau (see 3.2.1), and A corresponds to a subset of the space of the conditioning variable X_J [73].

The function also computes a test statistic to evaluate the null hypothesis:

$$H_0 : \text{The Kendall's taus between the variables do not change over time,}$$

opposed to the alternative, which indicates relevant changes in the τ. For a p-value lower than 0.05, we reject the null hypothesis, indicating that the identified partitions reflect significant changes in Kendall's tau. For the bidimensional case, the test statistic is equal to:

$$\tau_n := \hat{\mathbf{W}}_{1,2}^T T^T (T \hat{\Delta} T^T)^{-1} T \hat{\mathbf{W}}_{1,2},$$

where $\hat{\mathbf{W}}_{1,2} := \sqrt{n}\left(\hat{\tau}_{1,2|X_J \in A_{1,J}} - \tau_{1,2|X_J \in A_{1,J}}, \ldots, \hat{\tau}_{1,2|X_J \in A_{m,J}} - \tau_{1,2|\mathrm{X}_J \in A_{m,J}}\right)^\top$, the random vector representing the differnce between the estimated τ and its true value for all subsets $A_{i,J}$ considered, while $\hat{\Delta}$ is the empirical variance-covariance matrix. T is a matrix that represents a subset $\mathcal{S}$ of q pairs of indexes $(k_i, l_i) \in \{1, \ldots, m\}^2$, with m being the sample size: each row i of T has all components equal to zero, exept the $k_i - th$ and the $l_i - th$ one. The test statistic is asymptotically distributed as a chi-squared with $m-1$ degrees of freedom [73]. For the multivariate version of the test statistic refer to the original paper of Derumigny, Fermanian, and Min.

Additionally, the function allows to have unbiased p-values by splitting the sample in two parts (generally using 50% of the sample is advised). One part is used to construct relevant partitions,

while the other is used to construct the test statistic. This approach ensures that the results are not contaminated, as it prevents the use of the same data for both partition construction and test statistic calculation.

6.7 Simulation Study

The aim of this simulation study is to illustrate the steps required to create a copula-GARCH model, capable of capturing the behavior of multivariate time series. The focus will be to highlight the advantages of introducing copulas when we want to understand, and model, dependencies in a time series model. In particular, we compare the predictions obtained by assuming that the two simulated variables are independent, and modelling each separately, with the predictions obtained from fitting a copula-GARCH. The study, will be later extended to include the case of a time-varying copula. In this simulation, the analysis is restricted to bivariate time series, but the results can easily be extended to higher dimensions, and the copula can be substituted with a Vine copula.

6.7.1 Case 1: Constant Copula

We simulate a customized bivariate time series of $N = 500$ observations. The choice of the sample size, as for the simulation studies of the previous Chapters, is aimed at mirroring the expected amount of data that will be used in the real-data application discussed in Chapter 8, as it has been studied previously that this number guarantee satisfactory results for parameter estimation.

Both variables are simulated from an AR(1)-GARCH(1,1) (see Def. 24) with the following structure:

$$\begin{cases} X_{t+1,1} = \phi_{1,1} X_{t,1} + \epsilon_{t+1,1} \\ X_{t+1,2} = \phi_{1,2} X_{t,2} + \epsilon_{t+1,2}, \end{cases} \tag{6.1}$$

where the autoregressive parameters are $(\phi_{1,1}, \phi_{1,2}) = (0.9, 0.8)$.

The volatility of each variable is obtained from the following equations, with parameters $(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}) = (1, 0.3, 0.5)$, and $(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}) = (1, 0.2, 0.6)$:

$$\begin{cases} \sigma^2_{t+1,1} = \alpha_{0,1} + \alpha_{1,1}\epsilon^2_{t,1} + \beta_{1,1}\sigma^2_{t,1} \\ \sigma^2_{t+1,2} = \alpha_{0,2} + \alpha_{1,2}\epsilon^2_{t,2} + \beta_{1,2}\sigma^2_{t,2}. \end{cases} \tag{6.2}$$

And the residuals are given by:

$$\begin{cases} \epsilon_{t+1,1} = \sigma_{t+1,1} w_{t+1,1} \\ \epsilon_{t+1,2} = \sigma_{t+1,2} w_{t+1,2}, \end{cases} \tag{6.3}$$

where the vector $(w_{t+1,1}, w_{t+1,2})$ is obtained by generating a bivariate copula $(u_{t,1}, u_{t,2}) \sim Clayton(5)$, and then by transforming the uniformly distributed variables with the transformation $\epsilon_{t,i} = F^{[-1]}(u_{t,i})$, $i = \{1, 2\}$, where $F(\cdot)$ represents the CDF of the marginals. We will consider as marginals two t-student with parameters $\nu_1 = 8$ and $\nu_2 = 4$, hence we will have $\epsilon_{t,i} \sim t(\nu_i)$. In Figure 6.1 we provide the plot of the simulated data.

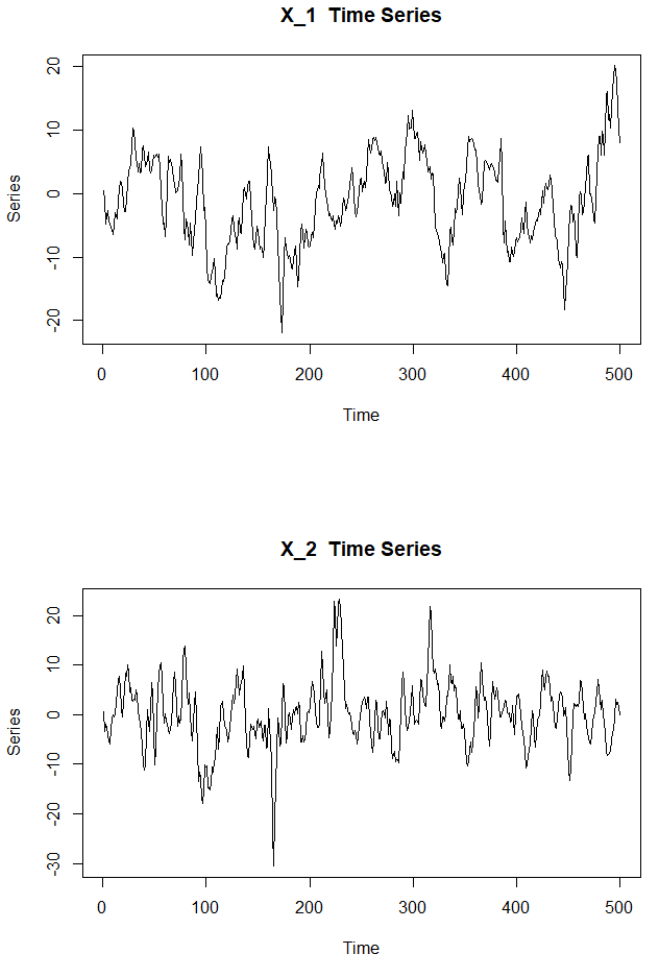


Figure 6.1: *Time series of $N = 500$ observations simulates using the the formulas provided in 6.1.*

The resulting residuals are shown in Figure 6.2, where we can recognise the typical feature of Clayton copulas, which are characterized by higher left-tail dependency.

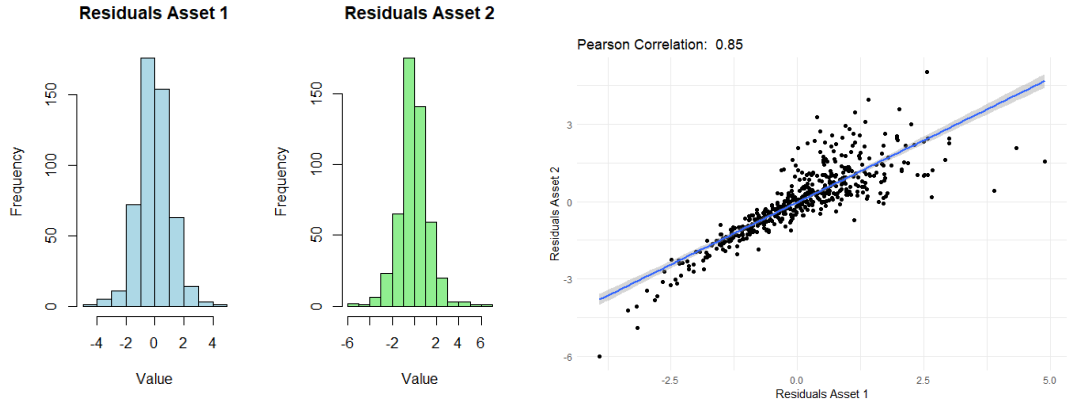


Figure 6.2: *Distribution and correlation of the residuals used to simulate the time series of Figure 6.1.*

We start by making some preliminary analysis on the simulated time series. Subsequently we will fit

a model, considering two separate cases. The first simply fits a time series to both variables, the second follows the step outlined in Section 6.1. Both methodologies will then be used to simulate one-day-ahead predictions for 30 days, which will be carefully evaluated and compared.

6.7.1.1 Preliminary Analysis

In Figure 6.3 we have the plots showcasing the ACF, and PACF of both time series, which have similar results. In fact, for both plots the ACF plot displays many significant values, which decay exponentially: this might indicate that there is not a moving average component for both time series, which reflects the original model. Moreover, the PACFs suggest that the order of the AR components should be equal to $p = 1$, since they both have one significant lag.

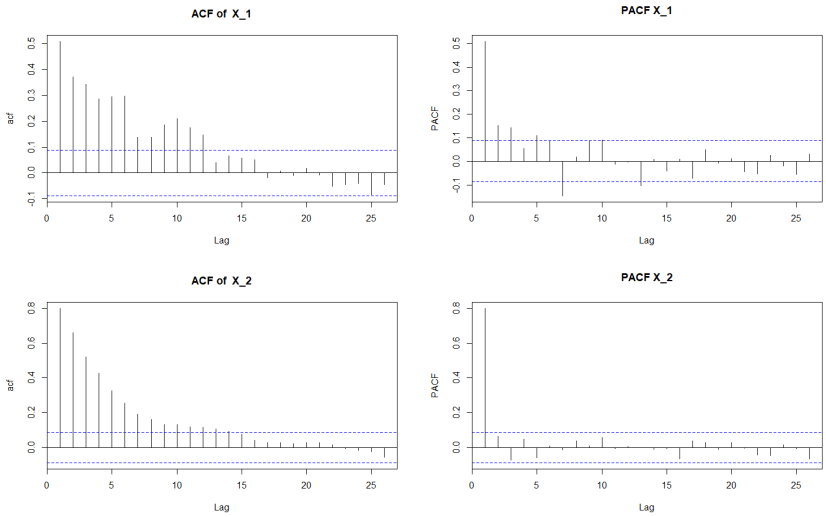


Figure 6.3: *ACF, and PACF of the two time series simulated form an AR(1)-GARCH(1,1).*

These observations are supported by Table 6.1, where for both variables the p-value of the McLeod-Li test suggests that the time series might be heteroskedastic. Moreover, the low p-value of the Dickey-Fuller test suggests stationarity of the data considered.

True Distr.	Dickey-Fuller	McLeod-Li
$X_{t,1}$	0.01	0
$X_{t,2}$	0.01	0

Table 6.1: *P-values of Augmented Dickey-Fuller to verify the stationarity of the GARCH time series, and of the McLeod-Li for heteroskedasticity.*

6.7.1.2 Marginal Fitting

We then proceed to fit an AR(1)-GARCH(1,1) model to the simulated time series to estimate its parameter (we assume the order of the time series known). The estimated parameters are presented in Table 6.2, where they are compared with the original parameters.

Variable	Original Param.	Estimated Param
$X_{t,1}$	$(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}) = (1, 0.3, 0.5)$	(1.131, 0.249, 0.565)
	$\phi_{1,1} = 0.9$	0.928
$X_{t,2}$	$(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}) = (1, 0.2, 0.6)$	(0.751, 0.236, 0.660)
	$\phi_{1,1} = 0.8$	0.792

Table 6.2: *Parameters obtained fitting an AR(1)-GARCH(1,1) model to the simulated bivariate time series.*

From the table we can observe how the estimated parameters appear rather close to the original values, with a maximum absolute difference of 0.249 for $\hat{\alpha}_{0,2}$, while the other estimated parameters appear much closer to the original ones. To assess the goodness of this fit, we proceed with some diagnostic of the fitted model on the obtained residuals.

We start by analyzing the ACF and QQ-plot of the residuals, shown in Figure 6.4. The very low ACF values, indicate a good fit, while the QQ-plot are not exactly aligned with the normal line, suggesting that the residuals might not follow a normal distribution: this is a good indication that we need a more complex model to study the distribution of the innovations.

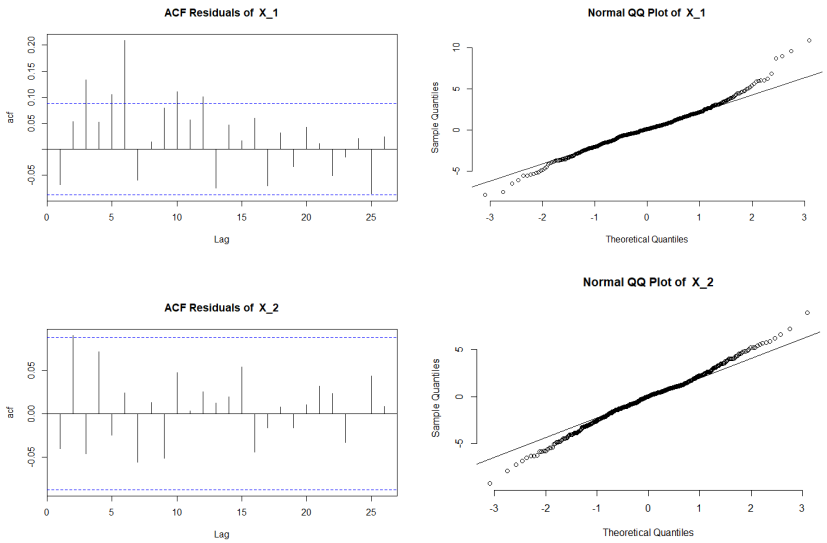


Figure 6.4: *ACF,and QQ-plot of the residuals of the two fitted AR(1)-GARCH(1,1) time series.*

Moreover, in Table 6.3, we provide the p-values of the Ljung-Box test, the ADF test, and the McLeod-Li test. The Ljung-Box test suggests that the residuals of both variables should be independent. The ADF result in a p-value of 0.01 for both variables, indicating stationarity. However, the McLeod-Li's p-values are smaller than the significance level, suggesting that the residuals could still exhibit heteroskedasticity. While the results of the Ljung-Box and ADF tests might indicate a good fit, the McLeod-Li test result suggests that the fit might not be perfect. This confirms the need for further refinement to achieve a better fit for the original dataset. With this observation in mind, we will proceed to fit a copula to the innovations in Section 6.7.1.4.

True Distr.	Ljung-Box	Dickey-Fuller	McLeod-Li
$X_{t,1}$	0.124	0.01	$2.99*10^{-13}$
$X_{t,2}$	0.361	0.01	$8.04*10^{-11}$

Table 6.3: *P-values of Augmented Dickey-Fuller to verify the stationarity of the residuals of the fitted time series, of the McLeod-Li for heteroskedasticity, and of the Ljung-Box test.*

6.7.1.3 Predictions Independent Variables

While the GoF suggested that the obtained fit might not be perfect, we still proceed to use the fitted model to simulate one-day-ahead predictions, leveraging the results obtained by fitting an AR(1)-GARCH(1,1) model, hence by assuming that the two variables are independent. The results are shown in Figure 6.5.

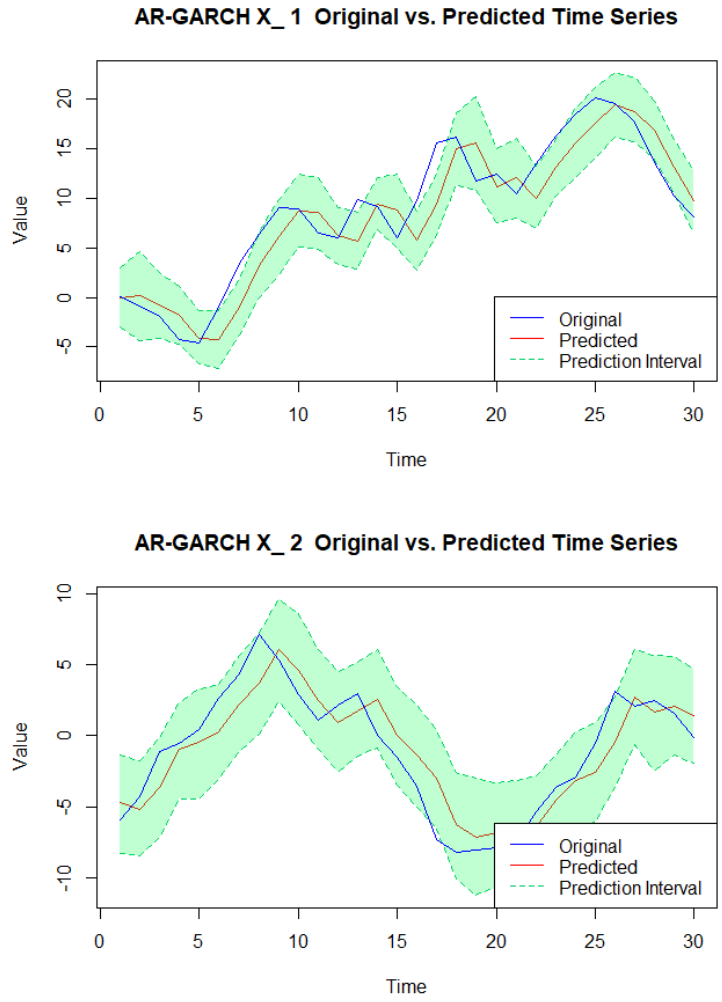


Figure 6.5: *1 day ahead days ahead prediction of an AR(1)-GARCH(1,1) time series, computed from the time series fit. The predictions are compared with the true realization of the original dataset. The green area represents the 90% prediction interval at each time point, given the past.*

The predictions seem to mimic the behaviour of the original variable quite well, especially for $X_{t,2}$. Moreover, in Table 6.4 we have the $rRMSE$ of the predictions, which are both rather low, and the percentage of observations within the prediction interval. The first variable falls within the prediction

interval only 70% of the times, which is lower than the expected 90%, hinting that some improvement can be done; the second variable however detains much better results.

True Distr.	rRMSE	Perc Inside
$X_{t,1}$	0.310	70%
$X_{t,2}$	0.195	93.3%

Table 6.4: *rRMSE of the forecasts, and percentage of times that the observed value lays within the prediction interval.*

6.7.1.4 Copula Fitting on Residuals

Starting from the residuals obtained in Section 6.7.1.2, instead of assuming independence of the innovations, we use the filtered data to fit a copula model, following the steps discussed in Section 6.1.

We start by fitting a marginal distribution to both set of residuals, and transform the residuals obtained above to a uniformly distributed dataset. For both variables, we obtained that the best marginal fit is a t-student with parameters $\hat{\nu}_1 = 9.135$, and $\hat{\nu}_2 = 5.268$.

To make sure the copula is not time dependant, we test the third assumption discussed in Section 6.1.1 on the transformed dataset, using the procedure presented in Section 6.6. We obtain a p-value of 0.89, that suggests the copula does not change with time.

We therefore proceed to fit a copula model assuming the family known. The result is a Clayton copula with estimated parameter $\hat{\theta} = 4.281$ (the original was $\theta = 5$). To evaluate the goodness of fit, we compute the p-value of the GoF test for bivariate copula models discussed in Section 6.4.2, which results to be equal to 0.18, indicating that parametric family of Clayton copula cannot be rejected. Hence, we should have a good fit.

6.7.1.5 Predictions

Lastly, following the steps discussed in Section 6.5, we leverage the information gained from fitting an AR(1)-GARCH(1,1), and a copula to the obtained innovations, to forecast 30 1-day ahead predictions. The results are shown in Figure 6.6.

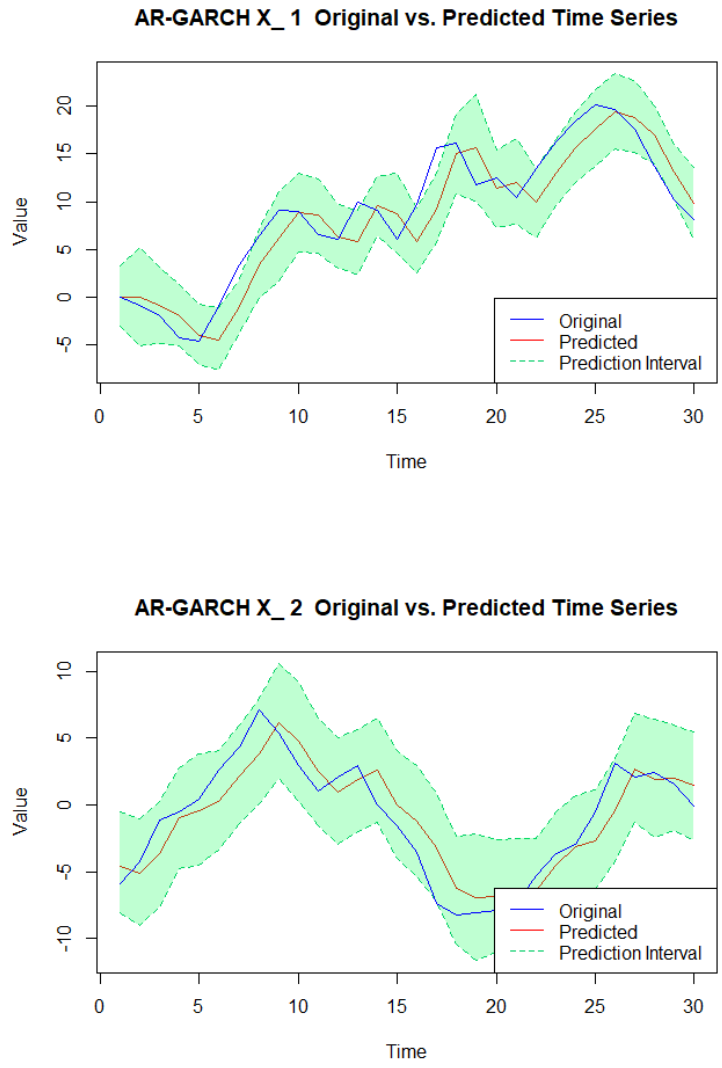


Figure 6.6: *1 day ahead days ahead prediction of an AR(1)-GARCH(1,1) time series, computed fitting a copula-GARCH model. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

In Table 6.5, we present the $rRMSE$ results from the true realization, along with the percentages of times the realization falls within the 90% prediction interval. The results are still satisfying, and comparable to the ones reported in Table 6.4, hence the ones obtained by only fitting a time series model. The true improvement from the previous case can be seen by comparing the perectages relative to the 90% prediction interval, which appears much better and colser to the expected one for the copula-GARCH case.

True Distr.	rRMSE	Perc Inside
$X_{t,1}$	0.249	86.6%
$X_{t,2}$	0.212	96.6%

Table 6.5: *rRMSE of the forecasts, and percentage of times that the observed value lays within the prediction interval.*

The predictions for both cases yield rather satisfactory results and exhibit similarities. Figure 6.7 shows that the predicted values are very similar to each other for both variables, with only small variations between the predictions. This similarity can be attributed to the high value of the autoregressive (AR) component ($\phi_{1,1} = 0.9$), which causes the predictions to be heavily influenced by past values. Additionally, similar results are expected as the advantage of copulas lies in modeling the dependence structure, rather than influencing the marginal predictions. This is evidenced by the improved interval percentages, and will be further examined in the following Section.

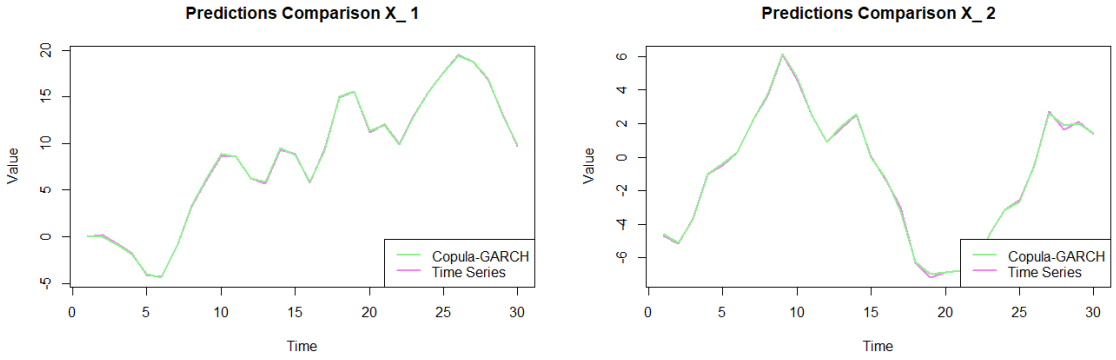


Figure 6.7: *Comparison between the 1 day ahead days ahead prediction of an AR(1)-GARCH(1,1) time series derived by fitting only a time series to the original dataset (blue), or with one that also fitted a copula to the filtered residuals, hence a copula-GARCH model (red).*

6.7.1.6 Additional Results

To further validate the advantages of applying copula-GARCH to forecasting, we decide to compare the distributions of the 95% percentile of the one-day-ahead predictions generated by the two methods, with the one of the points generated from the true distribution.

In particular, starting from the true realization of $\mathbf{X}_{470}$, we simulated $N = 1000$ points using the true distribution, using the parameters obtained from the $AR(1) - GARCH(1,1)$ fit, and lastly from the copula-GARCH model fitted above. We then selected the 95% of each set of simulated data. The results are compared in Figure 6.8.

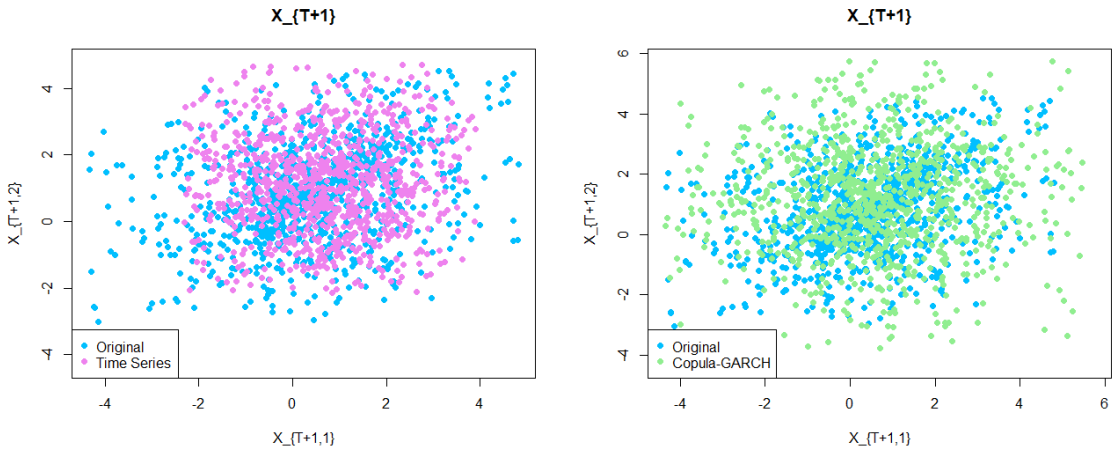


Figure 6.8: *95% percentile distribution clouds from one-day-ahead predictions using $AR(1) - GARCH(1,1)$ and copula-GARCH models, compared with points sampled from the true distribution.*

Both methods appear to have values that are aligned with the ones of the true model. However one advantage of copula-GARCH is quite evident: the fitted model is much better at imitating the behaviour of the original dataset when considering extreme events. In fact, the forecasts produced from the fitted time series are more concentrated toward the mean, while the copula-GARCH's predictions appear more spread out, which is aligned with the behaviour of the original dataset.

Moreover, if we compute the CVaRs of the simulated clouds of points, provided in Table 6.13, we can see how the result for the copula-GARCH appears to be much closer to the one obtained from the true model, highlighting the importance of correctly modelling the underlying copula. This particular advantage will be useful in the optimisation application that will be discussed in Chapter 8.

Method	$\text{CVaR}_{95\%}$
Original	1.983
AR-GARCH	1.093
Copula-AR-GARCH	1.772

Table 6.6: *Conditional Value at Risk obtained from the 95% percentile distribution clouds from one-day-ahead predictions using $AR(1) - GARCH(1,1)$, and a copula-GARCH model, compared with points sampled from the true distribution.*

For this particular case the advantages of fitting a copula are clear, but not obvious. However, in Section C.2 of the Appendix C we provide further examples, which consider different combinations of parameters, and different time series family, which further support the advantages of using copula-GARCH models.

6.7.2 Case 2: Time Dependant Copula

We repeat the simulation, by generating $N = 500$ points of a bivariate AR(1)-GARCH(1,1) time series, characterized by the same parameters as the one described in Eq. (6.2), with the only exception that the underlying copula used to simulate the residuals is now time varying.

The vector of residuals $(\epsilon_{t,1}, \epsilon_{t,2})$ is obtained by generating a bivariate copula $(u_{t,1}, u_{t,2}) \sim Gumbel(5)$,

for the first $N = 250$ observations, and subsequently a $t - Student(5)$ copula, for the remaining ones. The marginals will again be distributes as two t-student with parameters $\nu_1 = 8$ and $\nu_2 = 4$. In Figure 6.9 the plot of the simulated data is given.

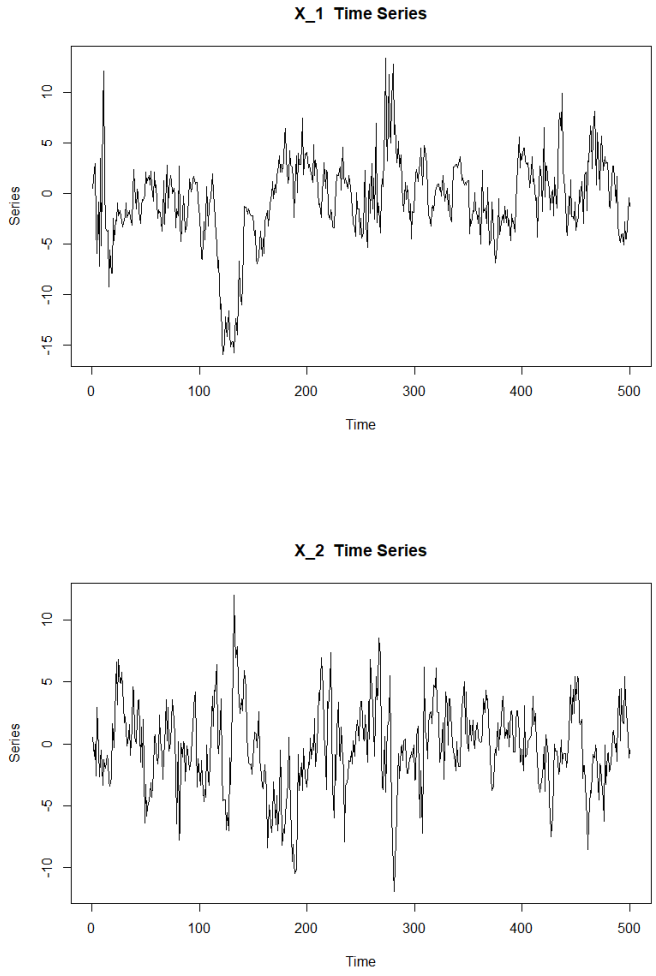


Figure 6.9: *Time series of $N = 500$ observations simulated from a AR(1)-GARCH(1,1) time series, with time varying residual generated from a bivariate copula.*

The resulting distribution of the residuals are shown in Figure 6.10.

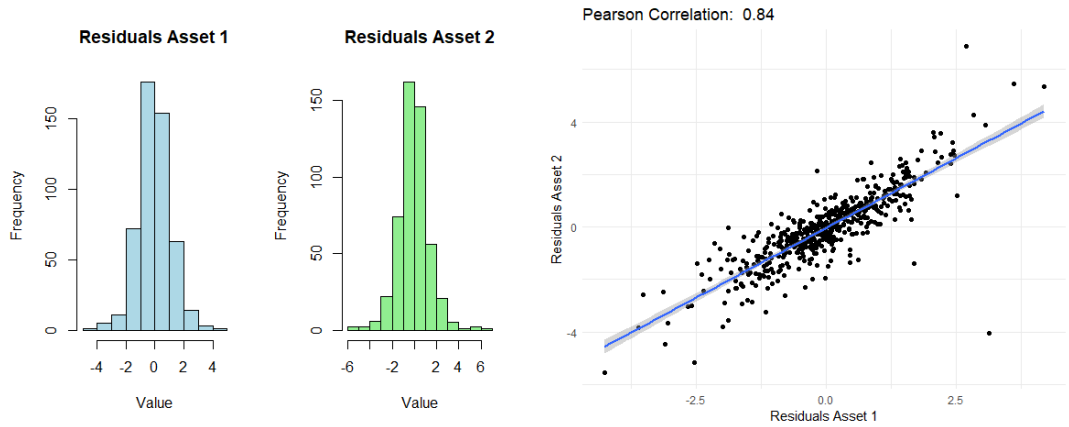


Figure 6.10: *Distribution and correlation of the residuals used to simulate the time series of Figure 6.9.*

6.7.2.1 Preliminary Analysis

In Figure 6.11 we find the plots showcasing the ACF, and PACF of both time series. For both time series, the ACF appear significant for many lags, and decays exponentially, while the both PACFs only have one significant lag: from this observation, we could conclude that the time series only have an AR component of order 1, hence an $AR(1)$.

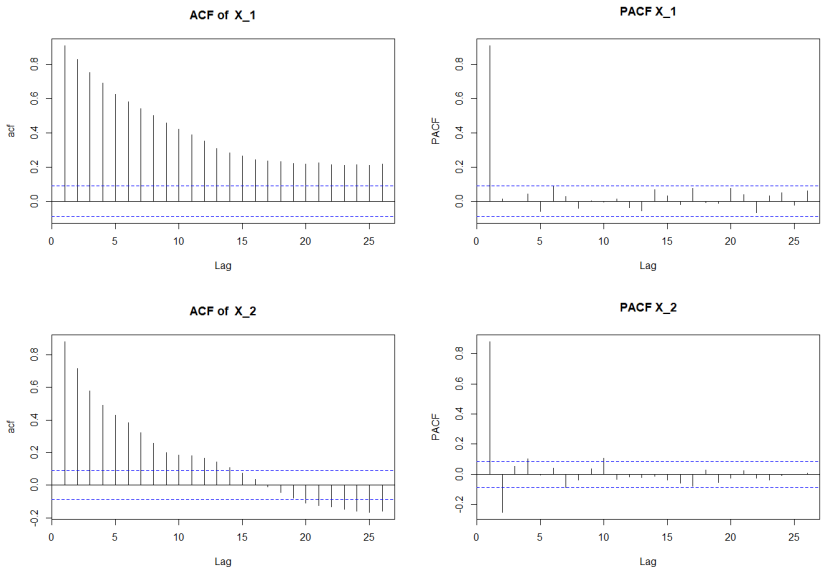


Figure 6.11: *ACF, and PACF of the two simulated time series form a AR(1)-GARCH(1,1).*

These observations are supported Table 6.7, where the tests suggest heteroskedasticity, and stationarity. We should therefore also consider a GARCH component when fitting a time series.

True Distr.	Dickey-Fuller	McLeod-Li
$X_{t,1}$	0.025	0
$X_{t,2}$	0.01	0

Table 6.7: *P-values of Augmented Dickey-Fuller to verify the stationarity of the AR-GARCH time series, and of the McLeod-Li for heteroskedasticity.*

6.7.2.2 Marginal Fitting

We then proceed to fit an AR(1)-GARCH(1,1) model to the simulated time series to estimate its parameter, assuming the order of the time series known. The estimated parameters are presented in Table 6.8.

Variable	Original Param.	Estimated Param
$X_{t,1}$	$(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}) = (1, 0.3, 0.5)$	(0.601, 0.297, 0.561)
	$\phi_{1,1} = 0.9$	0.911
$X_{t,2}$	$(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}) = (1, 0.2, 0.6)$	(0.835, 0.171, 0.683)
	$\phi_{1,1} = 0.8$	0.865

Table 6.8: *Parameters obtained fitting an AR(1)-GARCH(1,1) model to the simulated bivariate time series.*

From the table we can observe how, with the exception of the estimated values for parameters α_0, the estimations result rather accurate, in accordance with the results discussed for the time series simulation study. To assess the goodness of the fit, we therefore proceed with some diagnostic of the fitted model on the obtained residuals.

We start by analyzing the ACF and QQ-plot of the residuals, shown in Figure 6.12. The very low ACF values, indicate a good fit, while the QQ-plot suggests that the residual of $X_{t,2}$ might not follow a normal distribution: this is a good indication that we need a more complex model to study the distribution of the innovations. The residuals of the first variable, however appear rather alienated with the normal line, hinting that they might be normally distributed.

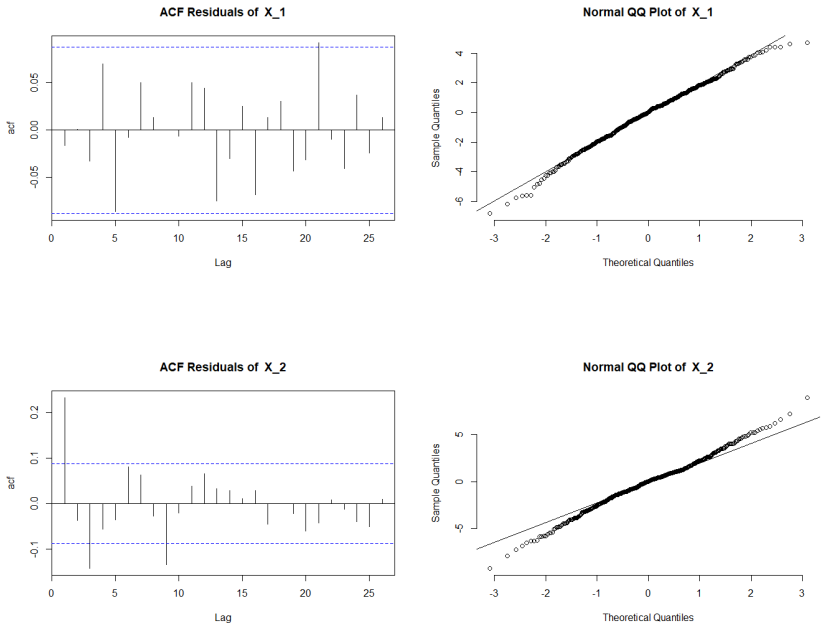


Figure 6.12: *ACF,and QQ-plot of the residuals of the two fitted AR(1)-GARCH(1,1) time series.*

Moreover, in Table 6.9 we provide the p-values of the Ljung-Box test, of the ADF test, and of the McLeod-Li test. The Ljung-Box, suggests that both set of residuals should be independent. The ADF results in a p-value of 0.01, indicating stationarity. However, both McLeod-Li p-value are lower that 0.05, indicating heteroskedasticity of the residuals: coherently with expectations, this might be

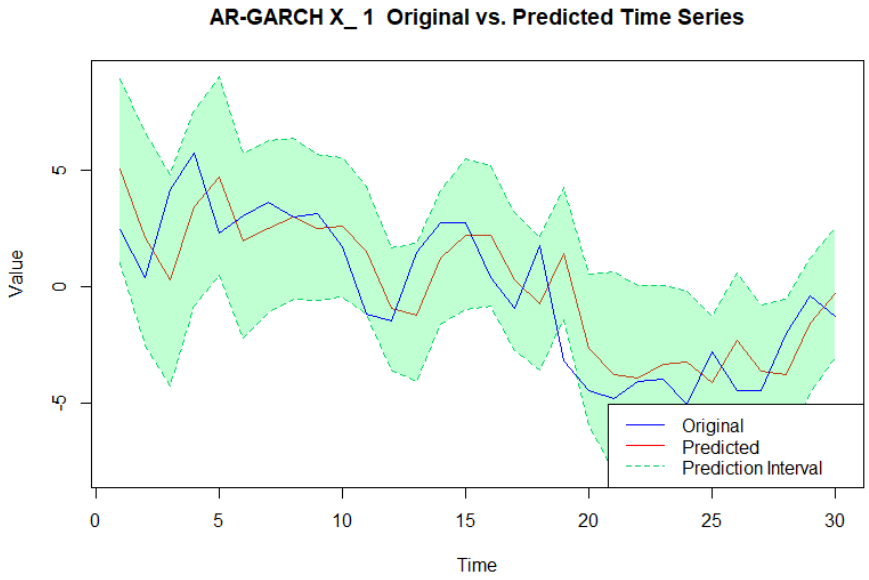
an indicator that we might need further modelling to achieve a good fit for our time series, able to incorporate all information.

True Distr.	Ljung-Box	Dickey-Fuller	McLeod-Li
$X_{t,1}$	0.716	0.01	$9.50*10^{-05}$
$X_{t,2}$	0.167	0.01	$3.95*10^{-07}$

Table 6.9: *P-values of Augmented Dickey-Fuller to verify the stationarity of the residuals of the fitted time series, of the McLeod-Li for heteroskedasticity, and of the Ljung-Box test.*

6.7.2.3 Predictions Independent Variable

Leveraging the results obtained by fitting a AR(1)-GARCH(1,1) model, hence by assuming that the two variables are independent, we proceed to use the fitted model to simulate one-day-ahead predictions. The results are shown in Figure 6.15.



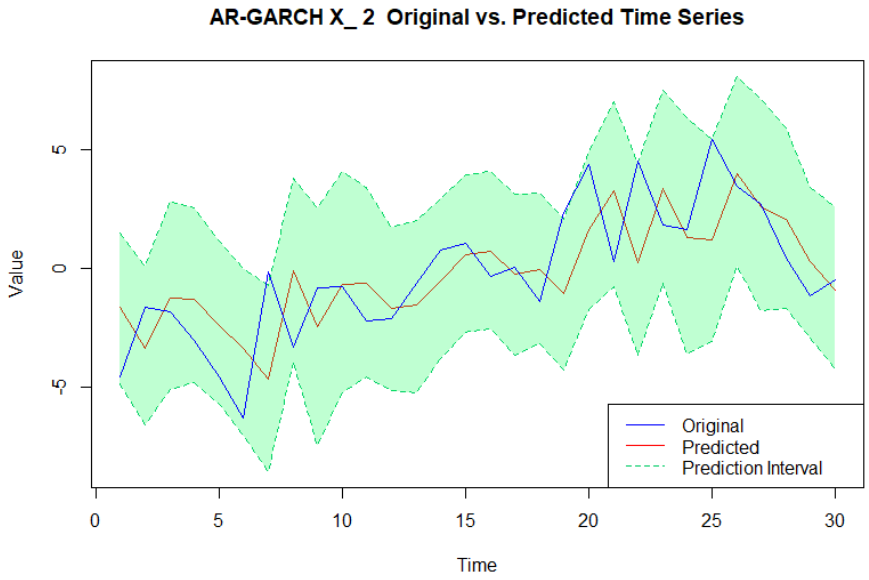


Figure 6.13: *30 days ahead prediction of the two AR(1)-GARCH(1,1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

The predictions seem to mimic the behaviour of the original variable quite well. Moreover, in Table 6.10 we provide the $rRMSE$ of the predictions, which are both rather low. The percentage of observations within the prediction interval, which is expected to be around 90% are also listed: the first variable has a satisfactory result, however, the interval for the second variable falls a little bit lower than its expected value.

True Distr.	rRMSE	Perc Inside
$X_{t,1}$	0.126	96.6%
$X_{t,2}$	0.195	83.3%

Table 6.10: *rRMSE of the forecasts, and percentage of times that the observed value lays within the prediction interval.*

The results obtained by only fitting the time series, and assuming the residuals independent resulted rather satisfactory in this case, but this might be explained by the choice of the AR component, which strongly influences the forecasts. In the following Section, we will analyze the effect of introducing copula models to fit the residuals.

6.7.2.4 Copula Fitting on Residuals

Starting from the residuals obtained in Section 6.7.2.2, we use the filtered data to first fit the marginal distribution of each variable's residuals, and subsequently fit a copula model to the transformed dataset, following the steps discussed in Section 6.1.

Once the data are transformed, we test the time dependence of the copula, using the procedure discussed in Section 6.6. The decision tree, represented in Figure 6.14, identifies two partitions, each identified by a specific Kendall's tau value.

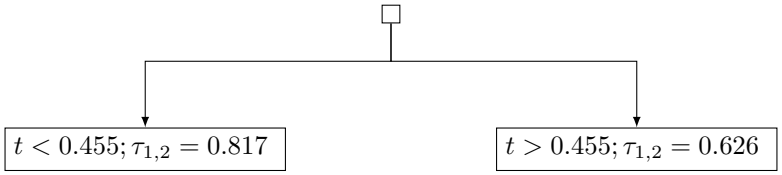


Figure 6.14: *Decision Tree assessing the time dependence of the simulated underlying copula, where time is considered an ordered vector with values in* $[0, 1]$*. The Kendall's Tau value corresponding to each partition is also provided.*

The p-value of the test is equal to 0.002, which suggests the copula structure to be time dependant. Two partitions are identified, one corresponding to 45.5% of the dataset, and the other to the remaining observations. Considering that the original dataset was simulated by simulating half of the residuals from a Gumbel, and half from a t-Student, we can consider the obtained result to be satisfying.

In the following sections we will compare the results obtained by ignoring this information, and therefore by fitting a single copula to the entire dataset, with the ones obtained by fitting different copulas to the partitioned dataset.

6.7.2.5 Predictions Constant Copula

As already anticipated, we start by fitting one copula to the whole dataset, ignoring its time dependence. The result is a Clayton copula with estimated parameter $\hat{\theta} = 1.01$, which is clearly different from the original copula used. To evaluate the goodness of fit, we compute the p-value of the Goodness of Fit test for bivariate copulas, which results to be equal to 0, indicating that parametric family of Clayton copula must be rejected, supporting the conclusion that the fitted model is not correct. However, we still proceed to use the fitted model to forecast the last 30 days, and evaluate the efficiency of the model. This will help us to draw some conclusions on the importance of fitting a good model, once the results will be compared with the ones obtained from fitting a time varying copula.

Following the steps discussed in Section 6.5, we leverage the information gained from fitting an AR-GARCH, and a copula to forecast 30 1-day-ahead predictions. The results are shown in Figure 6.15.

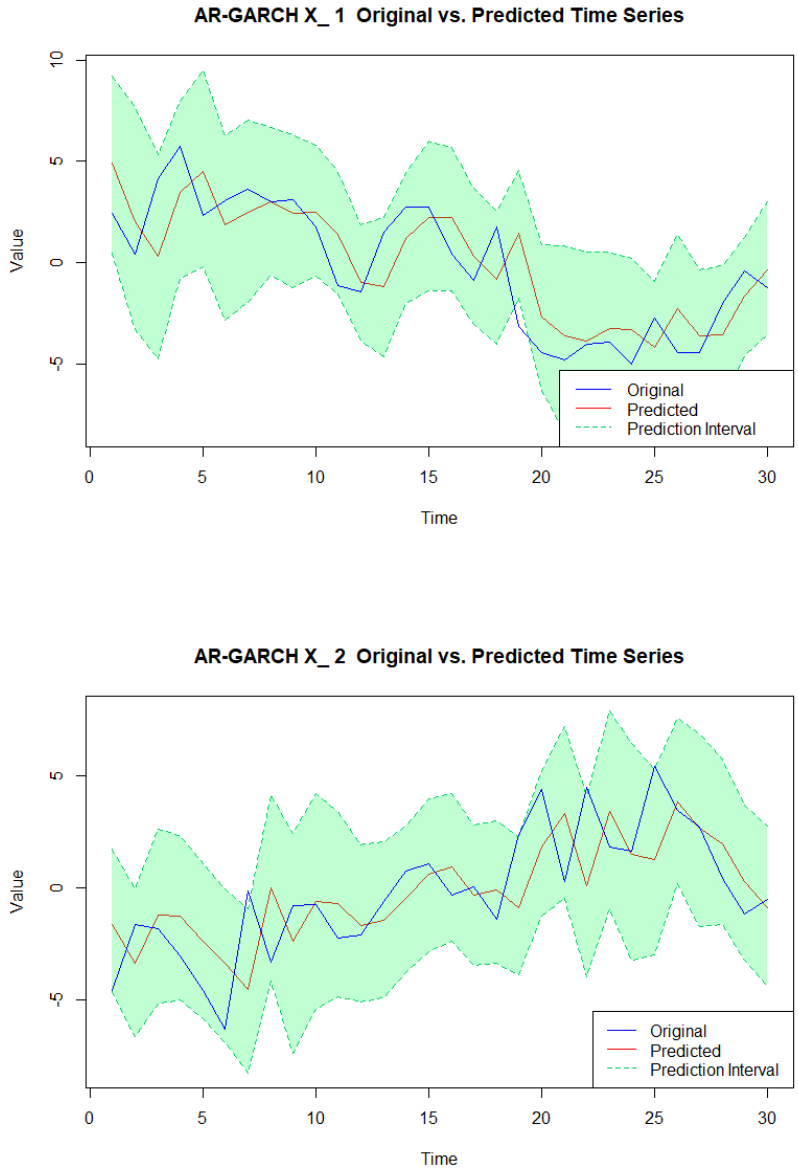


Figure 6.15: *30 days ahead prediction of the two AR(1)-GARCH(1,1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

The prediction, despite being generated with residuals simulated from an incorrect copula, appear rather good. They follow the behaviour of the original time series for both variables, and the errors obtained are very good, as it can be seen from Table 6.11. In particular, the forecasts for $X_{t,2}$ showcase an improvement from the results obtained by assuming independence of the residuals in the percentage of observations, as it corresponds to the expected 90%.

True Distr.	rRMSE	Perc Inside
$X_{t,1}$	0.113	96.7%
$X_{t,2}$	0.223	90%

Table 6.11: *rRMSE of the forecasts, and percentage of times that the observed value lays within the prediction interval.*

6.7.2.6 Time Varying Copula Fitting on Residuals

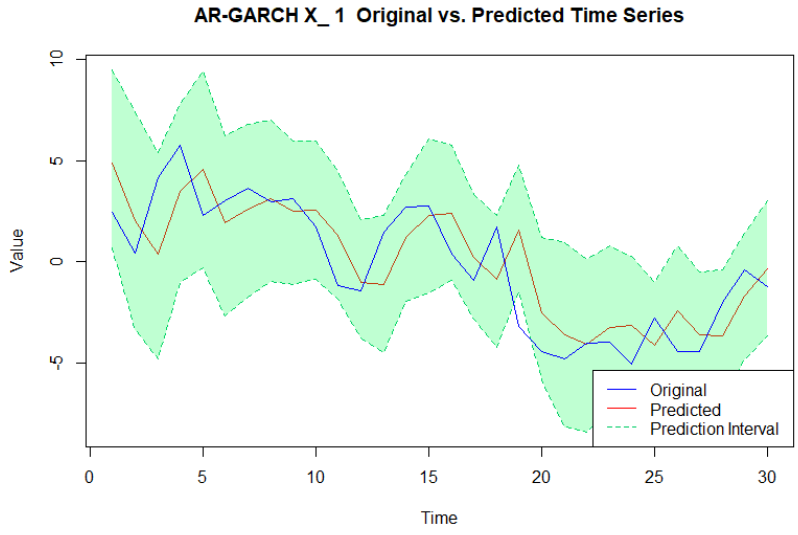
Lastly, from the residuals obtained in Section 6.7.2.2, we want to fit a copula model. This time, however, we do not assume that the underlying copula is constant over time, but rather that changes, as we have discussed how the hypothesis testing suggested the copula to be time dependant.

To address this time dependency, we fit a copula to each partition identified in Section 6.7.2.4. Based on the identified change-points, we split the dataset and fit two separate copula models to the transformed dataset. The results are a Gumbel of parameter $\hat{\theta} = 3.76$ for the first partition, and a t-Student of estimated parameters $\hat{\theta} = 0.83$, $\hat{\nu} = 5.71$. The results are rather close to the original parameters.

Lastly, we assess the goodness of fit of the the second fitted copula (the one we are interested in for the predictions), by using the test discussed in 6.4.2, which results in a p-value of 0.65, an indicator of a good fit.

6.7.2.7 Predictions Time Varying Copula

Following the steps discussed in Section 6.5, we leverage the information gained from fitting a time varying copula-AR(1)-GARCH, to forecast 30 1-day-ahead predictions. The results are shown in Figure 6.16.



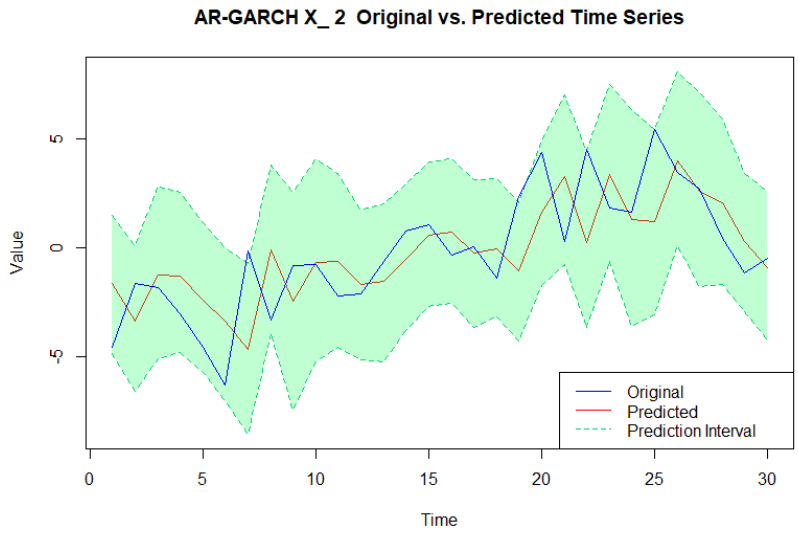


Figure 6.16: *30 days ahead prediction of the two GARCH(1,1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

The $rRMSE$ values shown in Table 6.12 are comparable to the ones obtained from fitting a constant copula to the residuals, and to the ones obtained by only fitting a time series. On the other hand, we can notice an improvement in the intervals, which are both above the expected 90%. We will investigate the reason behind this improvement in Section 6.7.3.

True Distr.	rRMSE	Perc Inside
$X_{t,1}$	0.112	96.7%
$X_{t,2}$	0.211	93.3%

Table 6.12: *rRMSE of the forecasts, and percentage of times that the observed value lays within the prediction interval.*

6.7.3 Additional Results

The predictions for all cases provided rather satisfactory results and exhibit similarities. Figure 6.17 shows that the predicted values are very similar to each other for both variables, with only small variations between the three methods.

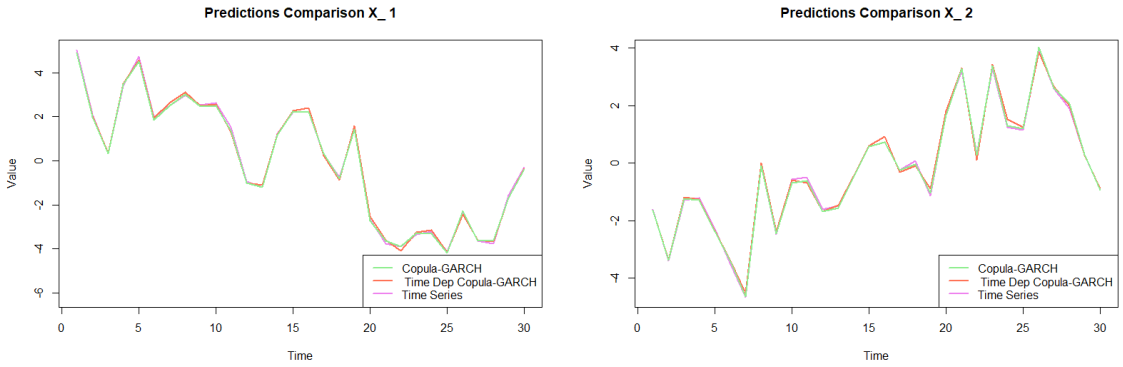


Figure 6.17: *Comparison between the 1 day ahead days ahead prediction of an AR(1)-GARCH(1,1) time series derived by fitting only a time series to the original dataset (blue), or with one that also fitted a copula to the filtered residuals, hence a copula-GARCH model (red).*

The similarity can be attributed to the high value of the autoregressive (AR) component ($\phi_{1,1} = 0.9$), which causes the predictions to be heavily influenced by past values. Additionally, similar results are expected as the advantage of copulas lies in modeling the dependence structure, rather than influencing the marginal predictions. This is evidenced by the improved percentages.

Moreover, as already seen for Case 1, in Section 6.7.1.6, the advantages of fitting a copula to the transformed residuals, translates into an improvement of the ability of capturing extreme events. This ability is evident from Figure 6.18, where it is possible to observe how only the time varying model (orange) is able to produce predictions that mimic the behaviour of the original model, while using the wrong copula (green) does not provide any advantage, and in particular assuming independence of the time series (pink) is reflected in more condensed simulated values, which are not able to imitate the extreme events of the original dataset.

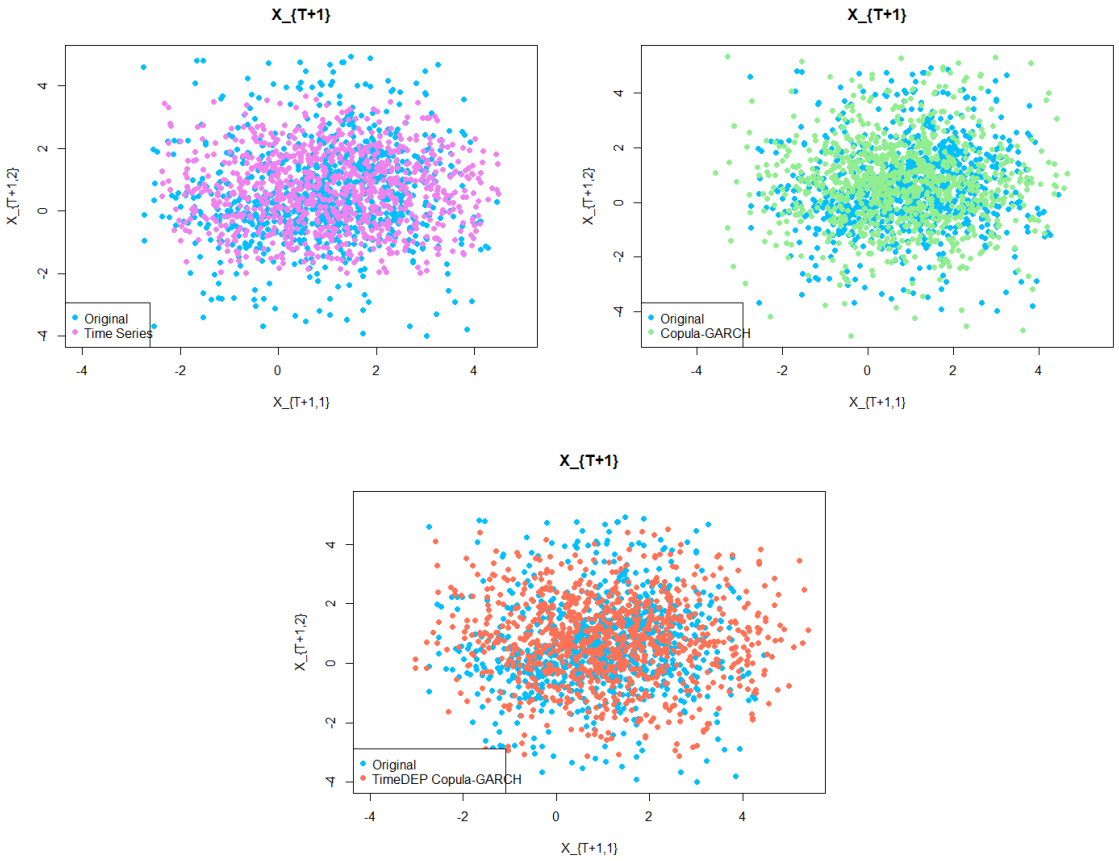


Figure 6.18: *95% percentile distribution clouds from one-day-ahead predictions using $AR(1) - GARCH(1,1)$, a copula-GARCH model, and a time varying copula-GARCH model, compared with points sampled from the true distribution.*

To validate this observation, we also computed the CVaRs for the different simulated clouds, as reported in Table 6.13. The results clearly indicate that the correct copula-GARCH model produces a risk estimate that closely matches the one derived from the original dataset. Although the constant copula-GARCH model also performs relatively well, it is still less accurate than the correct copula-GARCH model. The CVaR obtained from the time series model is less satisfactory.

Method	$CVaR_{95\%}$
Original	1.296
AR-GARCH	1.036
Copula-AR-GARCH	1.596
Time Dep. Copula-GARCH	1.233

Table 6.13: *Conditional Value at Risk obtained from the 95% percentile distribution clouds from one-day-ahead predictions using $AR(1) - GARCH(1,1)$, a copula-GARCH model, and a time varying copula-GARCH model, compared with points sampled from the true distribution.*

7

Mean-CVaR Portfolio with Copula-GARCH Method

In the previous Chapter 6, we linked the concept of time series to the one of copulas, through the introduction of copula-GARCH models. The results obtained can also be extended to vine models, by fitting a vine model to the residuals instead of a copula. However, one step still needs to be discussed. In this Chapter, we will connect these models to the discussion about Portfolio Optimization presented in Chapter 2. In particular, we will build upon the CVaR simplified model of equation (2.28), to derive a new conditional formulation. This formulation incorporates the information gained from fitting copula-GARCH models, enhancing our understanding of portfolio optimization in dynamic market conditions.

The procedure presented in this Chapter expands on the simulation discussed in Section 2.6.7, and represents the final piece of the theoretical framework, allowing the reader to have a clear understanding of all the steps necessary for the practical application of Chapter 8.

7.1 New Conditional CVaR Formulation

In the time series context, the construction of A portfolio is CONDUCTED taking into account the information contained in the time series up to time T, to predict the behaviour of the portfolio at a specific time $T+h, h = 1, 2, \dots$. If we have a multivariate time series $(\mathbf{X}_t)_{t=1,\dots,T} = (X_{t,1}, \dots, X_{t,d})_{t=1,\dots,T}$, we will compute the risk measures conditioning on Ω_T, the information set at T [74]. Specifically, we can derive a new definition of Conditional Value at Risk at time $T+h$, which is computed conditionally on the information set at time T:

$$CVaR_{\beta,T+h}(L)|\Omega_T = \frac{1}{1-\beta}\int_{L(\mathbf{w},\mathbf{X}_{T+h})\geq VaR_{\beta,T+h}(L|\Omega_T)} L(\mathbf{w}, x_{T+h}) f(x_{T+h}|\Omega_T) dx_{T+h}, \tag{7.1}$$

where also the $VaR_{\beta,T+h}(L|\Omega_T)$ is obtained conditioning on the past.

Similarly, we can proceed with the simplification discussed in equation (2.28) of Chapter 2, and obtain the discretized formulation of the Conditional Value at Risk, conditional to Ω_T:

$$\tilde{F}_{\beta,T+h}(L,\alpha)|\Omega_T = \alpha + \frac{1}{Q(1-\beta)}\sum_{k=1}^{Q}(L(\mathbf{w},\hat{\mathbf{X}}_{k,T+h}) - \alpha)\mathbb{1}_{L(\mathbf{w},\hat{\mathbf{X}}_{k,T+h})\geq\alpha}. \tag{7.2}$$

where $\{\hat{\mathbf{X}}_{k,T+h}\}_{k=1,\dots,Q}$ represent the Q sampled vectors of returns at time $T+h$, simulated from the information gained from the observations up to time T. This discretized version of the CVaR allow for the derivation of the Monte Carlo allocation method that will be discussed in the following Subsection, and is a time series adaptation of Problem (2.29).

7.1.1 Monte Carlo Allocation Method

Starting from the copula-GARCH model fitted on the observations at time $1,\dots,T$, we can sample a set of return vectors at time $T+h$, $\hat{\mathbf{X}}_{1,T+h},\dots,\hat{\mathbf{X}}_{Q,T+h}$, from which we can derive the following linear optimization problem, through the employment of the auxiliary variable $u_{k,T+h}$:

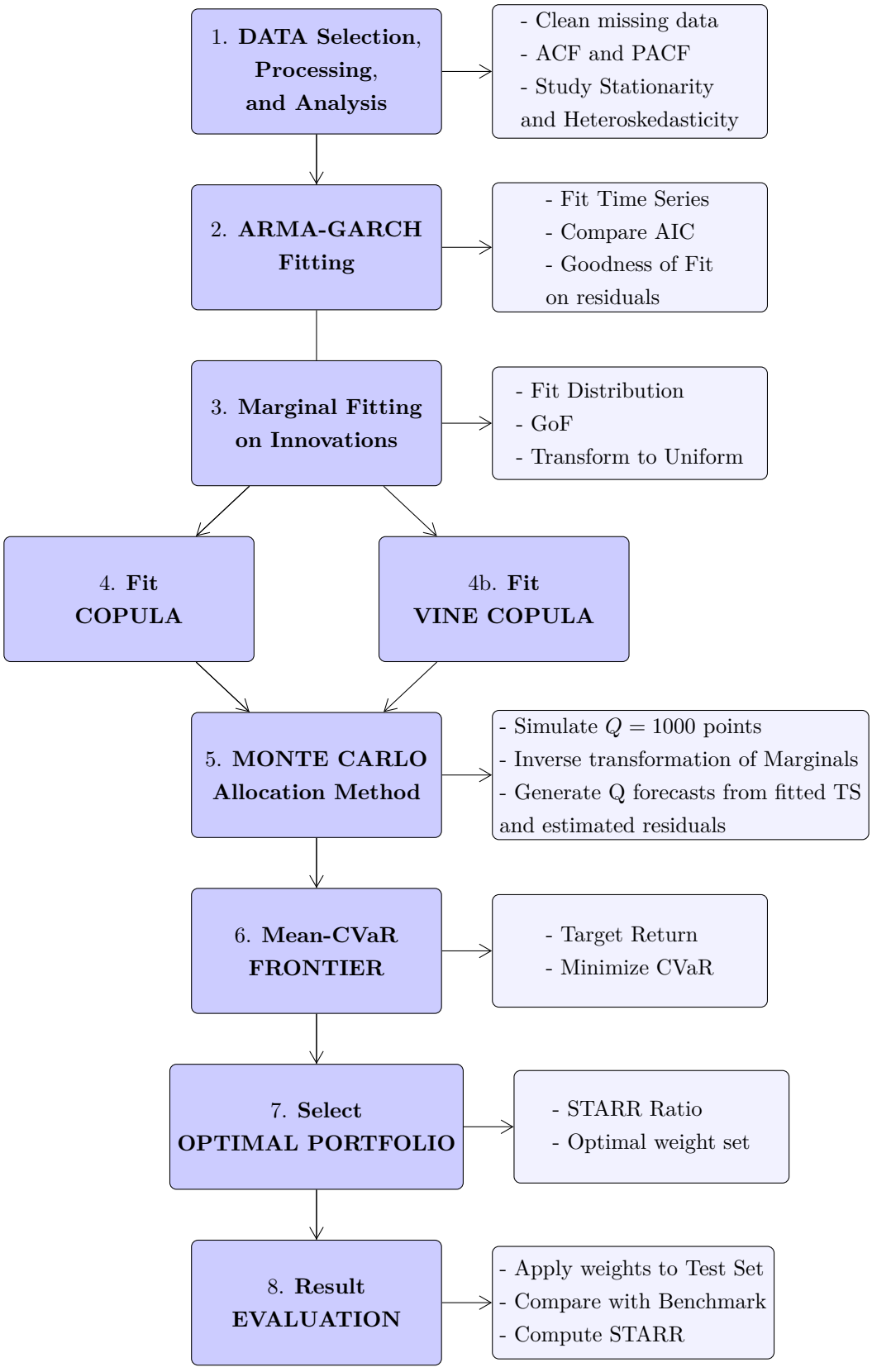
$$\begin{cases} \min_{w,\alpha,u}\ \alpha + \frac{1}{Q(1-\beta)}\sum_{k=1}^{Q} u_{k,T+h} \\ s.t.\ L(\mathbf{w},\hat{\mathbf{X}}_{k,T+h}) + \alpha + u_{k,T+h} \geq 0\ \forall k = 1,\dots,Q \\ s.t.\ u_{k,T+h} \geq 0\ \forall k = 1,\dots,Q \\ s.t.\ \mathbf{w}^T\boldsymbol{\mu}_{T+h} = x_0 \\ s.t.\ \sum_{i=1}^{N} w_i = 1 \\ s.t.\ w_i \geq 0 \\ s.t.\ \alpha \in [0,1]. \end{cases} \tag{7.3}$$

Iterating this problem to a set of target returns x_0, allow us to obtain the set of points of the efficient frontier at time $T+h$. The difference in this new formulation from the problem discussed in [16], is that we now have a dynamic model, which will change over time and evolve as new information is added.

7.2 Steps

In Chapter 8, we will integrate all the methodologies discussed in the previous chapters to create an optimal portfolio, by fitting a Copula-GARCH model to stock returns, and constructing the efficient frontier using Monte Carlo Simulations. We will iterate the process for various partitions of the dataset, enabling us to account for changes in the underlying structure. The main steps to be undertaken are firstly schematized in Section 7.2.1, and subsequently explained more in detail right below, following the structure presented in [75].

7.2.1 Overview



7.2.2 More in Detail

The steps that will be followed in Chapter 8 are the following.

1. Select d financial assets that will compose the portfolio. It is important to choose instruments characterized by readily available data, without too many missing data, and with a sufficiently large history. Perform data cleaning, and processing.

2. Follow the steps discussed in Section 5.2 to fit an ARMA-GARCH model. Compute the corresponding innovations $e_{t,i}$ at each time point t considered.

3. Fit a marginal distribution to the set of residuals of each variable, using a MLE estimator, and use the estimated model to project the innovations to the unit hypercube, applying the transformation $u_{t,i} = \hat{F}_i(e_{t,i})$, $i = 1, \dots, d$, where $\hat{F}_i$ represents the estimated CDF for variable i.

4. Fit a copula or a vine model on the uniformly distributed variables $u_{t,i}$, by using the methods described in Sections 3.4 and 4.4.1. Evaluate the goodness of fit of the resulting model.

5. Generate $Q = 1000$ uniform random vectors $(\hat{u}_{k,t,1}, \dots, \hat{u}_{k,t,d})_{k=1,\dots,Q}$ from the estimated copula or vine model.

6. Transform the simulated uniformly distributed variables by applying the inverse of the marginal distribution estimated in step 3, to obtain a set of standardized residuals $(\hat{e}_{k,t,1}, \dots, \hat{e}_{k,t,d})_{k=1,\dots,Q} = (\hat{F}_1^{-1}(\hat{u}_{k,t,1}), \dots, \hat{F}_d^{-1}(\hat{u}_{k,t,d}))_{k=1,\dots,Q}$.

7. Use the estimated residuals obtained in the previous step, to forecast the returns at time $t + h$, by applying the fitted time series model obtained at step 2:

$$\hat{\mathbf{X}}_{1,T+h}, \dots, \hat{\mathbf{X}}_{q,T+h}.$$

8. Use the forecasted returns to perform a Monte Carlo allocation, and obtain the corresponding Mean-CVaR efficient frontier by optimizing problem (7.3).

9. Select the portfolio on the frontier that maximises the STARR ratio as the optimal portfolio (refer to Section 2.4.2).

10. Evaluate the resulting portfolio, by comparing it to a set of benchmarks. The comparison is carried out by computing the main metrics of the different portfolios, as well as by studying the out-of-sample efficiency of each of them [76]. In particular, we compute statistics such as the mean and variance of the different portfolios. Additionally, we compare the STARR ratios, the CVaRs, and the cumulative returns. Since the goal is to obtain a balanced portfolio, we primarily focus on the STARR ratio to determine the model's success. The cumulative return at time t is obtained by applying the selected set of weights $\mathbf{w}$, to the log-return vector $\mathbf{r}(t_i)$ at each time t_i

of the test set, $T_i = [t_1, \dots, t]$, as follows:

$$R(t) = \sum_{t_i=1}^{t} \mathbf{w}^T \mathbf{r}(t_i). \tag{7.4}$$

7.3 Simulation Study

As for the previous Chapters, we conclude this Chapter with a a simulation study illustrating the steps discussed above, to provide the reader with a full overview of the methodologies used in this thesis. In particular, we will discuss the construction of the portfolios once the estimation of the copula-GARCH is concluded, and the evaluation of the optimal portfolios. We will assume that the model estimation was properly conducted, and proceed with steps (7) to (10) of Section 7.2.1. In particular, we will discuss two separate cases: a portfolio constructed from two set of returns simulated from an AR(1)-GARCH(1,1), and one with the returns derived from a GARCH(1,1).

7.3.1 AR(1)-GARCH(1,1)

In this first simulation, we pick up from the results obtained in Section 6.7.1. We simulate two set of returns of $N = 500$ data points from an AR(1)-GARCH(1,1) model with the same parameters as those used in the previous Chapter in Section 6.7.1. The residuals of these returns are assumed to follow a Gumbel copula with parameter equal to 5. We then assume to have fitted a copula-AR-GARCH model to the training set, consisting of $N = 470$ points. Using this fitted model, we perform a Monte Carlo simulation to construct an optimal portfolio based on the two sets of returns, following the problem outlined in (7.3).

The results from this optimal portfolio will be compared to:

- an equal weight portfolio;
- a portfolio obtained by applying Markovitz's mean-var optimization (refer to 2.3.1);
- a portfolio derived by applying the Monte Carlo allocation method to only the time series fit of the dataset, hence neglecting the copula fit of the residuals.

We start by setting the risk free rate equal to $r_f = 1.7$, and by computing the equal weight portfolio. We then proceed by finding the optimal weights for the mean-variance portfolio on the training set, following the steps outlined in Chapter 2. The results are compared in Figure 7.1, where both Markowitz's portfolio and the Equal Weight one lie on the frontier, hence both resulting optimal.

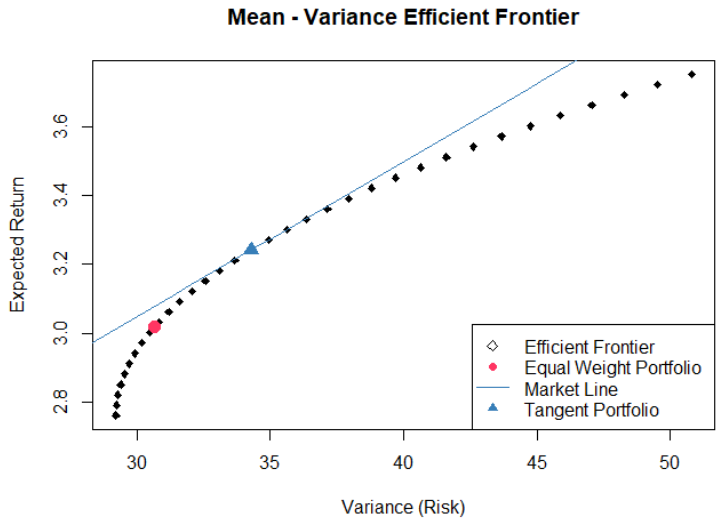


Figure 7.1: *Mean-Variance efficient frontier, computed optimising a portfolio of two sets of returns, derived from a copula-GARCH model. In blue we find Markovitz portfolio, while in red the EW portfolio.*

Subsequently, we simulate $Q = 1000$ points from the fitted copula-GARCH, and from the AR-GARCH fit, starting from the values of the real data at $T = 470$. The resulting clouds of simulated points are shown in Figure 7.2. We can notice how the copula-GARCH points present a more distinguished shape, which is strongly influenced by the underlying copula, while the points simulated from the time series fit appear more concentrated around the starting point. This plot clearly demonstrates that only the copula-GARCH model effectively captures the dependence structure of the variables, which is distinctly positive, while the points simulated from the AR-GARCH appear independent.

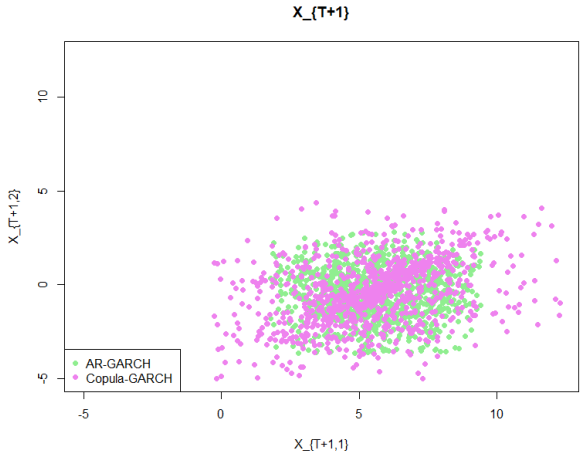


Figure 7.2: *95% percentile distribution clouds from one-day-ahead predictions using $AR(1) - GARCH(1,1)$, and the copula-AR-GARCH model.*

The simulated points are then used to construct an optimal mean-CVaR frontier using the steps illustrated in Section 7.1 for $h = 1$, hence for the following day, and to identify the portfolio with the best STARR ratio. The result for the copula-GARCH case is shown in Figure 7.3, while the one for the time series case is shown in Figure 7.4. The weights, and corresponding results will be commented on the following Section, but it is already visible how the Copula-GARCH one opts for a more risky

portfolio.

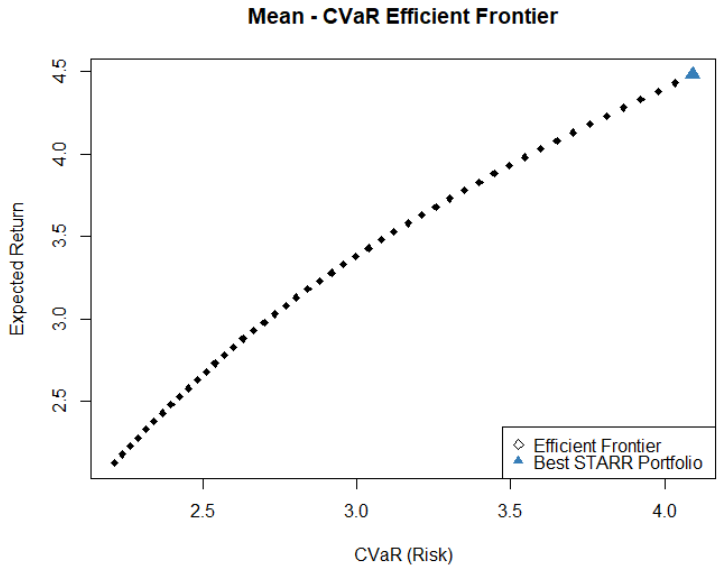


Figure 7.3: *Mean-CVaR efficient frontier, computed with a Monte Carlo Allocation method, where we generate* $Q = 1000$ *data from a fitted copula-GARCH model. The results are obtained optimising a portfolio of two sets of returns, derived from a copula-GARCH model.*

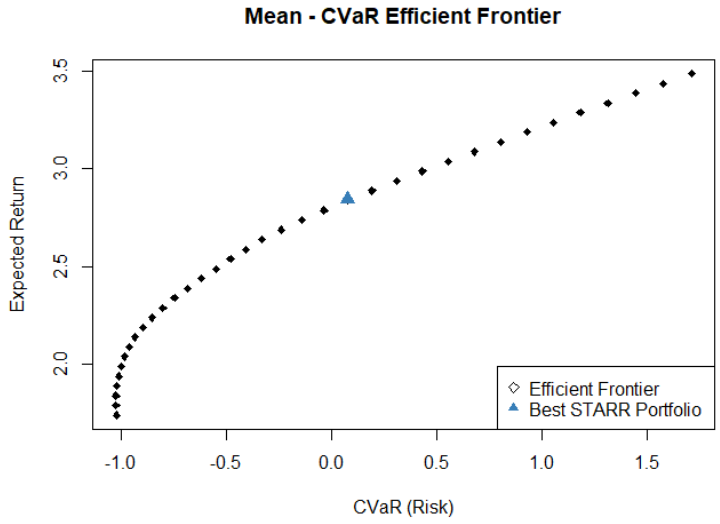


Figure 7.4: *Mean-CVaR efficient frontier, computed with a Monte Carlo Allocation method, where we generate* $Q = 1000$ *data from a fitted AR(1)-GARCH(1,1) model. The results are obtained optimising a portfolio of the two simulated sets of returns.*

7.3.1.1 Evaluate the Portfolio

The portfolios obtained from each method are compared in Table 7.1. In particular, as already anticipated, the copula-GARCH portfolio appears more extreme, preferring to invest mainly on the first asset, which results in a greater risk appetite: the portfolio is characterized by higher return, and higher CVaR. The Mean-var portfolio detains the best Sharpe ratio, while the copula-GARCH portfolio presents the

best STARR ratio.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(50\%, 50\%)$	3.019	9.412	4.3%	14.0%
Mean-var	$(59.15\%, 40.85\%)$	3.241	9.428	**4.5**%	16.3%
Copula-GARCH	$(99.12\%, 0.88\%)$	4.213	14.307	3.3%	**17.6**%
AR-GARCH	$(28.35\%, 71.65\%)$	2.492	10.919	2.6%	7.3%

Table 7.1: *Table comparing the portfolios obtained by optimizing the weights of two returns simulated from a copula-GARCH model, using different methods. The table reports the weights and the results of the risk measures and ratios computed numerically on the training set of $N = 470$ points.*

Subsequently, we proceed with an out-of-sample efficiency evaluation of the portfolios considered. We apply the selected weights from each method to the $N = 30$ points of the test set, plotting the obtained cumulative returns computed using equation (7.4), the Conditional Value at Risks, and ultimately comparing the STARR ratios. The results are plotted in Figure 7.5 and Figure 7.7, while in Figure 7.6 the differences of the first two plots are highlighted.

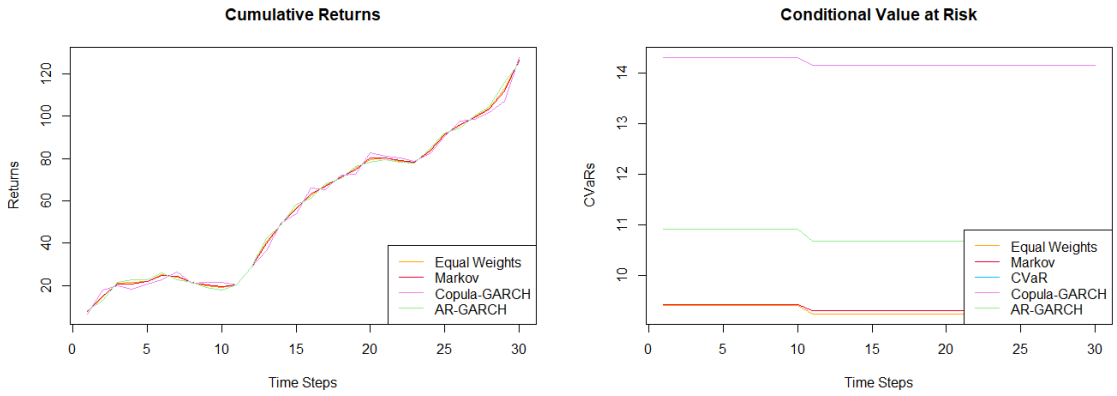


Figure 7.5: *Evolution of the Cumulative Returns and Conditional Value at Risks obtained by applying the optimal weights from each method to $N = 30$ data points of the test set.*

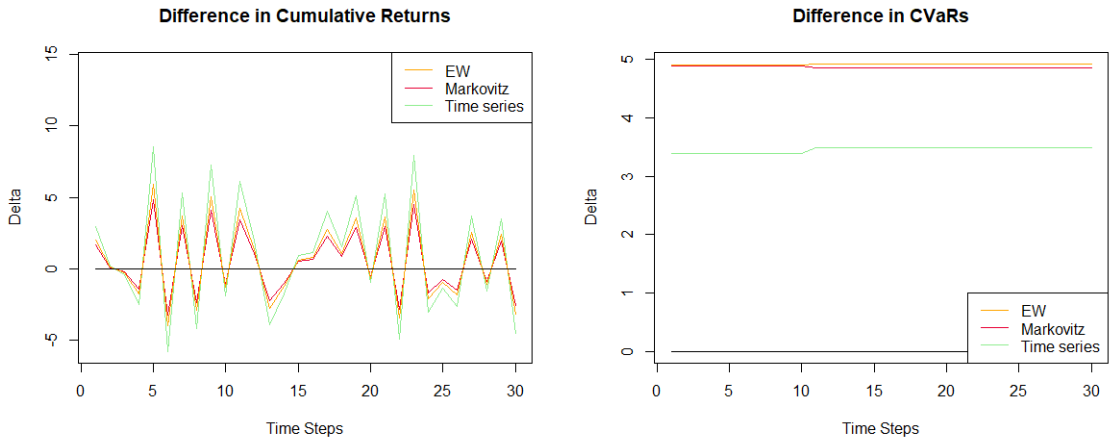


Figure 7.6: *Difference of the evolution of the Cumulative Returns and Conditional Value at Risks obtained by applying the optimal weights from each method to $N = 30$ data points of the test set.*

In the first period of the plot above, we can see how the cumulative returns obtained for the different methodologies are rather comparable, and the results vary at different times considered. This is even more clear by observing the plot of the difference of cumulative returns in Figure 7.6, where the returns of all other portfolios are subtracted from the copula-GARCH one: the difference is oscillating above and below zero, without a clear best. On the contrary, the copula-GARCH appears to be riskier for the entire period considered, as all differences appear positive in the CVaR plot.

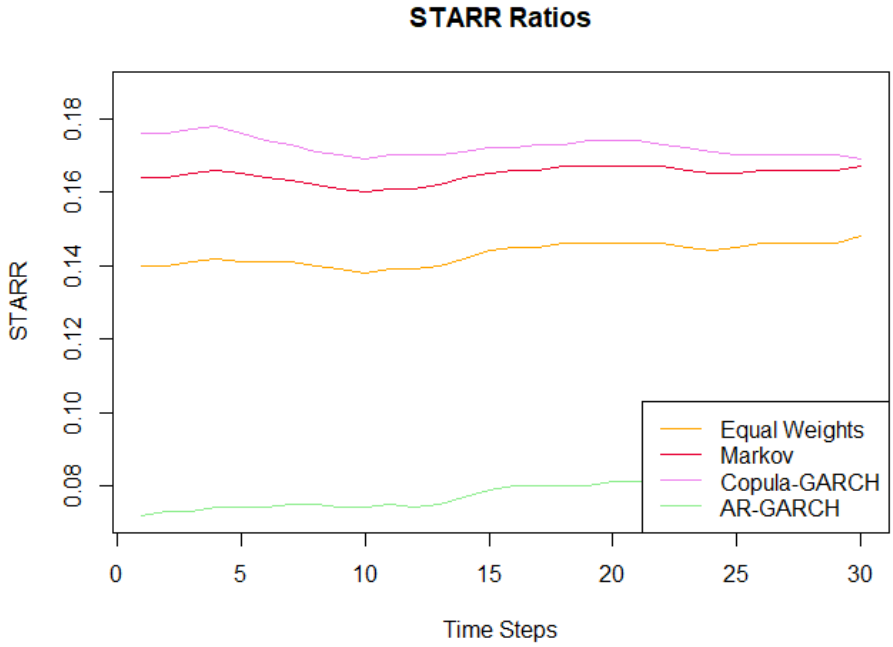


Figure 7.7: *Evolution of the STARR ratios obtained by applying the optimal weights from each method to $N = 30$ data points of the test set.*

Lastly, we compare the STARR ratios of each portfolio across the test set period, to determine which one is the best performing one. Clearly, for all 30 days considered, the copula-GARCH portfolio (in pink) outperforms the others, as it has the highest ratio. To confirm this improvement, we compare the considered metrics at time $T = 30$ of the test set in Table 7.2. The GARCH portfolio appears to have the highest cumulative return, the copula-GARCH one however confirms to be the riskiest one, but most importantly the best one in terms of STARR ratio, outperforming the other ones, with approximately a 1.20%. improvement from the second best, which would be Markowitz's portfolio.

Method	Cumulative Return	CVaR	STARR
Equal Weight	118.48	9.242	14.8%
Mean-Variance	117.90	9.301	16.7%
Copula-GARCH	115.34	**14.159**	**16.9**%
GARCH	**119.86**	10.68	8.5%

Table 7.2: *Cumulative returns, CVaRs and STARR ratios for each method at the last day of the test set, hence at $T = 500$.*

7.3.2 GARCH(1,1) Case

For the second example, we repeat the procedure illustrated in the fist simulation, simulating two return vectors from a GARCH(1,1), characterized by the following parameters: $(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}) = (1, 0.7, 0.2)$, and $(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}) = (1, 0.8, 0.1)$. The residuals are again obtained simulating the data from a Gumbel of parameter equal to 6, and by transforming the uniform data to two t-students of parameters $\nu_1 = 3$ and $\nu_2 = 4$. We split the dataset in a training set of $N = 470$ observations, and a test set of $N = 30$ points, for a total of 500 simulated points. The risk free rate is again set equal to $r_f = 1.7$.

In Figure 7.8, we present the mean-Variance frontier, and compare its optimal portfolio (blue triangle) with the equal weight portfolio, which lies as well on the frontier, hence both resulting optimal.

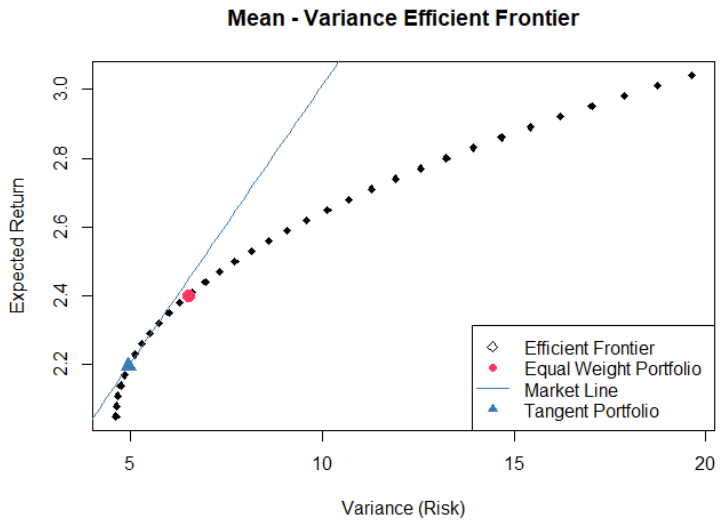


Figure 7.8: *Mean-Variance efficient frontier, computed optimising a portfolio of two sets of returns, derived from a copula-GARCH model. In blue we find Markovitz portfolio, while in red the EW portfolio. The blue line represents the market line.*

We then simulate $Q = 1000$ points from the fitted copula-GARCH, and the fitted GARCH(1,1), starting from the values of the real data at $T = 470$. The resulting clouds of simulated points are shown in Figure 7.9. The copula-GARCH simulated points exhibit a more elongated shape, derived by its ability to incorporate While the points simulated from the time series appear independent and more concentrated around a central point, which is comparable to the results obtained in the AR-GARCH case previously discussed.

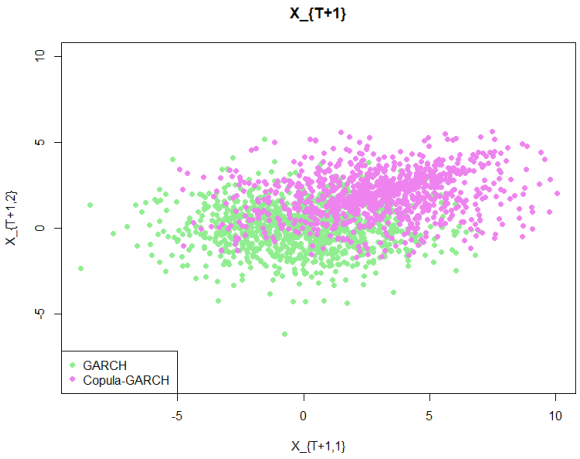


Figure 7.9: *95% percentile distribution clouds from one-day-ahead predictions using $GARCH(1,1)$, and the copula-GARCH model.*

The simulated points are then used to construct a optimal mean-CVaR frontier using the steps illustrated in Section 7.1 for $h = 1$, and to identify the portfolio with the best STARR ratio. The resulting frontier for the copula-GARCH case is shown in Figure 7.10, while the time series case is presented in Figure 7.11.

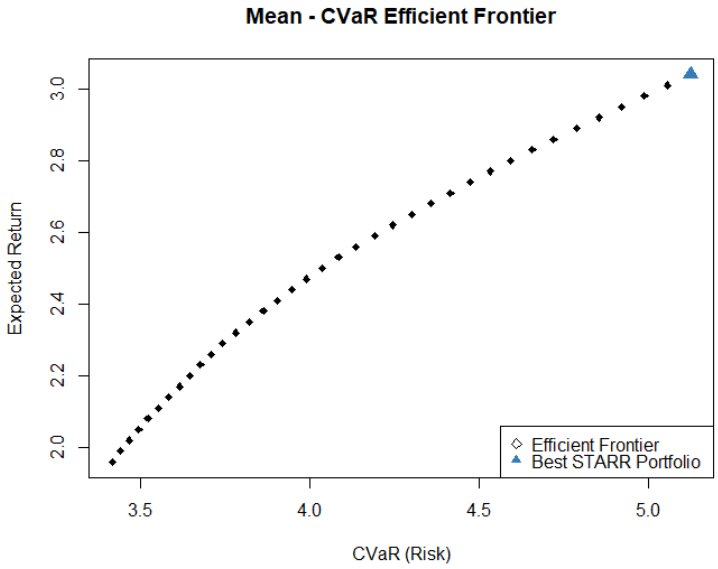


Figure 7.10: *Mean-CVaR efficient frontier, computed with a Monte Carlo Allocation method, where we generate $Q = 1000$ data from a fitted copula-GARCH model. The results are obtained optimising a portfolio of two sets of returns, derived from a copula-GARCH model.*

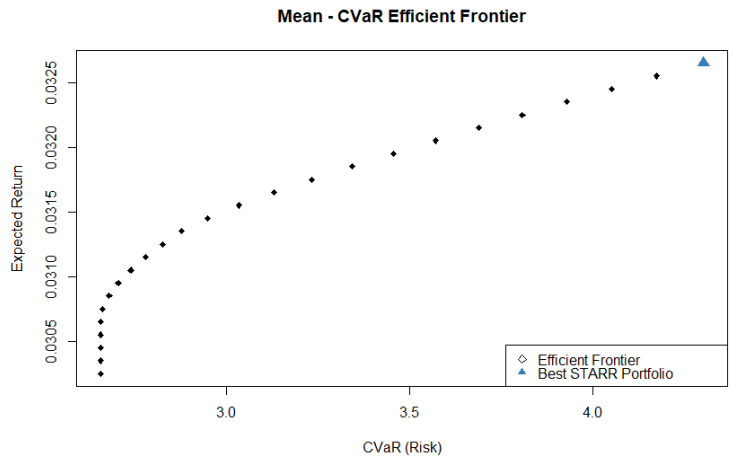


Figure 7.11: *Mean-CVaR efficient frontier, computed with a Monte Carlo Allocation method, where we generate $Q = 1000$ data from a fitted GARCH(1,1) model. The results are obtained optimising a portfolio of the two simulated sets of returns.*

7.3.2.1 Evaluate the Portfolio

The portfolios obtained for the different methods are compared in Table 8.18. In particular, we compare the weights of the different portfolios, noting how both portfolios obtained applying the Monte Carlo allocation method appear more extreme, but selecting opposite variables. Each set of weights is then applied to the original dataset, to compare the different measures.

The GARCH(1,1) portfolio appears to have lower return than the other ones, but higher risk (CVaR) than the first two: this indicates that the obtained portfolio is not optimal, and the procedure involving only time series fitting might not be appropriate for this case. This underlines the importance of fitting a good model able to incorporate all the information of the original one results important to achieve good results, which was already discussed in the simulation study of the previous Chapter (Section 6.7). On the other hand, the copula-GARCH portfolio appears to have higher risk, but also characterized by greater return. The Equal Weight portfolio detains the best Sharpe ratio, while the best STARR ratio is obtained by the copula-GARCH portfolio, highlighting again the efficacy of this model.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(50\%, 50\%)$	2.401	3.023	**10.8**%	23.2%
Markovitz	$(35.14\%, 64.86\%)$	2.197	3.045	10.1%	16.3%
Copula-GRACH	$(98.10\%, 1.90\%)$	3.062	5.198	6.7%	**26.2**%
GARCH	$(0.05\%, 99.95\%)$	1.714	4.809	0.2%	0.3%

Table 7.3: *Table comparing the portfolios obtained by optimizing the weights of two returns simulated from a copula-GARCH model, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the training set of $N = 470$ points.*

We continue with an out-of-sample efficiency evaluation of the portfolios considered. We apply the selected weights to $N = 30$ points of the test set, computing the cumulative returns, the CVaRs and the STARR ratio at each time step, and plotting the obtained results in Figures 7.12 and 7.14.

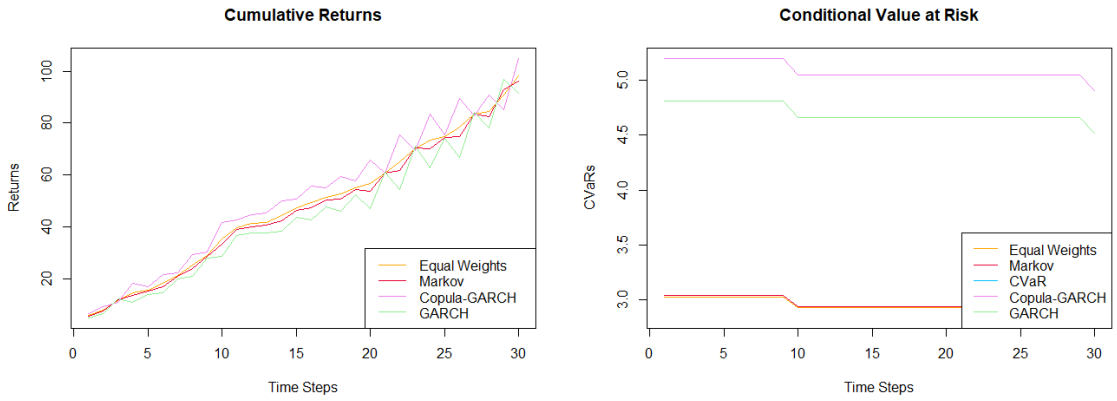


Figure 7.12: *Evolution of the Cumulative Returns and Conditional Value at Risks obtained by applying the optimal weights from each method to $N = 30$ data points of the test set.*

In the first plot above, we can see how the copula-GARCH portfolio is characterized by greater cumulative returns throughout the entire test period considered, with a small exception toward the end, this is even more clear from the plot of the differences shown in Figure 7.13, where the differences of the cumulative returns are shown. Moreover, coherently with the results discussed for the AR-GARCH case, the copula-GARCH portfolio also remains the riskiest, having the greatest CVaR for all the $N = 30$ days considered. While the GARCH portfolio appears clearly under-performing the other ones, as it still has high CVaR, while consistently showing the lowest cumulative returns.

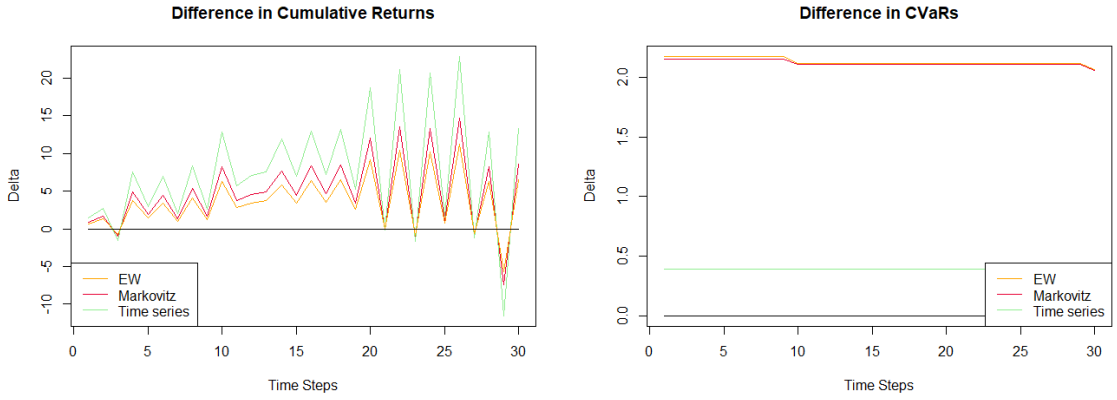


Figure 7.13: *Difference of the evolution of the Cumulative Returns and Conditional Value at Risks obtained by applying the optimal weights from each method to $N = 30$ data points of the test set.*

Lastly, we compare the STARR ratios of each portfolio, to determine the best performing one. For all 30 days considered, the copula-GARCH portfolio (in pink in Figure 7.14) outperforms the others, as it is characterized by the highest ratio.

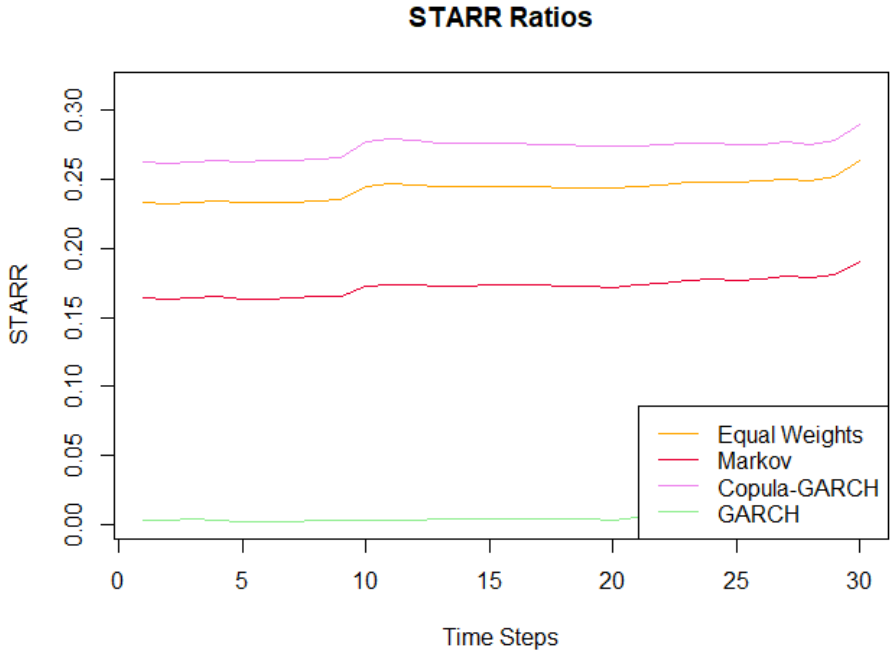


Figure 7.14: *Evolution of the STARR ratios obtained by applying the optimal weights from each method to $N = 30$ data points of the test set.*

To confirm the superiority of the copula-GARCH portfolio, we compare the considered metrics at time $T = 500$ in Table 7.4. Once again, the GARCH portfolio appears to have the lowest cumulative return, but still high CVaR, which result in a very poor STARR value. On the other hand, the copula-GARCH portfolio, the riskiest, stands out as the best in terms of the STARR ratio, outperforming the other ones, with approximately a 10.32% improvement from the second best, which would be the Equal Weight portfolio.

Method	Cumulative Return	CVaR	STARR
EW	98.29	2.839	26.4%
Mean-Var	96.28	2.842	19.0%
COPULA	**104.83**	**4.905**	**29.0**%
GARCH	91.51	4.518	1.1%

Table 7.4: *Cumulative returns, CVaRs and STARR ratios for each method at the last day of the test set, hence at $T = 500$.*

8

Application to Financial Returns

In this Chapter, we will combine the theory previously discussed, and apply the models to a real dataset. Specifically, the closing prices of four of the main international stock indexes will be modeled to construct an optimal portfolio. A Vine copula-ARMA-GARCH model will be constructed to describe the relationship between four of the main stock indexes: the German index (DAX), the French Index (CAC), the Japanese Nikkei 225 (N225), and the American S&P500 (GSPC).

We will start by conducting some data preparation, and proceed to fit an appropriate time series to the transformed dataset. The filtered dataset will be used to fit a Vine model (Chapter 4), using the procedure introduced in Chapter 6. Lastly, the copula-GARCH fit of the time series will be used to construct an efficient frontier on the Mean-CVaR plane, as discussed in Chapter 7.

8.1 Data Selection, Collection, and Processing

An important aspect of portfolio optimisation involves choosing a diverse basket of financial instruments. From the investor's prospective, the goal is to assemble a well-rounded portfolio composed of different instruments, thereby achieving diversification — a fundamental principle in finance, crucial for risk reduction. As discussed in Chapter 2, investors tailor their selection to align with their risk preferences, but a generally good approach would be to alternate secure stocks, such as state indexes of more developed countries, and riskier ones.

Instead of focusing on diversification, for the real application described in this Chapter we will prioritize data quality. Hence, we will prioritize a selection of stocks with clean data, minimal data gaps, and a substantial data history. This will ensure the availability of reliable information for a reasonable time frame, allowing the creation of a statistically significant and more stable model.

8.1.1 Data Selection

We start by collecting the daily closing prices p_i of the four indexes selected: the German index (DAX), the French Index (CAC), the Japanese Nikkei 225 (N225), and the American S&P500 (SP500). The time series are considered in the period between October 23^{rd}, 2014 and July 5^{th}, 2024.

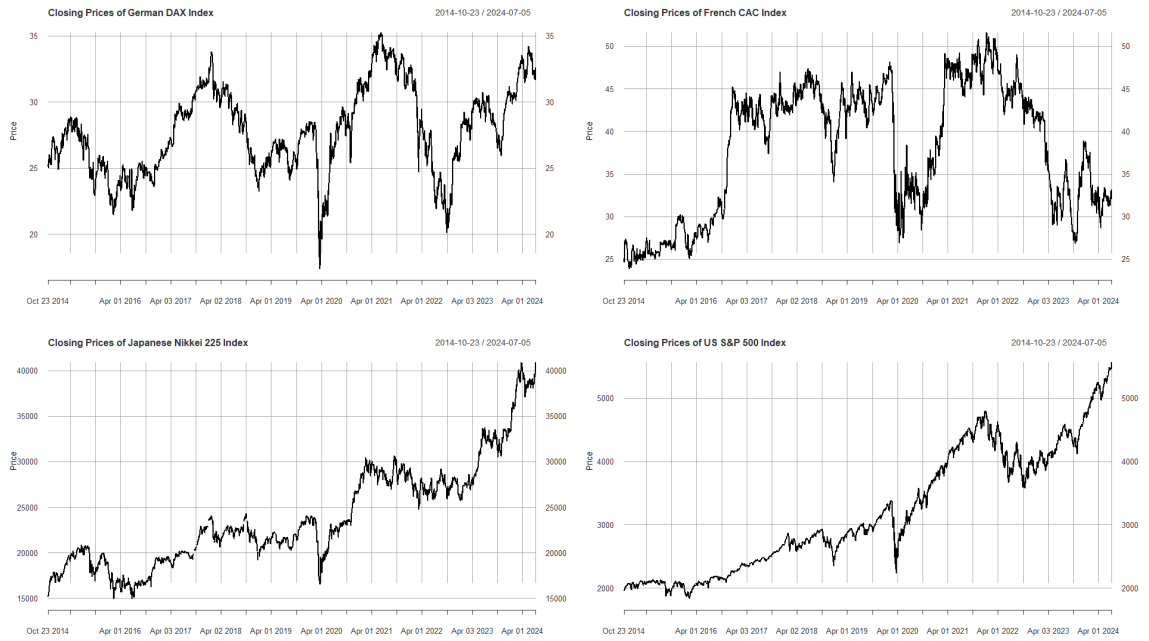


Figure 8.1: *Closing prices of the German index (DAX), French Index (CAC), Japanese (N225), and American (SP500), computed in the period from* 23/10/2014 *to* 5/7/2024*.*

8.1.2 Data Cleaning

Once the basket of assets is chosen, the next step involves data pre-processing to ensure a clean dataset. This requires identifying the largest time window at which all stocks have data available, and merging the data into a single dataframe. We then look for missing data: a few missing data points may not pose a significant issue. Various methods, such as using the last available value, the first value available afterward, or employing some type of interpolation, can be used. However, we prefer to simply remove the dates with missing data, as this ensures that we do not wrongly influence the volatility, and that the data retains its underlying structure, without altering the properties of the time series.

The observations for the weekends are not available and are therefore neglected. Moreover, the Japanese index presented 21 missing data. The issue has been resolved by deleting the dates for all indexes considered: we are left with 2289 observations for each index.

8.1.3 Data Processing

Once the dataframe is cleaned, we are able to compute the daily log-returns for each asset by applying equation (2.3). The results are shown in Figure 8.2. Coherently with what we expected, the time series appear stationary, showing volatility clustering typical of financial series.

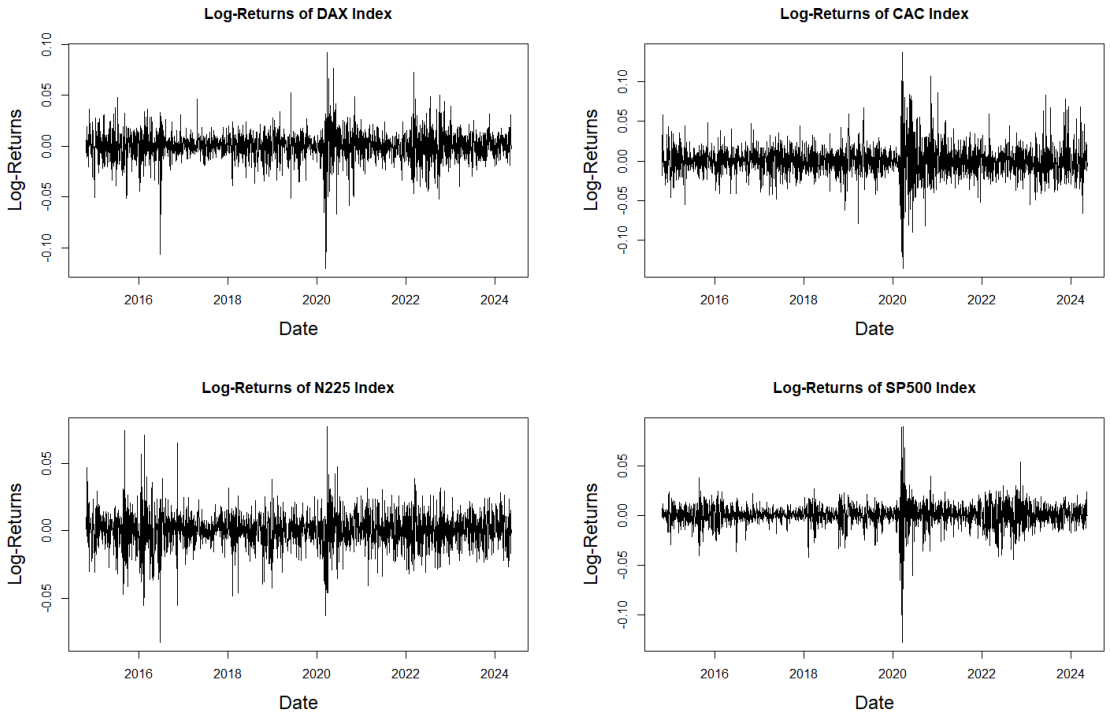


Figure 8.2: *Log-Returns of the closing price of the German index (DAX), French Index (CAC), Japanese (N225), and American (SP500), computed in the period from* 23/10/2014 *to* 5/7/2024.

8.2 Preliminary Analysis

To gain insights on the data and understand their characteristics, we proceed to analyze the dataframe. We start by computing the main statistics of each time series. Once we have a general overview of the data, we proceed to test some of the main features that are required for time series analysis. We therefore compute and plot the ACF and PACF of each time series (see Def. 25 and 26). Lastly, we proceed to test the stationarity (discussed in Sect. 5.3.1) of the log-returns, by performing a Dickey–Fuller Test (Sect. 5.3.1.1), and their heteroskedasticity (Sect. 5.3.3) with a McLeod-Li Test.

In Table 8.1 we provide the main summary statistics of the log-returns. All four data present a mean value very close to zero. Moreover, for most geographies considered, we observe a negative skew and a high kurtosis, which implies the presence of heavy tails and of a slightly left-skew; the only exception is the French index, which is characterized by a positive skew. Lastly, we performed a Jarque–Bera Test (JB), a statistical test that combines the values of Kurtosis and Skewness to test the normality of the returns: the high values indicate that the null hypothesis must be rejected, and hence that the distributions of all log-returns cannot be considered normal.

Index	Mean	St Dev	Skewness	Kurtosis	JB
DAX	$1.11 * 10^{-4}$	0.014	-0.785	8.675	7429
CAC	$1.06 * 10^{-4}$	0.020	0.101	7.007	4698
N225	$4.34 * 10^{-4}$	0.013	-0.097	4.266	1745
SP500	$4.58 * 10^{-4}$	0.012	-0.967	16.172	25352

Table 8.1: *Summary statistics of the considered log-Returns computed from the closing price of the German index (DAX), French Index (CAC), Japanese (N225), and American (SP500).*

Lastly, we provide in Figure 8.3 the scatter plots of all pair of variables, the histograms showing their distribution and the Kendall's tau values for each pair, from which we can extract that the German index and the American one are the most correlated ones, with a value of $\tau = 0.49$.

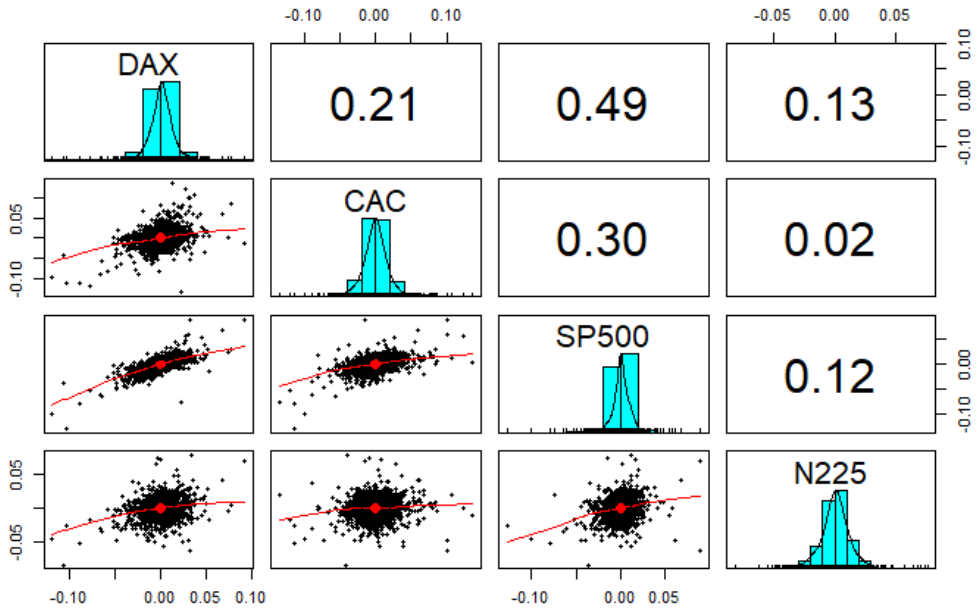


Figure 8.3: *Summary plot showing the histograms of the log-return of each index, the scatterplot of each pair combination, and Kendall's tau correlation of the different pairs.*

8.2.1 ACF and PACF

In Figure 8.4, the ACF and PACF plots of the log returns of the four indexes are shown. The ACFs of all indexes have extremely low values, indicating that the MA component should be zero for all variables,. The PACFs also present rather low values, with the exception of the CAC and SP500 log-returns, which suggest that the time series might have an AR component of order $p = 1$.

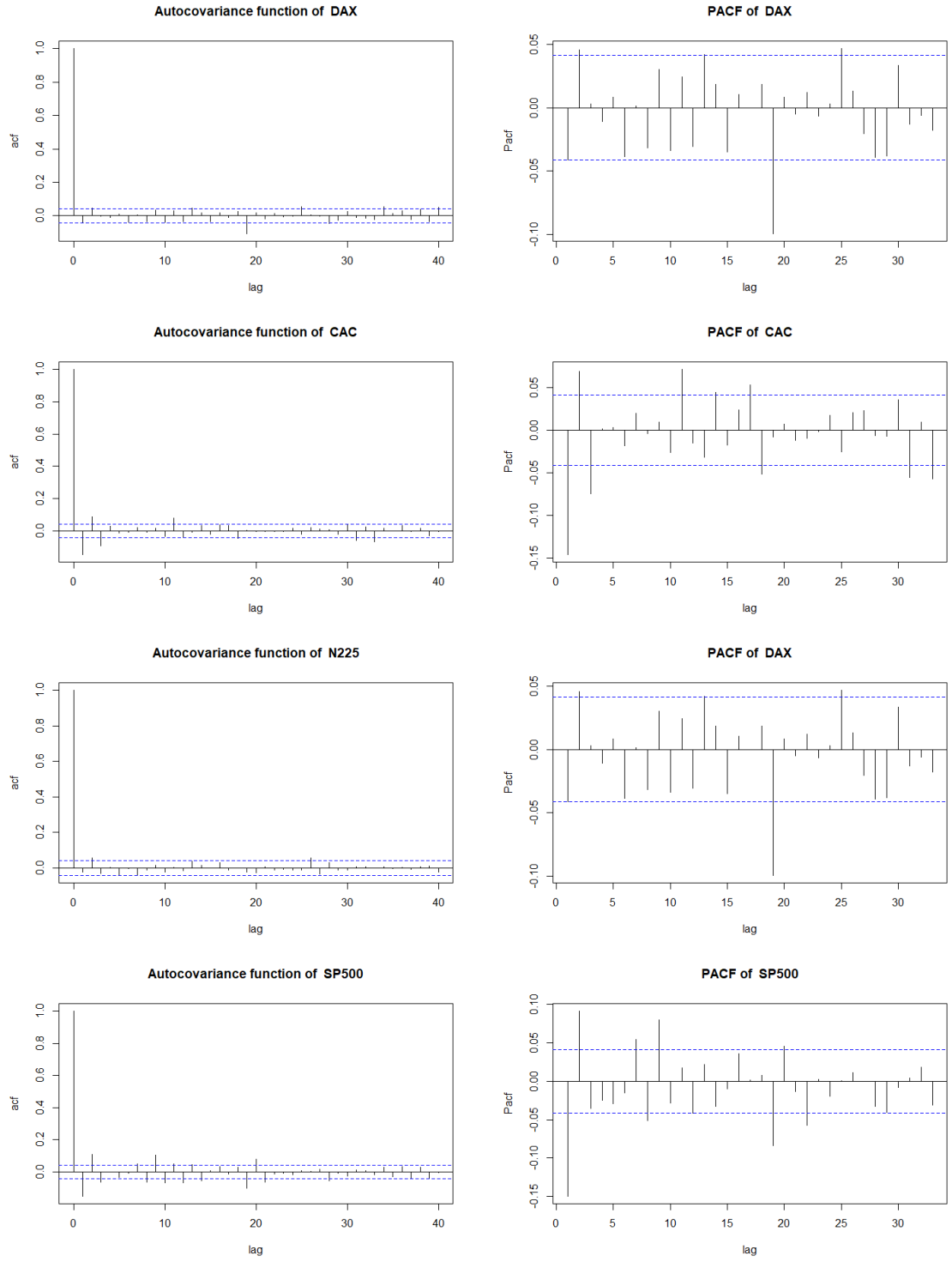


Figure 8.4: *ACF AND PACF of the log-returns of the closing price of the DAX index, CAC Index, N225, and the SP500, computed in the period from* 23/10/2014 *to* 5/7/2024.

8.2.2 Stationarity of the dataset

Furthermore, we perform an Augmented Dickey–Fuller Test (ADF) to assess the stationarity of the time series considered. The results are reported in Table 8.2, where all times series considered show a p-value lower than the 0.05 fixed significance level. Thus, we can reject the null hypothesis that the

time series have a unit root in favor of the alternative. The time series can be considered stationary.

Index	ADF p-Value
DAX	0.01
CAC	0.01
N225	0.01
SP500	0.01

Table 8.2: *P-values of the Augmented Dickey-Fuller test for stationarity of a time series.*

8.2.3 Heteroskedasticity of the Dataset

Lastly, in Table 8.3 the results of the McLeod-Li test are shown. For all time series considered, the p-value appear lower than the 0.05, suggesting that the log-returns of each index are heteroskedastic. When fitting a time series model, we will therefore include a GARCH component, to ensure that we correctly incorporate this property in the fit.

Index	McLeod-Li p-Value
DAX	6.43×10^{-10}
CAC	0
N225	0
SP500	0

Table 8.3: *P-values of the McLeod-Li test for heteroskedasticity of a time series.*

8.3 ARMA-GARCH Fitting

For each time series considered, we then proceeded to fit an appropriate time series model. To accomplish this, we conducted a grid search fitting all combinations of ARMA(i, j)-GARCH(k, k), with $i, j, k \in \{0, 1, \ldots, 4\}$. The parameters were fitted by applying a Quasi-Maximum-Likelihood approach, which is a simplification of the Maximum Likelihood method described in Subsection F.1.1 of the Appendix. The model with the lowest AIC value was chosen.

In Tables 8.4, 8.5, 8.6, and 8.7 we compare the fit of different ARMA-GARCH models. By examining the AIC values, we determine that the GARCH(1,1) model consistently provides the best fit. The selected combinations, and the corresponding fitted parameters are summarized in Table 8.8.

	AIC	DAX Parameters
ARMA(0, 0)-GARCH(1, 1)	**13351**	$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0, 0.101, 0.868)$
ARMA(1, 0)-GARCH(1, 1)	13355	$(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.036, 0, 0.099, 0.871)$
ARMA(0, 1)-GARCH(1, 1)	13355	$(\hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.036, 0, 0.099, 0.871)$
ARMA(1, 1)-GARCH(1, 1)	13358	$(\hat{\phi}, \hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0.537, -0.569, 0, 0.099, 0.870)$

Table 8.4: *AIC values obtained by fitting different time series models to the log-returns of the German DAX index. We highlighted the best values in bold.*

	AIC	CAC Parameters
ARMA(0, 0)-GARCH(1, 1)	**11976**	$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0, 0.079, 0.893)$
ARMA(1, 0)-GARCH(1, 1)	11989	$(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.078, 0, 0.073, 0.902)$
ARMA(0, 1)-GARCH(1, 1)	11989	$(\hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.078, 0, 0.074, 0.901)$
ARMA(1, 1)-GARCH(1, 1)	11992	$(\hat{\phi}, \hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.378, -0.302, 0, 0.726, 0.904)$

Table 8.5: *AIC values obtained by fitting different time series models to the log-returns of the French CAC index. We highlighted the best values in bold.*

	AIC	N225 Parameters
ARMA(0, 0)-GARCH(1, 1)	**13645**	$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0, 0.107, 0.850)$
ARMA(1, 0)-GARCH(1, 1)	13649	$(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.035, 0, 0.107, 0.849)$
ARMA(0, 1)-GARCH(1, 1)	13649	$(\hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.032, 0, 0.107, 0.849)$
ARMA(1, 1)-GARCH(1, 1)	13655	$(\hat{\phi}, \hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.658, -0.615, 0, 0.107, 0.850)$

Table 8.6: *AIC values obtained by fitting different time series models to the log-returns of the Japanese N225 index. We highlighted the best values in bold.*

	AIC	SP500 Parameters
ARMA(0, 0)-GARCH(1, 1)	**14828**	$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (0, 0.203, 0.772)$
ARMA(1, 0)-GARCH(1, 1)	14834	$(\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.050, 0, 0.021, 0.771)$
ARMA(0, 1)-GARCH(1, 1)	14835	$(\hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.048, 0, 0.020, 0.771)$
ARMA(1, 1)-GARCH(1, 1)	14841	$(\hat{\phi}, \hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1) = (-0.715, -0.664, 0, 0.208, 0.768)$

Table 8.7: *AIC values obtained by fitting different time series models to the log-returns of the American SP500 index. We highlighted the best values in bold.*

Index	Model	$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$
DAX	ARMA(0, 0) + GARCH(1, 1)	$(0, 0.100, 0.869)$
CAC	ARMA(0,0) + GARCH(1, 1)	$(0, 0.079, 0.891)$
N225	ARMA(0, 0) + GARCH(1, 1)	$(0, 0.107, 0.850)$
SP500	ARMA(0, 0) + GARCH(1, 1)	$(0, 0.204, 0.771)$

Table 8.8: *Time Series Model chosen for each variable.*

From the fitted model, we extracted the residuals $\epsilon_{i,t}$, $i = \{\text{DAX, CAC, N225, SP500}\}$, which are computed as the standardized difference between the true value and the fitted one (see Def. 5.3). The plot of the time series of the residuals is shown in Figure 8.5,

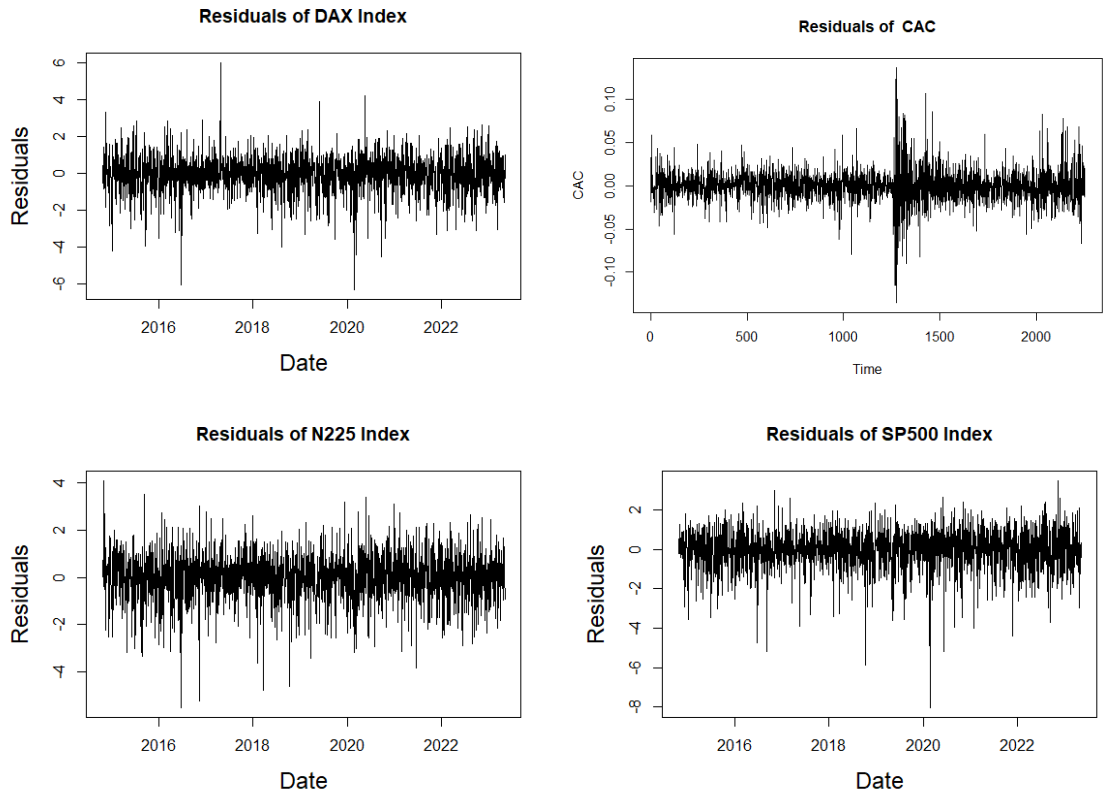


Figure 8.5: *Residuals of the fitted time series of the Log-returns of the closing price of the German index (DAX), French Index (CAC), Japanese (N225), and American (GSPC), computed in the period from* 23/10/2014 *to* 5/7/2024.

8.3.1 Residual Analysis

To assess the Goodness of Fit of the models selected in the previous Section, we perform some GoF analysis on the obtained residuals.

Staring from the ACF and QQ plots, shown in Figure 8.6. The very low ACF values, indicate a good fit, while the QQ-plot are not aligned with the normal line, suggesting that the residuals might not follow a normal distribution: this is a good indication that we need a more complex model to study the distribution of the innovations.

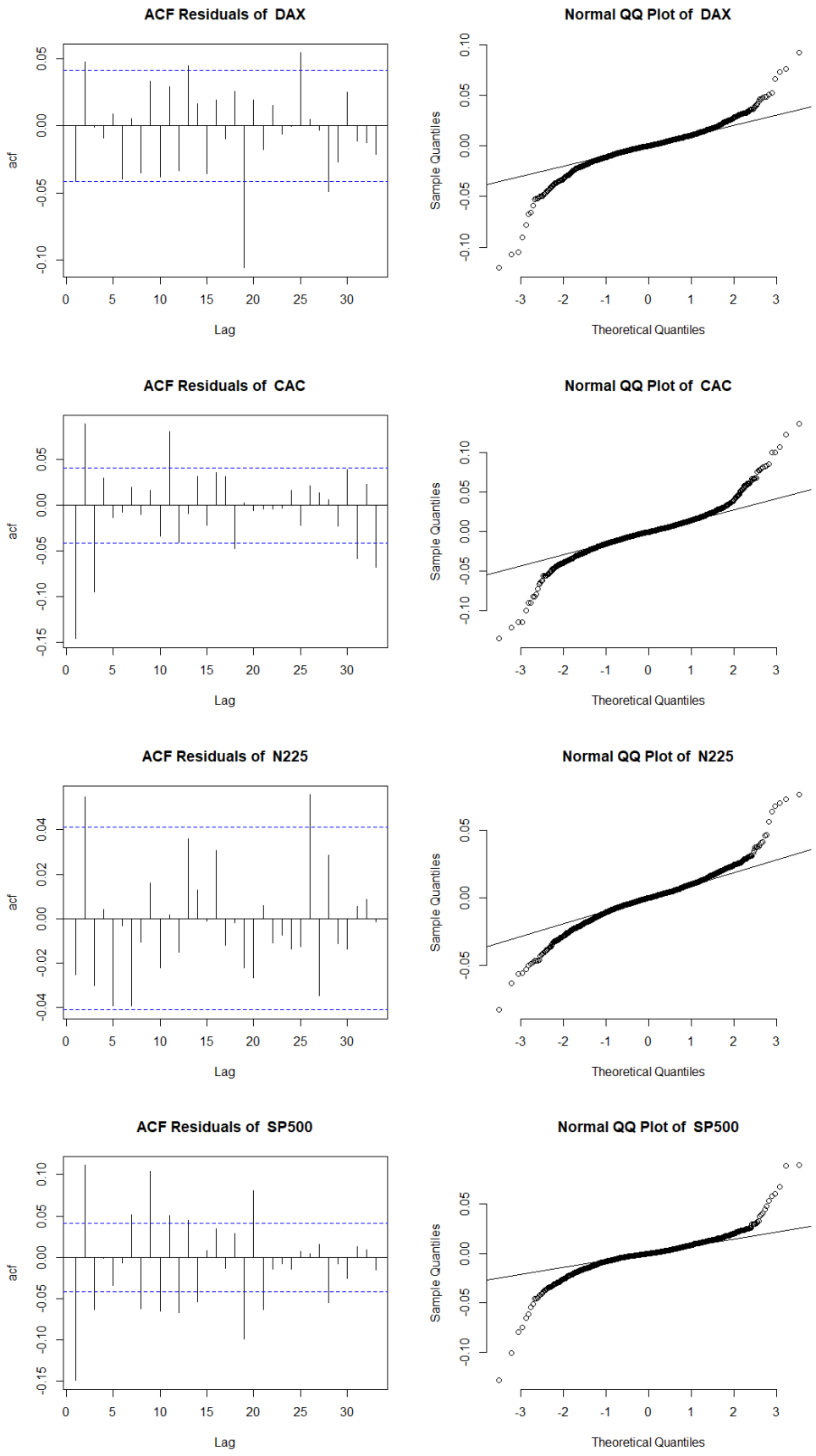


Figure 8.6: *ACF,and QQ-plot of the residuals of the fitted time series of the Log-returns of the closing price of the German index (DAX), French Index (CAC), Japanese (N225), and American (GSPC), computed in thof the two fitted AR(1)-GARCH(1,1) time pseriod from* 23/10/2014 *to* 5/7/2024.*es.*

Moreover, in Table 8.9, we provide the p-values of the Ljung-Box test, the ADF test, and the McLeod-Li test. The Ljung-Box test suggests that the residuals should be independent only for the SP500 case. The ADF result in a p-value of 0.01 for all variables, indicating stationarity. However, the McLeod-Li's p-values are smaller than the significance level, suggesting that the residuals could still exhibit heteroskedasticity.

The results obtained are in accordance to those of the simulation study of Section 6.7: while the ADF results might indicate a good fit, the McLeod-Li test result suggests that the fit might not be perfect. This confirms the need for further refinement to achieve a better fit for the original dataset. In this context, the application of copulas and vine models might be beneficial. With this observation in mind, we proceed to transform the dataset into a uniformly distributed one, and to fit a vine copula in the following Sections.

Index	Ljung-Box	Augmented Dickey-Fuller	McLeod-Li
DAX	4.42×10^{-12}	0.01	0
CAC	1.03×10^{-12}	0.01	0
N225	0	0.01	0
SP500	0.21	0.01	0

Table 8.9: *P-values of Augmented Dickey-Fuller to verify the stationarity of the residuals of the fitted time series, of the McLeod-Li for heteroskedasticity, and of the Ljung-Box test.*

8.4 Marginal Fitting

Lastly, we transform the residuals to uniform scale, by fitting a marginal distribution, and by applying the corresponding Cumulative Distribution Function (CDF).

The distribution of each marginal, shown in Figure 8.7, appear bell-shaped, rather symmetrical, but slightly heavy tailed. These are all indication that a t-Student might be an appropriate choice to fit to each marginal. We therefore proceed to fit a t-Student to each variable of the dataset, the results are provided in Table 8.10. For each variable, we also test the GoF by computing the p-value of the Kolmogorov-Smirnov test (see F.4.1). All p-values indicate that the t-student family cannot be rejected.

Index	Degree of Freedom	Mean	St Deviation	p-value KS
DAX	3.247	2.63×10^{-4}	0.009	0.275
CAC	3.196	-2.19×10^{-4}	0.013	0.900
N225	3.899	2.78×10^{-6}	0.009	0.563
SP500	2.914	-4.23×10^{-5}	0.007	0.053

Table 8.10: *Fitted t-Students distribution to each time series considered, and the corresponding p-value of the Kolmogorov-Smirov test, assessing the GoF.*

Additionally, to support the choice of a t-Student for the marginals we plotted the fitted distributions against the distribution of each marginal (in red) in Figure 8.7, where we can observe how the fitted models describe rather accurately the original distributions. Moreover, in Figure 8.9, we present the Q-Q plots for the t-Student distribution. The plots show that the data points closely follow the theoretical t-Student line, confirming the appropriateness of the t-Student distribution for modeling the marginals.

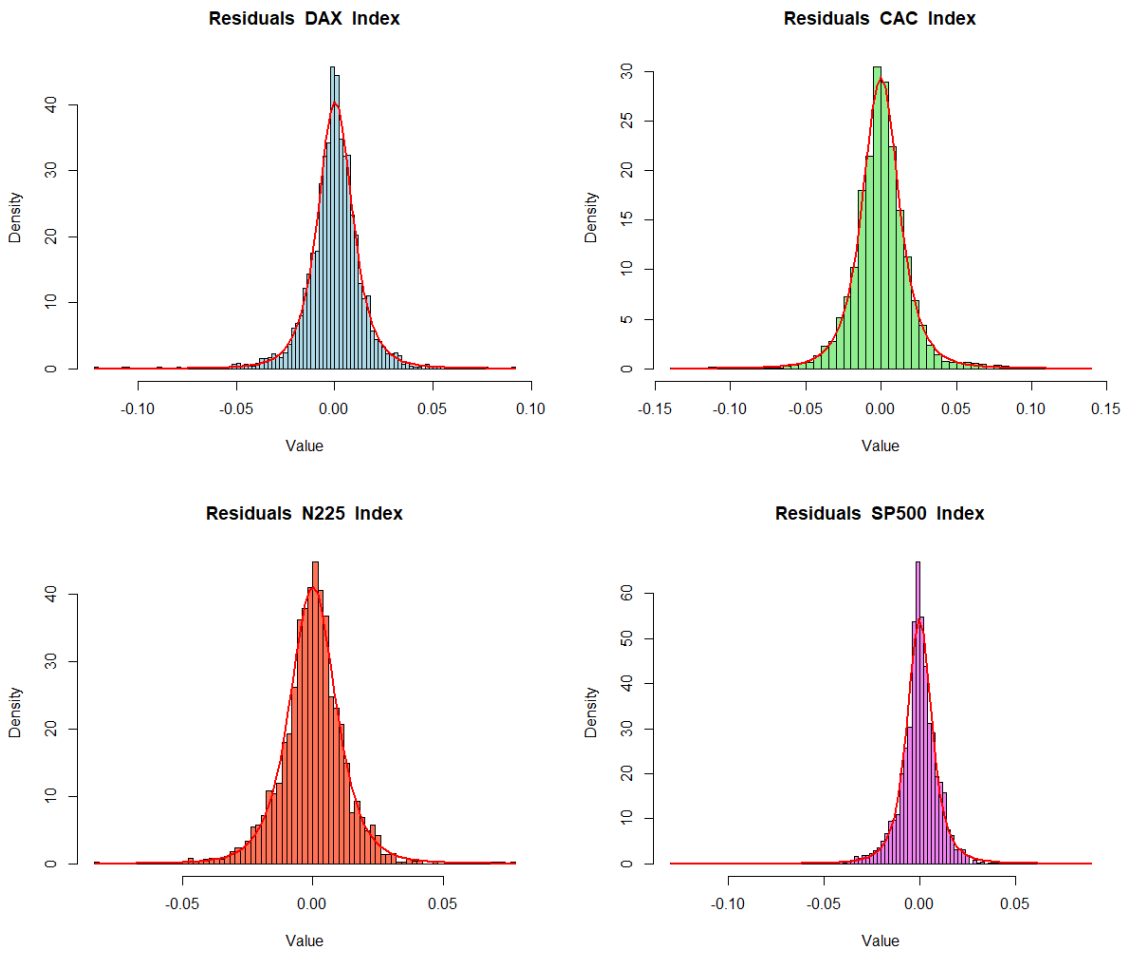


Figure 8.7: *Comparison of the fitted t-Student distributions (in red) with the empirical density distributions of each marginal, for each of the four index considered.*

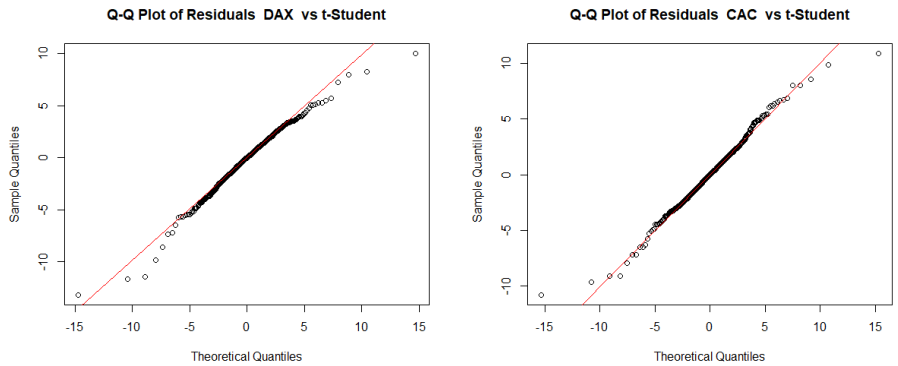


Figure 8.8: *Q-Q plots testing that the distribution of the residuals of the fitted time series models is a t-student.*

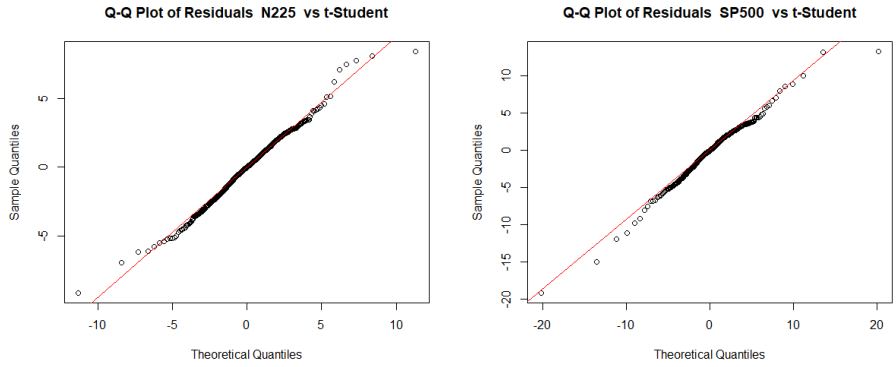


Figure 8.9: *Q-Q plots testing that the distribution of the residuals of the fitted time series models is a t-student.*

The fitted marginals are used to transform the data into uniformly distributed variables, by applying the CDF as $u_{t,i} = F_i(e_{t,i})$ for $t = 1, \ldots, T$ and $i = \{DAX, CAC, SP500, N225\}$. The obtained results are summarized in Figure 8.10.

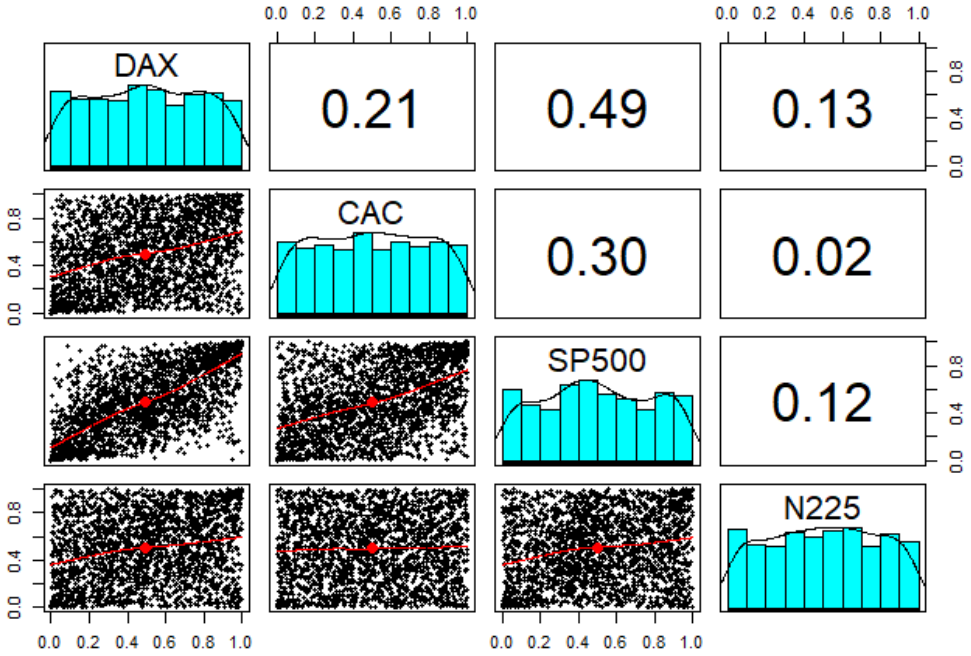


Figure 8.10: *Summary plot showing the histograms of the ranked data of each index, the scatterplot of each pair combination, and Kendall's tau correlation of the different pairs.*

Moreover, to check that each transformed variable is indeed uniformly distributed, a Kolmogorv-Smirnov test is performed. All the obtained p-values, which are reported in Table 8.11, are above the 0.05, indicating that all variables can be considered uniformly distributed, as we do not have enough evidence to reject the H_0 hypothesis. The dataset is now ready to be used to fit a vine model, as it will be discussed in the following Section.

Index	p-value KS
DAX	0.275
CAC	0.900
N225	0.563
SP500	0.144

Table 8.11: *P-value of the Kolmogorov-Smirov test, assessing that the transformed dataset have indeed a uniform distribution.*

8.5 Check Time Dependence

Additionally, we test the time dependence of the considered time series, by testing the third assumption discussed in Section 6.1.1 on the transformed dataset. By applying the decision tree presented in Section 6.6 used to identify relevant partitions by comparing the Kendall's taus, we are able to identify a set of 14 partitions, illustrated in Figure 8.11. The p-value of the test is equal to 0.02, suggesting that indeed we should consider time dependence when fitting a vine structure to the dataset.

Additionally, it is interesting to notice how period characterized by strong economic stress, such as 2018, or 2020, display partitions of short duration, meaning that the underlying structure changes rather quickly.

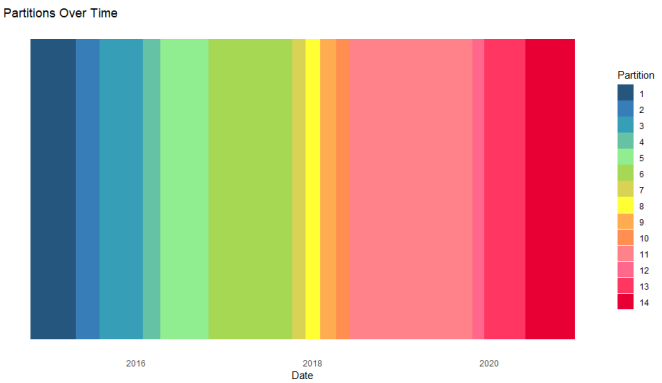


Figure 8.11: *Identified partitions, each indicating time periods with the same underlying vine copula model.*

However, in Section 4.6.1, we discussed the convergence of the estimators for Vine copulas and concluded that a sample size of $N = 500$ points is necessary to ensure a reliable model. The partitions identified depicted in Figure 8.11, sometimes contain only 50 points, which would be insufficient to produce a good model.

To address this, we re-ran the test with a constraint on the minimum length of each partition. The results are described in the following Subsection.

8.5.1 Updated Partitions

Although a minimum of $N = 500$ would be ideal, it is not always feasible. Therefore, we compromised with a minimum of 170 observations per partition. This number should be a sufficient sample size to ensure a satisfactory fit, while still allowing the identification of different partitions.

The new decision tree has a p-value of 0.03, which supports the time dependence of the underlying vine structure. 5 partitions are identified by the following decision tree, shown in Figure 8.12 along with the corresponding Kendall's tau value.

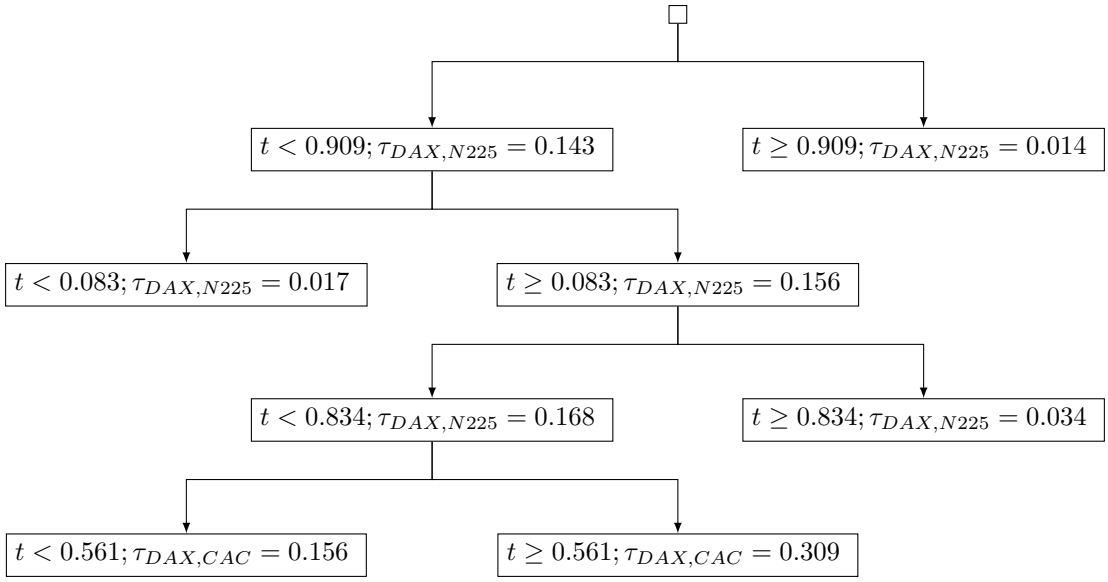


Figure 8.12: *Decision Tree identifying 5 partitions on the four dimensional dataset, where time t is identified by values ranging from 0 to 1, which represent the period between dates 24/10/2014 and 05/07/2024. The values of the Kendall's taus used to identify the partitions are also provided.*

The partitions identified are presented in Table D.1, and represented in Figure 8.13. It is noteworthy that the second and third partitions contain a substantial number of data points, which should ensure robust results when fitting a vine model. In contrast, the remaining partitions consist of fewer data points, potentially limiting the reliability of the models. Despite this, we proceed with fitting a copula-GARCH model to each partition. It is important to acknowledge these differences and consider their implications for the reliability of the results.

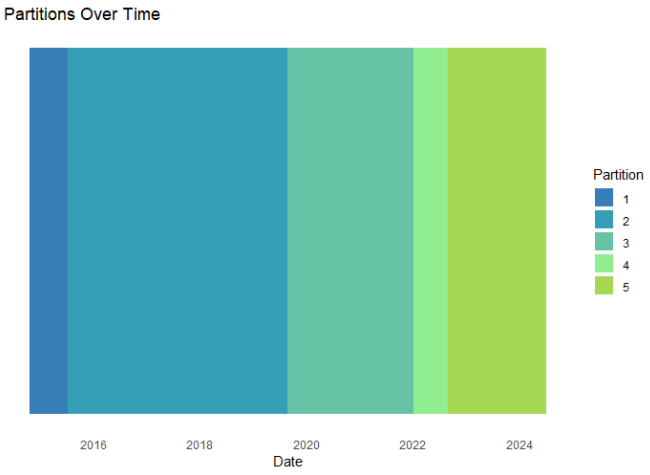


Figure 8.13: *Identified partitions, each indicating time periods with the same underlying vine copula model.*

Additionally, it is interesting to highlight in Table D.1 how some of the identified breaks between

the partitions reflect changes in the global economy and crises. In particular, the third partition starts in March 2020, which corresponds to the start of the COVID-19 pandemic, period that signified a significant shock wave in the economy. Moreover, October 2022, which marks the beginning of the fourth partition, coincides with the escalation of the war in Ukraine, which had a profound impact on global markets, particularly in energy and agricultural sectors, leading to increased volatility and economic uncertainty [77]. In addition, August 2015, where the break between partitions 1 and 2 occur, could be associated with the start of China's stock market turbulence; while June 2023 (start of the last partition) is close to the [78].

	Dates	Length
Partition 1	24/10/2014 - 10/08/2015	187
Partition 2	11/08/2015 - 03/03/2020	1076
Partition 3	04/03/2020 - 11/10/2022	615
Partition 4	12/10/2022 - 29/06/2023	170
Partition 5	30/06/2023 - 05/07/2024	241

Table 8.12: *Partitions of the dataset.*

Unfortunately, the results obtained by applying the copula-GARCH model to the partitions considered, which are reported and discussed in Appendix D, highlighted the importance of having a dataset with a sufficient number of points. In fact, a larger dataset ensures more accurate partitioning and stable estimation of the copula parameters, leading to more reliable results and conclusions, while the partitions with fewer data (1 and 4) led to poorer results. For this reason, we decided to re-run the analysis to a new set of customized partitions, which are discussed in the following Subsection.

8.5.2 Final Partitions

As already anticipated, the partitions discussed above have financial sense, as the identified breaks correspond to significant changes in the global economy. However, some partitions resulted in too few data points, leading to unsatisfactory results. To address this issue, we created a new set of customized partitions, ensuring a sufficient number of data points for each partition, while keeping the financial relevance of the break points.

We decided to keep the partition starting in 2020, which aligns with the outbreak of the COVID-19 pandemic. Additionally, we introduced a new break corresponding to the 2018 market crash, a significant moment for the global economy, which seemed to be particularly relevant in the identification of the first set of partitions (see Figure 8.11).

We therefore ended up with three separate partitions, illustrated in Figure 8.14 and detailed in Table 8.13. We also tested the relevance of the partitions by testing the Simplifying Assumption on the identified partitions (see Section 6.6), which resulted in a p-value of 0.03, which indicates that the underlying vine models should still be considered time dependant. The results obtained from these partitions will be discussed in the following Sections.

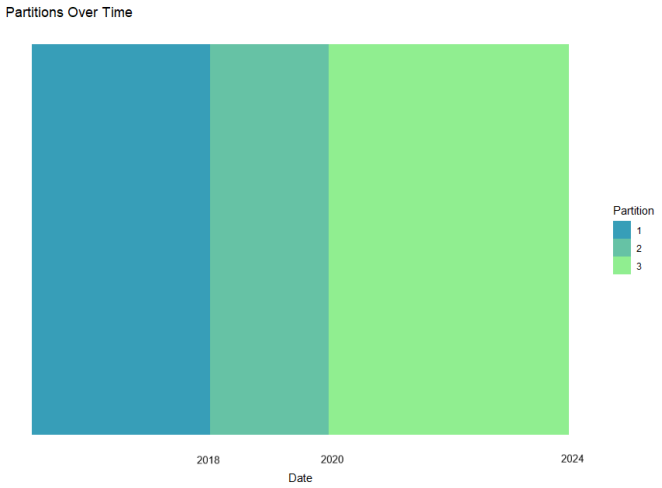


Figure 8.14: *Identified partitions, each indicating time periods with the same underlying vine copula model.*

	Dates	Length
Partition 1	24/10/2014 - 05/01/2018	757
Partition 2	06/01/2018 - 03/03/2020	505
Partition 3	04/03/2020 - 05/07/2024	1027

Table 8.13: *Partitions of the dataset.*

8.6 Portfolio Optimization

Having divided the dataset into different partitions, we will now proceed to test the procedure developed in this thesis for each of them. The process begins by splitting each partition into a training set and a test set. The length of the training set will vary depending on the partition: the priority is to guarantee a sufficient amount of points on the training set to provide a good fit (around 500, as it was already discussed in the previous Sections). We then aim at having at least $N = 30$ days on the test set, to have enough material to evaluate the performance of each portfolio. For larger partitions, we will use an even greater proportions of the data for testing (around 20 to 30 percent of the points).

Subsequently, we will fit a vine model to each training set, evaluate the goodness of fit of the model, and use the fitted model to determine the weights of the optimal portfolio with the methodology described in Section 7.1. This involves simulating from the fitted model $Q = 1000$ points, and optimizing the portfolio through a Monte Carlo allocation method. The obtained weights will be applied to the original dataset to compute key metrics (Mean, Variance, CVaR, Sharpe ratio, STARR ratio). The chosen weights will then be applied to the log-returns of the test set to evaluate the performance of the portfolios out-of-sample. Similarly to the example proposed in Section 7.3, the cumulative log-returns, the Conditional Value at Risks and the STARR ratios will be evaluated and compared to those of an equal weight portfolio and of the tangent portfolio of the Markowitz model obtained optimized the training set (refer to Section 2.3.1). Once again, we will consider the copula-GARCH successful if able to outperform the others in therm of STARR ratio.

The result will be discussed separately for each partition in the following Subsections.

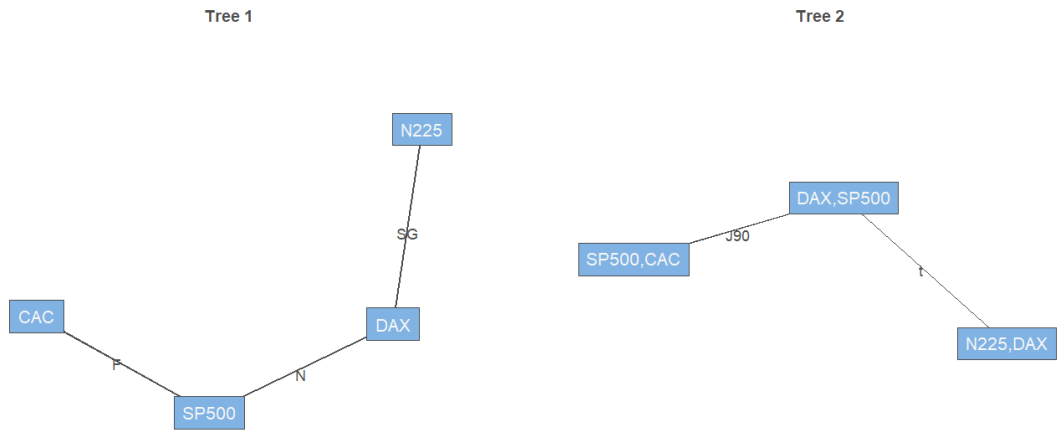
8.6.1 Partition 1

The first partition, is characterized by $N = 757$ data points, as it covers the period between 24/10/2014 and 05/01/2018. The amount of data should be sufficient to guarantee a satisfactory fit for the vine model. For our analysis, the dataset is split into a training set and a test set, the first consisting of $N = 623$ points (around 80% of the dataset), and the second of $N = 134$ points.

8.6.1.1 Evaluate the Portfolio

Firstly, we analyze the overall performance of the vine copula model on the train set, constructed by fitting a proper vine copula model to the transformed residuals of the train set of the first partition. The result is represented in Figure 8.15, and the corresponding parameters are given in Table 8.14. We can notice how the Kendall's tau of the bivariate copulas of the first tree are characterized by moderate positive relationship, which tends to disappear in the following trees, till we reach independence in the last tree. In addition, the highest τ value in the one between the German and the American indexes which, in the period before Wall Street stock market crash, were two of the strongest economies.

The goodness of fit of the vine is assessed using the White test discussed in Section 4.5.2, obtaining a p-value of .225 which indicates that we cannot reject the null hypothesis, and hence that we should consider it to be a good fit. In addition, we test for the simplifying assumption (refer to Definition 14), obtaining that it cannot be rejected for any of the bivariate copulas considered.



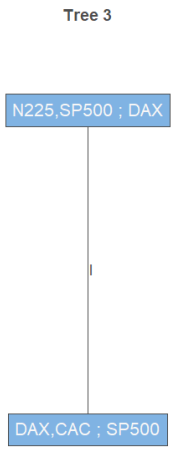


Figure 8.15: *Vine structure of the fitted Vine models of the Residuals, transformed using their empirical CDF.*

Tree	Edge	Distribution	Parameters
T_1	CAC, SP500	Frank	par $= 3.99$, $\tau = 0.39$
	DAX, SP500	Normal	par $= 0.73$, $\tau = 0.52$
	N225, DAX	Survival Gumbel	par $= 1.14$, $\tau = 0.13$
T_2	DAX, CAC; SP500	Joe 90	par $= -1.06$, $\tau = -0.03$
	N225, SP500; DAX	t Student	$\rho = 0.10$, $\nu = 4.83$, $\tau = 0.06$
T_3	N225, CAC; DAX, SP500	Independent	$\tau = 0$

Table 8.14: *Parameters of the vine copula model fitted on the first partition at* $T = 14/06/2017$.

The fitted vine copula discussed above is used to simulate $Q = 1000$ points, which are then transformed using the appropriate marginal distributions (the ones reported in Table 8.10), and used to construct the copula-GARCH model's mean-CVaR efficient frontier. The optimal portfolio, found by optimizing the STARR ratio among the frontier's points' has weights:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 64.35\%, 0, 35.65\%).$$

The results are compared with the optimal mean-variance portfolio, obtained by selecting the weights that guarantee the best Sharpe ratio among the ones of the Markowitz frontier (refer to Section 2.3.1.1), which corresponds to:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 71.33\%, 28.67\%, 0).$$

The portfolios obtained for the two methods are compared in Table 8.15, with an equal weight portfolio. In particular, we compare the portfolio's weights, noticing how both prefer investing the biggest percentage in the French index (CAC). The vine copula approach selects the Japanese index (N225) as the second preference, Markowitz's prefers to invest in the American SP500. Each set of weights is then applied to the train set of the original dataset, to compare the different performance metrics.

Markowitz' portfolio showcases a greater risk appetite, reflected by a higher CVaR value and higher mean return. The copula-GARCH portfolio has slightly smaller mean value, but also lower risk, re-

sulting in the best STARR ratio among the portfolios. This suggests a promising performance for out-of-sample evaluation. Lastly, the Equal Weight portfolio is characterized by more conservative mean and risk, which result in the best Sharpe ratio.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(25\%, 25\%, 25\%, 25\%)$	0.48	21.00	**0.61**%	2.27%
Mean-Var	$(0, 71.33\%, 28.67\%, 0)$	**0.71**	26.84	0.53%	2.65%
copula-GARCH	$(0, 64.35\%, 0, 35.65\%)$	0.70	25.65	0.58%	**2.75**%

Table 8.15: *Table comparing the portfolios obtained by optimizing the weights of the four indexes, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the partition of $N = 623$ points of the first partition.*

8.6.1.2 Out of Sample Evaluation

We continue with an out-of-sample efficiency evaluation of the portfolios considered. We apply the selected weights from each method to the test set, composed of the last $N = 134$ points of the partition. At each point, computing the cumulative returns, shown in Figure 8.16, the Conditional Value at Risk of Figure 8.18 and the STARR ratio in Figure 8.19. Moreover, in Figures 8.17 and 8.20, we provide the differences of Cumulative returns and STARR Ratio respectively.

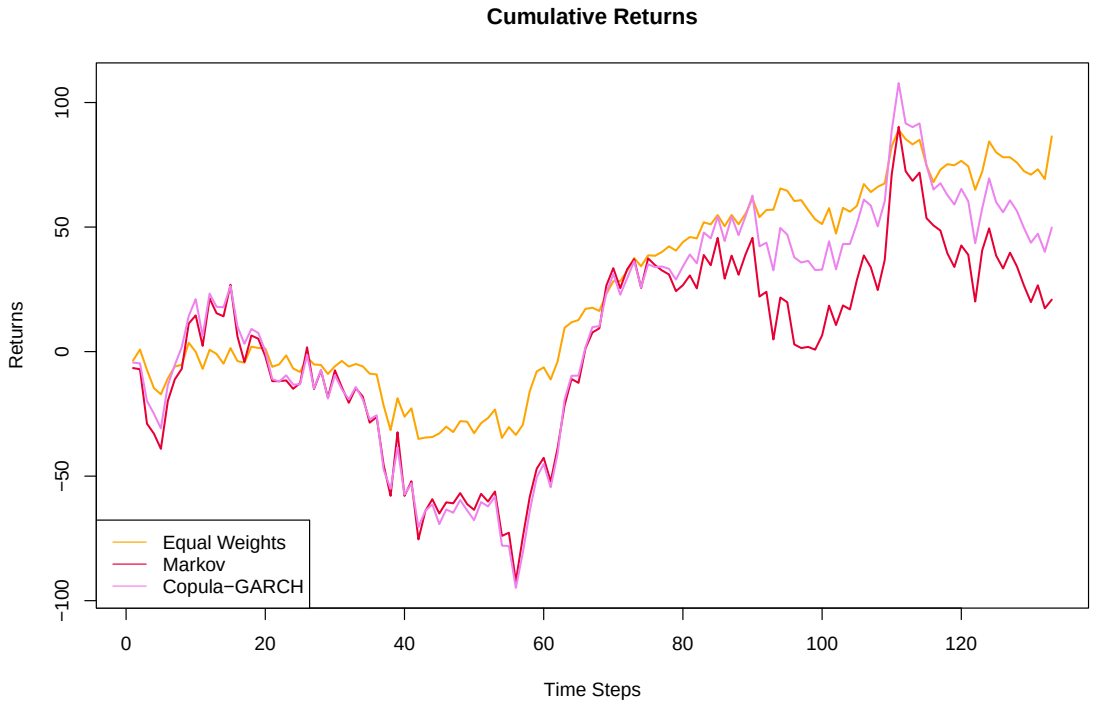


Figure 8.16: *Evolution of the Cumulative Returns obtained by applying the optimal weights to the test set of $N = 134$ points.*

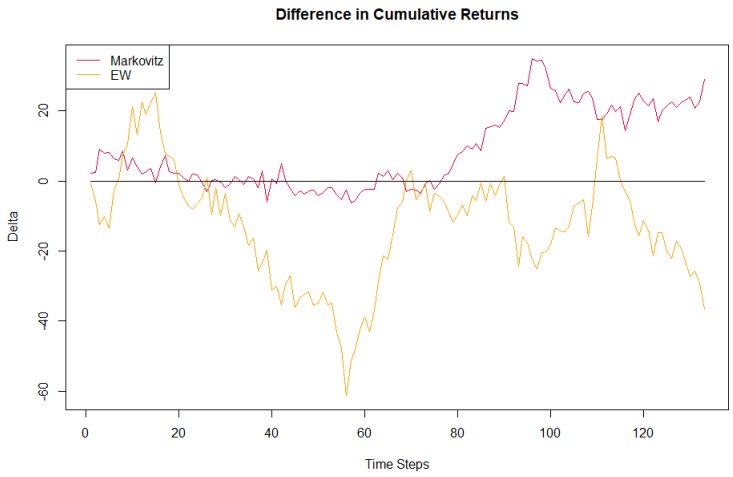


Figure 8.17: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one on the set of the first partition.*

In terms of cumulative returns, it is hard to determine the best-performing portfolio, as all portfolios exhibit similar behavior. However, by examining the differences illustrated in Figure 8.17, it becomes clearer that the copula-GARCH portfolio outperforms Markowitz's portfolio. The Equal Weights portfolio generally shows a higher cumulative return but is occasionally outperformed by the copula-GARCH portfolio.

Given that our goal is to achieve a good balance between return and risk, evaluating the portfolio's efficacy requires more than just analyzing returns. Thus, we also consider other performance metrics.

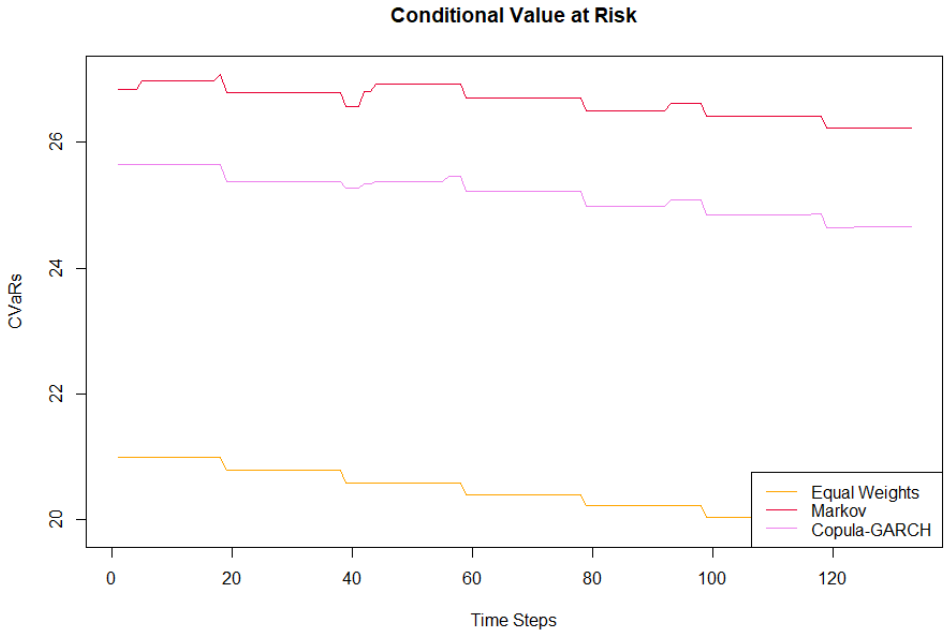


Figure 8.18: *Evolution of the Conditional Value at Risk obtained by applying the optimal weights to the test set of* $N = 134$ *points.*

From the CVaR plots shown in Figure 8.18, we can see how coherently with the previous observations, the mean-variance portfolio remains the riskiest, with a Conditional Value at Risk (CVaR) that remains above 26. In contrast, the Equal Weight (EW) portfolio appears more conservative, showcasing the lowest CVaR value across the entire period considered.

We conclude this analysis by examining the performance of each portfolio in terms of STARR ratios, as shown in Figure 8.19. The differences reveal that the copula-GARCH portfolio consistently achieves the best ratio throughout the entire test set, indicating that the portfolio successfully meets its goal and maintains its superiority over all 134 points considered. On the other hand, it is interesting to notice the switch between the other two portfolios: for the first 90 days, Markowitz's portfolio holds the second-best position, which is subsequently taken over by the EW portfolio.

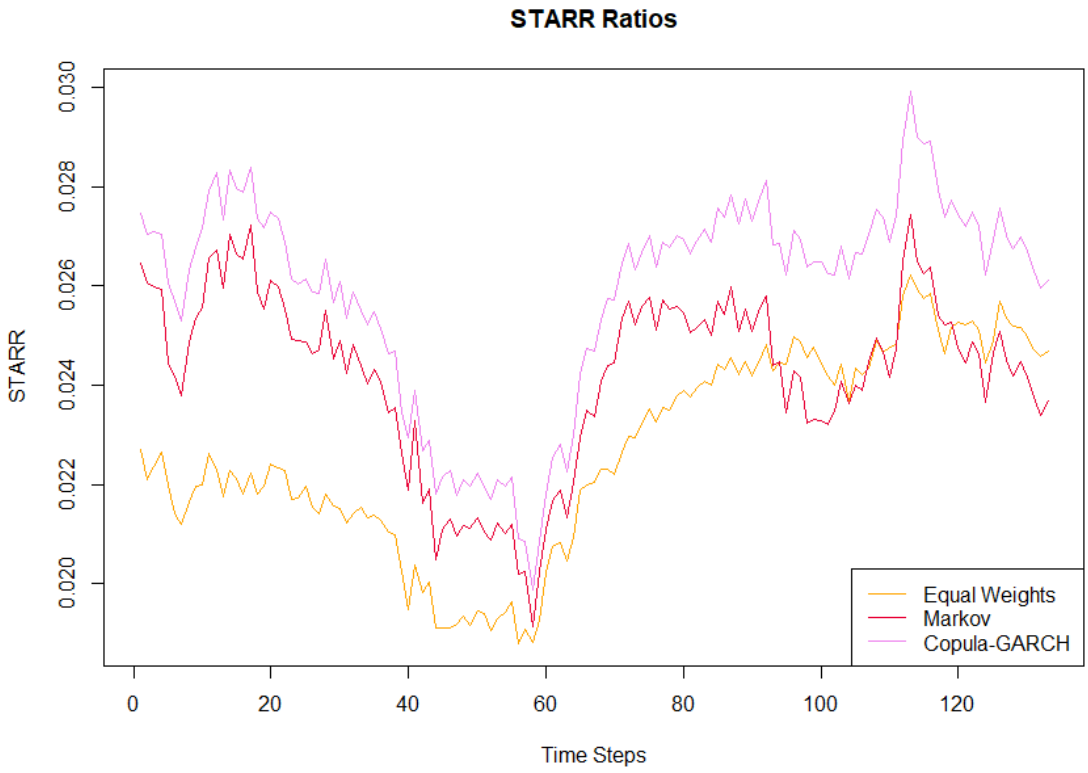


Figure 8.19: *Evolution of the STARR Ratios obtained by applying the optimal weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one to the test set of $N = 134$ points.*

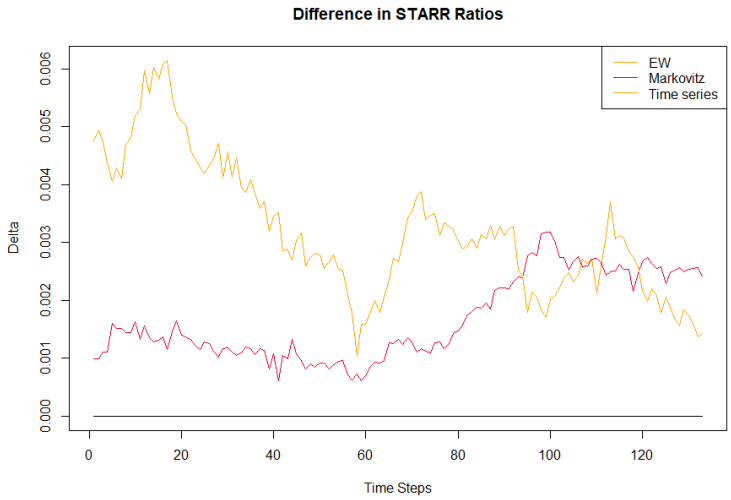


Figure 8.20: *Difference between the STARR ratios obtained from the weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one on the set of the first partition.*

Lastly, to confirm the out-performance of the copula-GARCH portfolio, we compare the considered metrics at time $T = 757$, the last day of the test set, in Table 8.16. The copula-GARCH indeed holds the best STARR ratio, with approximately a 4.3%. improvement over the second best, which would be the Equal Weight portfolio. Markowit's portfolio appears to have the lowest performance, showing low cumulative return, higher risk, and the lowest STARR ratio among those considered. The Equal Weight portfolio performs rather well, with the highest cumulative return and relatively low risk. However, this performance, derived from the stability and diversification (being the only one investing in all four stocks) of the EW portfolio, will not be consistent in the following partitions.

Method	Cumulative Return	CVaR	STARR
EW	**86.40**	19.86	2.44%
Mean-Var	20.86	26.23	2.32%
copula-GARCH	49.79	24.65	**2.57**%

Table 8.16: *Cumulative returns, CVaRs and STARR ratios for each method at the last day of the test set, hence at $T = 05/01/2018$.*

8.6.2 Partition 2

The second partition contains $N = 505$ observations, covering the period from 06/01/2018 to 03/03/2020. To ensure a good model fit, given that the number of observations is marginally sufficient, the dataset is divided into the following sets: $N = 475$ days for the training set, and $N = 30$ days for the test set. This choice guarantees the around 500 observations needed to fit a good vine model, and enough points to evaluate the model's performance.

8.6.2.1 Evaluate the Portfolio

In this Section we will discuss the portfolios obtained by fitting a model to the entire partition 2. The resulting weights will then be applied on an out of sample test set.

The trees corresponding to the C-vine fitted on the transformed residuals of the second partition are

represented in Figure 8.21, while the corresponding parameters are given in Table 8.17. The Kendall's tau on the first tree reveal a moderate positive dependence in all bivariate copulas. In contrast, the conditional trees show relatively low Kendall's tau values, indicating weaker conditional dependence. The nodes in the final tree are independent. These results are similar to those from the previous partition, however the Kendall's tau values appear decreased in the second partition, indicating weaker dependence.

Additionally, a p-value of 0.975 of White's Goodness of Fit test indicates that the fit should be good; moreover, testing the simplifying assumption resulted in not enough evidence to reject the assumption, which allows for a much more simple model.

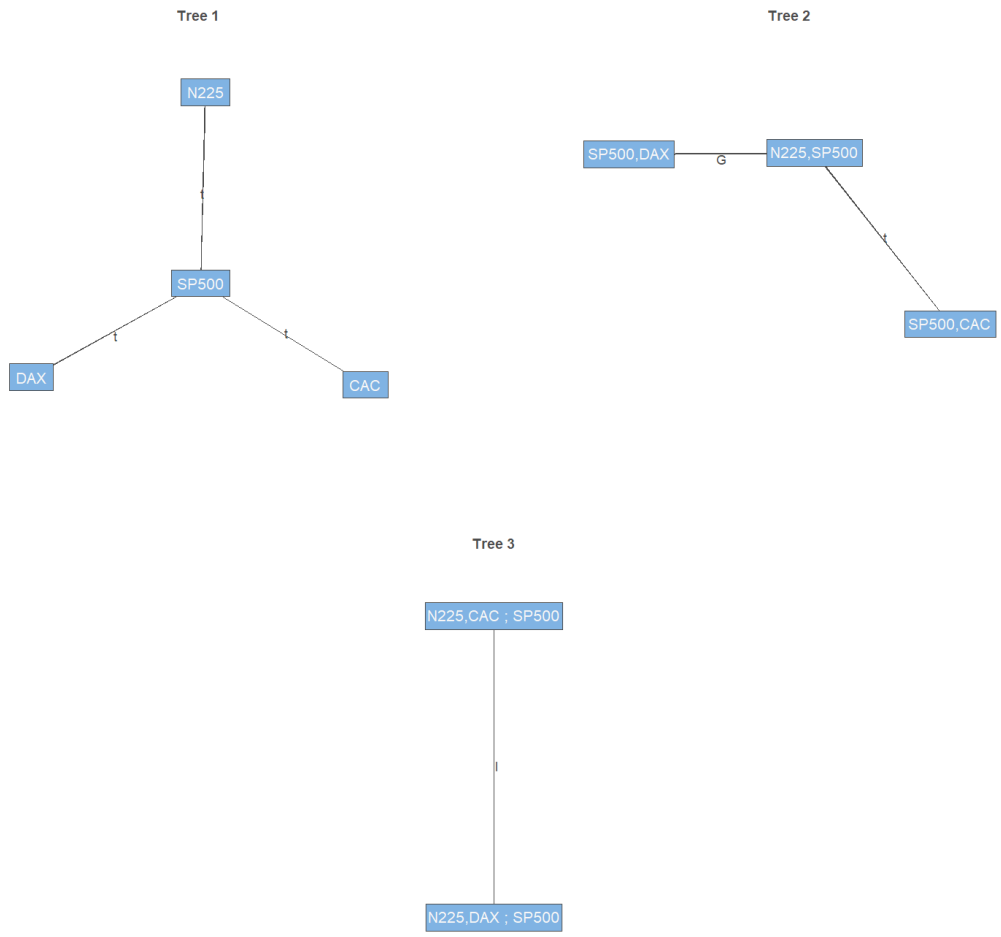


Figure 8.21: *Vine structure of the fitted Vine models of the Residuals, transformed using their empirical CDF.*

Tree	Edge	Distribution	Parameters
T_1	SP500, DAX	t Student	$\rho = 0.64$, $\nu = 2.55$, $\tau = 0.45$
	SP500, CAC	t Student	$\rho = 0.51$, $\nu = 3.29$, $\tau = 0.34$
	N225, SP500	t Student	$\rho = 0.32$, $\nu = 3.44$, $\tau = 0.20$
T_2	N225, DAX; SP500	Gumbel	par $= 1.10$, $\tau = 0.09$
	N225, CAC; SP500	t Student	$\rho = -0.10$, $\nu = 7.21$, $\tau = -0.06$
T_3	CAC, DAX; N225, SP500	Independent	$\tau = 0$

Table 8.17: *Parameters of the Vine copula fitted on the second partition at* $T = 14/01/2020$.

The fitted vine copula is used to simulate $Q = 1000$ points, which are then transformed and used to construct the efficient frontier for the copula-GARCH model. The optimal portfolio has weights:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 0, 78.91\%, 21.09\%).$$

Indicating that the optimal portfolio should be composed solely of the American and the Japanese indexes, avoiding the European ones.

The results are then compared with the optimal mean-variance portfolio, which corresponds to:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 1.41\%, 66.11\%, 32.48\%).$$

This portfolio aligns with the copula-GARCH model in terms of the selected indexes, with the minor addition of the French index, to which only 1.41% is allocated. The percentages are slightly different, but maintain a similar balance, suggesting that we can expect comparable results from both portfolios.

More in detail, we compare the two portfolios with an equal weights one in Table 8.18, where it is clear how the vine copula portfolio outperforms the other two methods considered. In fact, both the Sharpe ratio (5.78% better than Markowitz') and the STARR ratio (5.37% better than Markowitz') appear to be greater than the other two, confirming the positive performance of our method for the partition considered.

The vine copula portfolio appears to be more risky than the other ones, but also characterized by greater expected value. On the other hand, the equal weight portfolio showcases a very poor performance, as it has the lowest expected value, but also a higher risk (CVaR) that the mean-variance portfolio: this is a sign that the diversified and balanced equilibrium achieved by an equal weight might perform well (as seen in the previous partition), but can also be extremely unreliable, as we can conclude from this example.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(25\%, 25\%, 25\%, 25\%)$	0.092	20.046	0.16%	0.46%
Mean-Var	$(0, 1.4\%, 66.1\%, 32.5\%)$	0.278	19.636	0.49%	1.41%
copula-GARCH	$(0, 0, 78.9\%, 21.1\%)$	**0.320**	21.470	**0.52**%	**1.49**%

Table 8.18: *Table comparing the portfolios obtained by optimizing the weights of the four indexes, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the partition of* $N = 475$ *points of the second partition.*

8.6.2.2 Out of Sample Evaluation

We continue with an out-of-sample efficiency evaluation of the portfolios considered. We apply the selected weights from each method to the $N = 30$ points of the test set. We start by computing the cumulative returns plotted in Figure 8.22.

Figure 8.22: *Evolution of the Cumulative Returns obtained by applying the optimal weights to the test set of $N = 30$ points of the second partition.*

Figure 8.23: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one on the set of the second partition.*

In Figure 8.23, the differences between the cumulative returns are illustrated, showing that the vine-copula-GARCH portfolio outperforms the other two portfolios for most of the dataset, with only

few exceptions. This result is further validated by Table 8.19, where the cumulative returns of our portfolio at the end of the test set appear grater than those of the other two portfolios considered. This result is a confirmation that using a vine copula to model the underlying structure provides a long-term advantage. By understanding the relationships among the variables, the vine-copula-GARCH approach enables to achieve better results over extended periods, rather than short-term gains. We will now proceed to analyze the CVaR and the STARR ratio performance as well.

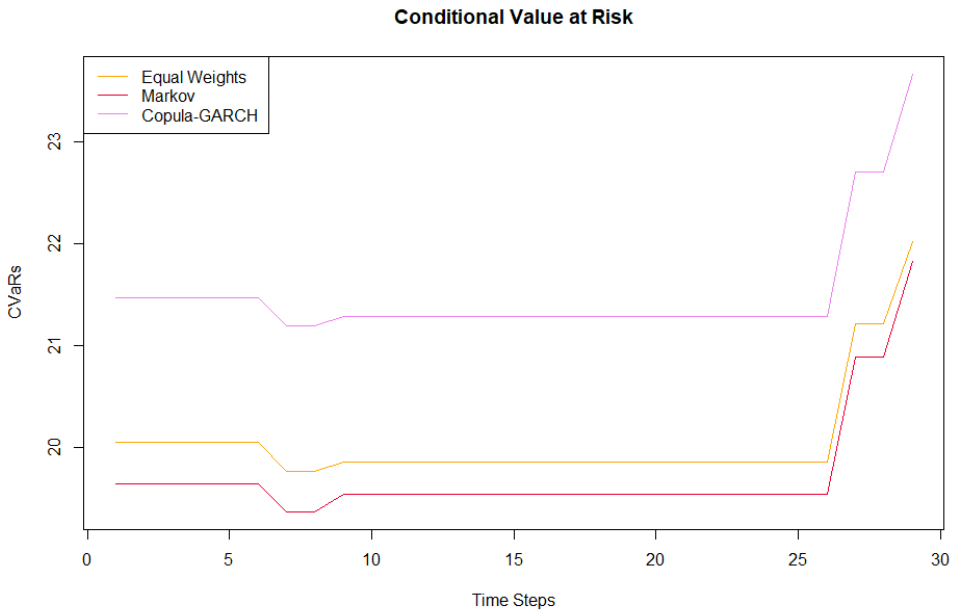


Figure 8.24: *Evolution of the Conditional Value at Risk obtained by applying the optimal weights to the test set of $N = 30$ points of the second partition.*

As already seen in Table 8.18, Figure 8.24 shows that the copula-GARCH portfolio is the riskiest, exhibiting the highest CVaR value throughout the entire test set. On the other hand, the Equal Weight portfolio, which was characterized by the lowest cumulative returns, appears to be riskier than the mean-variance one, confirming its poor performance in this partition.

Lastly, we compare the STARR ratios in Figure 8.25, and the corresponding differences in Figure 8.26. From both plots, it is clear that our portfolio outperforms the other two for the entire test set, displaying the highest ratio value. There is a drop in the last few days where the EW portfolio, which has a low but stable ratio value, nearly catches up. It is also interesting to note the similar behaviour of the copula-GARCH portfolio and Markowitz's one. Their ratios have close values and maintain a similar trend, particularly evident in the drop that does not affect the more balanced EW portfolio. This observation is coherent with expectations, as we previously discussed how the two portfolios had similar compositions.

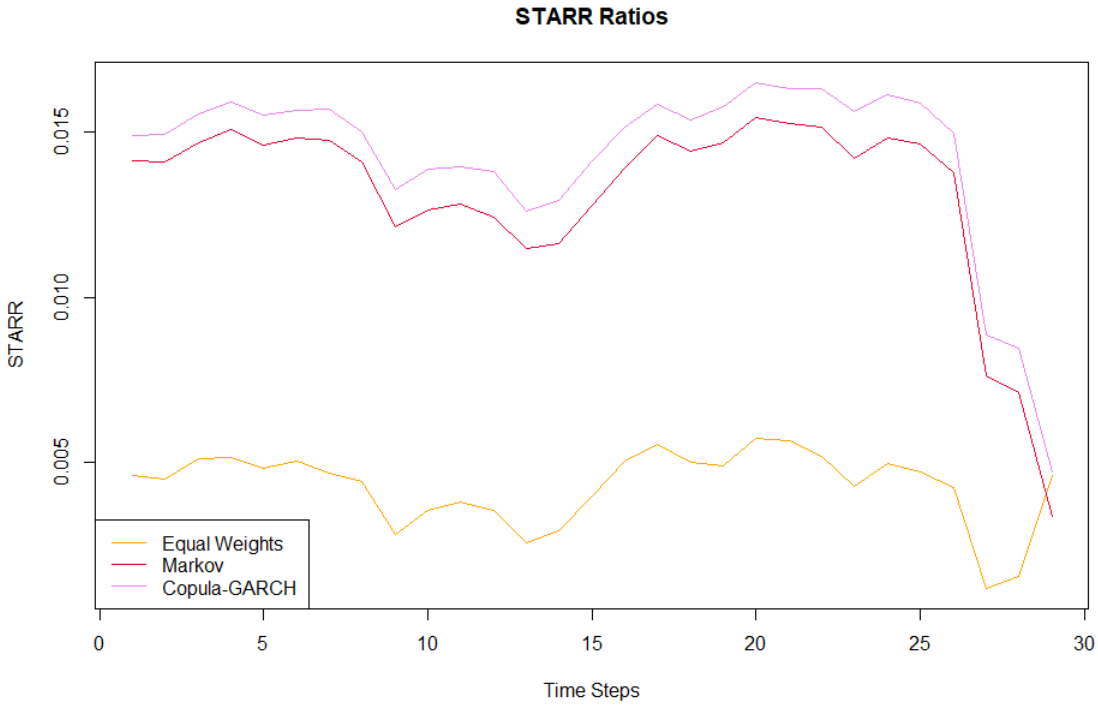


Figure 8.25: *Evolution of the STARR Ratios obtained by applying the optimal weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one to the test set of $N = 30$ points.*

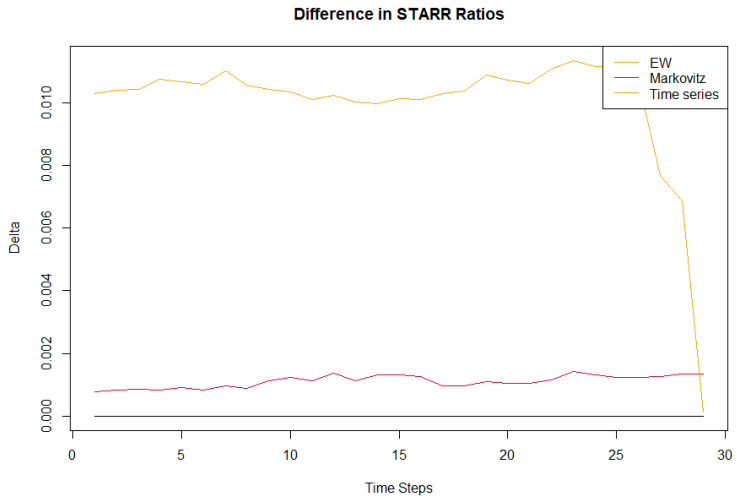


Figure 8.26: *Difference between the STARR Ratios obtained from the weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one on the set of the second partition.*

To confirm the observations previously stated, we provide the cumulative returns, CVaR values and STARR ratios of each portfolio on the test set at time $T = 30$ in Table 8.19, from which we can conclude that the vine copula-GARCH portfolio clearly outperforms the other ones, as it has higher returns, but also the highest STARR ratio.

Method	Cumulative Return	CVaR	STARR
EW	-115.1	22.02	0.46%
Mean-Var	-106.8	21.82	0.34%
copula-GARCH	**-104.4**	23.67	**0.47**%

Table 8.19: *Cumulative returns, CVaRs and STARR ratios for each method at the last day of the test set, hence at* $T = 03/03/2020$.

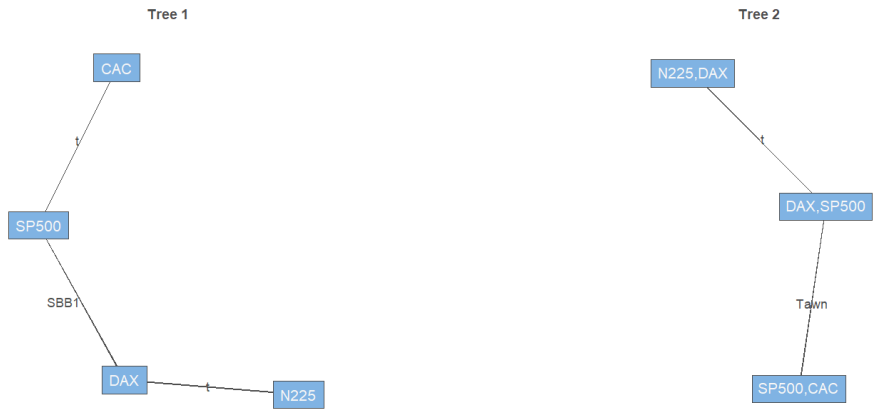
8.6.3 Partition 3

The third partition, is again characterized by a sufficient amount of point to guarantee a satisfactory fit. It is composed of $N = 1027$ observations, from 04/03/2020 to 05/07/2024, and the dataset is split in 731 days for the training set, and 296 for the test set, which correspond to around 30% of the partition.

8.6.3.1 Evaluate the Portfolio

The vine fitted on the third partition is characterized by the structure shown if Figure 8.27, and with the parameters of Table **??**. The fitted model obtains a p-value equal to 0.27 on the White test, which suggests a goodness of fit. Moreover, the simplifying assumption is not rejected for any copula considered.

From the Kendall's taus reported in table **??** we can notice that, similarly to the previous two partitions, we have a moderate positive correlation among the variables that compose the first tree, while the correlations in the second tree is almost zero, suggesting that there is not a strong conditional relationship among the different geographies. Lastly, the last tree, differently from the one fitted on the first partition, does not display independence, however the value of Kendall's tau is still extremely low.



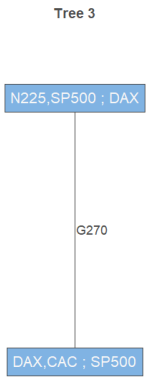


Figure 8.27: *Vine structure of the fitted Vine models of the Residuals, transformed using their empirical CDF.*

The fitted vine copula discussed above is used to simulate $Q = 1000$ points, which are then transformed and used to construct the mean-CVaR efficient frontier of the copula-GARCH case. The obtained optimal portfolio has weights:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 0, 31.22\%, 68.78\%).$$

The results are compared with the optimal mean-variance portfolio obtained by selecting the weights that guarantee the best Sharpe ratio, which corresponds to:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 0, 43.56\%, 56.44\%).$$

Both portfolios, as for the previous partition analysed, invest only in the American and Japanese indexes. However, this time the preference is shifted to the Japanese one for both portfolios. Lastly, as for the previous section, the percentages obtained from the two methods are rather close, so we expect them to have similar performances.

From Table 8.20, we observe that the equal weight portfolio is again extremely inefficient, with the lowest mean value and the highest risk. The mean-variance portfolio, despite being the most risk-exposed with the highest mean and highest CVaR, achieves the best Sharpe ratio, showing only a 1.48 % increase over the copula-GARCH portfolio. On the other hand, the copula-GARCH portfolio holds the best STARR ratio, 0.7% greater than Markowitz's portfolio.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(25\%, 25\%, 25\%, 25\%)$	0.165	33.297	0.08 %	0.49%
Mean-Var	$(0, 0, 43.56\%, 56.44\%)$	**0.371**	27.217	**0.270** %	1.36%
copula-GARCH	$(0, 0, 31.22\%, 68.78\%)$	0.368	26.830	0.266%	**1.37%**

Table 8.20: *Table comparing the portfolios obtained by optimizing the weights of the four indexes, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the partition of $N = 731$ points of the third partition.*

8.6.3.2 Out of Sample Evaluation

It is left to evaluate the performance of the vine copula fit on an out of sample dataset. We apply the weights obtained to $N = 300$ points of the test set, and compare the different metrics obtained with the other two method.

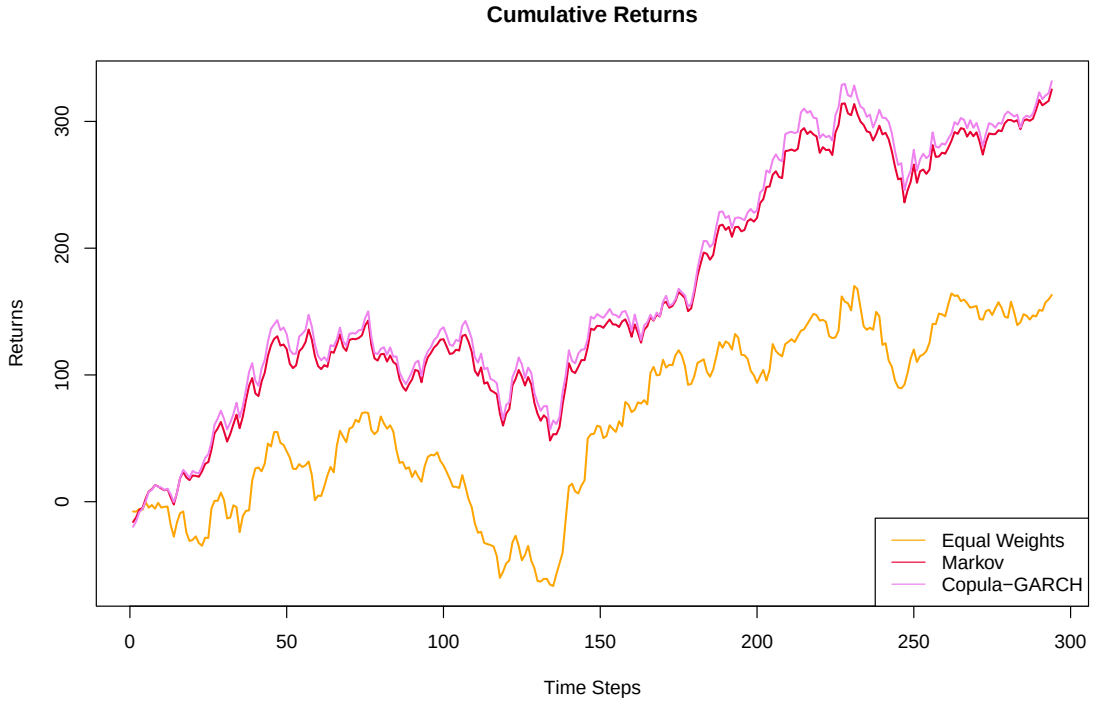


Figure 8.28: *Evolution of the Cumulative Returns obtained by applying the optimal weights to the test set of $N = 296$ points of the third partition.*

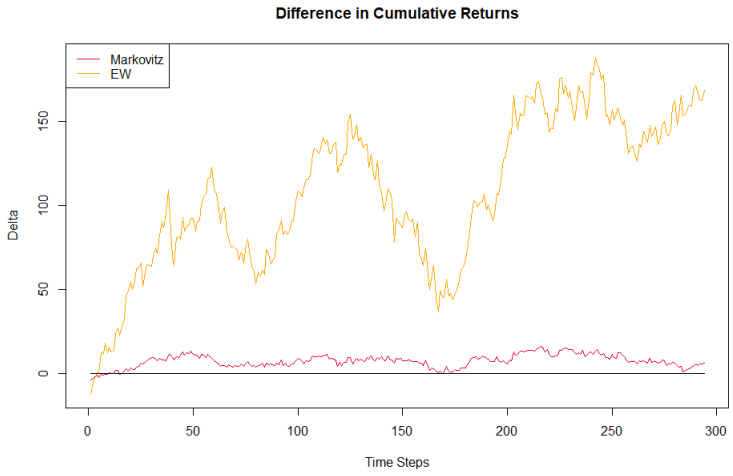


Figure 8.29: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one on the set of the third partition.*

In Figure 8.28 the cumulative returns of the three portfolios are compared. Both from this plot and

the corresponding differences in Figure 8.29, we can see how the vine copula-GARCH portfolio is able to maintain a higher return throughout the entire $N = 296$ points that compose the dataset, outperforming the other two, as the initial lower value is rapidly overcome. These results further confirm that using a vine copula to model the underlying structure provides a long-term advantage. By understanding the relationships among the variables, the vine-copula-GARCH approach enables to achieve better results over extended periods, rather than short-term gains. Moreover, coherently with expectations, the mean-var portfolio showcases a behaviour that is extremely similar to the copula one. On the contrary, the EW portfolio has a significantly worse performance.

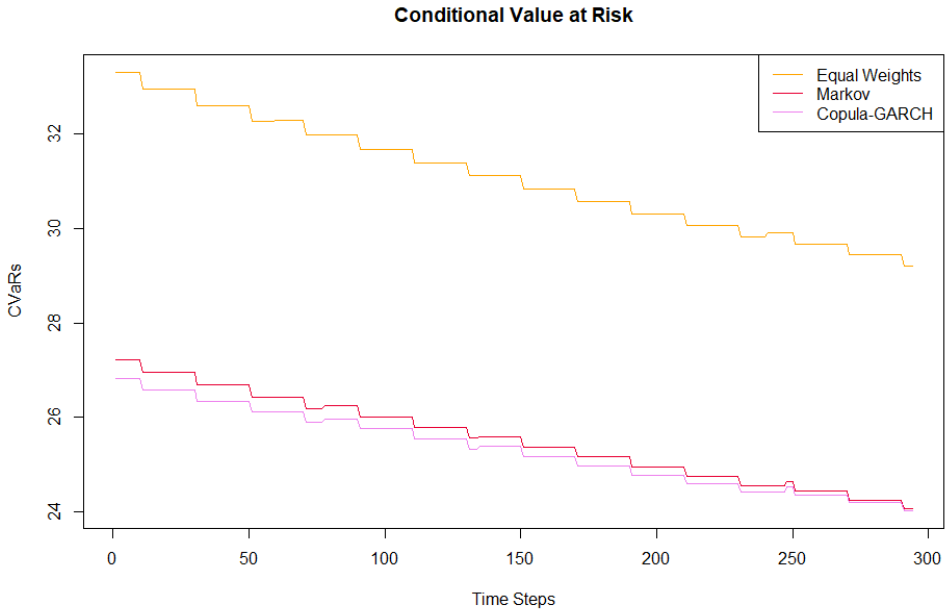


Figure 8.30: *Evolution of the Conditional Value at Risk obtained by applying the optimal weights to the test set of $N = 296$ points of the third partition.*

As already seen in Table 8.20, Figure 8.30 shows that the copula-GARCH portfolio is the least risky, exhibiting the lowest CVaR value throughout the entire test set. On the other hand, the Equal Weight portfolio, which was characterized by the lowest cumulative returns, also detains the highest risk, confirming how this method does not perform well in terms of balancing return and risk, as it does not take into account any information of the underlying assets.

Lastly, we compare the STARR ratios in Figure 8.31, and analyse the corresponding differences in Figure 8.32. From both plots, it is clear that the vine-copula-GARCH portfolio outperforms the other two for the entire test set, maintaining a positive difference. It is also interesting to note the similar behaviour of the copula-GARCH portfolio and Markowitz's one, which is derived by the similar compositions of the two portfolios. This was a trend that was discussed also in the previous partition, hence in Section 8.6.2. In fact, the ratio of both portfolios have close values and maintain a similar trend.

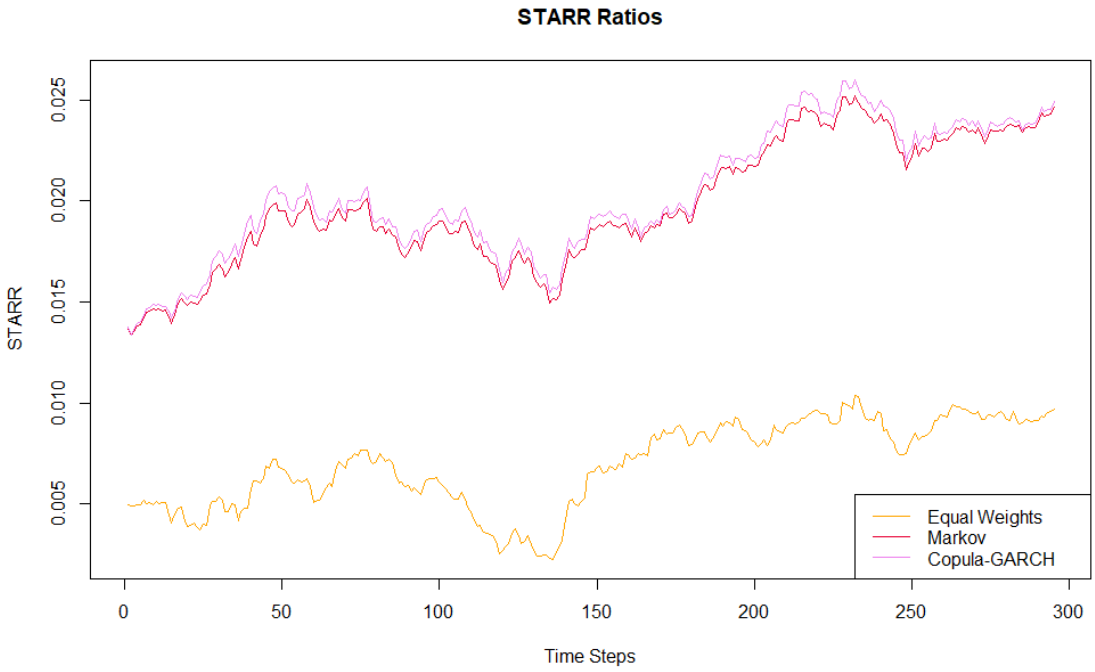


Figure 8.31: *Evolution of the STARR Ratios obtained by applying the optimal weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one to the test set of $N = 296$ points.*

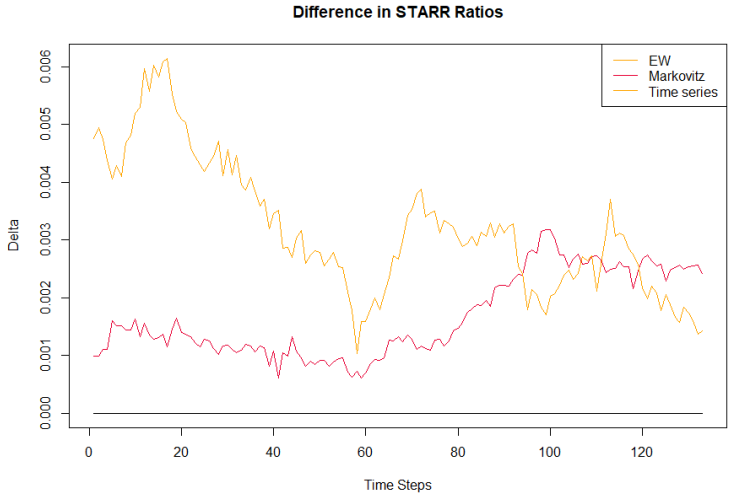


Figure 8.32: *Difference between the STARR Ratio obtained from the weights of the optimal copula-GARCH portfolio, of Markovitz' portfolio and of the EW one on the set of the third partition.*

To confirm the out-performance of the copula-GARCH portfolio, we compare the considered metrics at time $T = 05/07/2024$, the last day of the test set, in Table 8.21. The copula-GARCH indeed holds the best STARR ratio, with approximately a 1.20%. improvement over the second best, Markowitz's portfolio. The copula-GARCH portfolio also detains the best cumulative return value, confirming the superiority of the chosen method. Coherently with expectations, the values of the mean-variance portfolio are rather close to the former one, but slightly worse. Lastly, as already discussed throughout this section, the EW's performance is bad: it has the lowest return, as well as being exposed to the highest

risk, which results in a very poor STARR ratio value.

Method	Cumulative Return	CVaR	STARR
EW	163.1	29.21	0.01%
Mean-Var	325.2	24.07	2.46%
copula-GARCH	**331.7**	24.03	**2.49**%

Table 8.21: *Cumulative returns, CVaRs and STARR ratios for each method at the last day of the test set, hence at $T = 05/07/2024$.*

8.7 Conclusions

Our study highlights several points regarding the application of vine-copula-GARCH models for portfolio optimization.

Firstly, the long-term advantage of the vine-copula-GARCH approach is clear. For all partitions, we observed an improvement in the STARR ratios compared to portfolios constructed using other methods, demonstrating the robustness and effectiveness of this model in capturing complex dependencies over time. This improvement resulted also in higher cumulative returns for most of the test periods considered, meaning that a potential investor would have gained more if he decided to invest according to the proposed portfolio throughout the test set considered.

Moreover, a portfolio such as the Equal Weight one, which is not derived by taking into account the information of the underlying assets, can sometimes perform well (refer to Section 8.6.1). However, this success is inconsistent, and most of the time, understanding the relationships among the variables proves essential to guarantee satisfactory results. This is particularly evident in the other two partitions discussed in Sections 8.6.2 and 8.6.3.

Additionally, throughout all partitions considered, the vine copula fit showed consistent results in terms of Kendall's tau. Specifically, the first tree was always characterized by bivariate copulas with moderate positive correlations between the variables, while the conditional copulas of the subsequent trees had correlations that tended to zero.

Lastly, the initial analysis conducted on the original set of partitions (Section 8.5) highlighted the importance of having a dataset with a sufficient number of points, ideally $N = 500$. In fact, a larger dataset ensures more accurate partitioning and stable estimation of the copula parameters, leading to more reliable results and conclusions.

9

Conclusion

In this thesis, we proposed to analyze the advantages and disadvantages of introducing more sophisticated mathematical models in portfolio optimization. We explored the impact of using time-dependent vine models, applied to the residuals of time series models, for portfolio optimization. We compared the performance of these models with more traditional approaches like Markowitz's Mean-Variance approach and simpler models such as the Equal Weight portfolio, aiming to address the following questions introduced in the introduction:

What are the advantages of introducing copulas and vine models, in the context of portfolio optimisation theory? How can the properties of these models be used to improve and understand the optimisation process?

To answer these questions, we developed a vine-copula-GARCH model by introducing each component in different chapters of this thesis, like pieces of a puzzle that we carefully evaluated before assembling. This model incorporates various mathematical tools to understand and mimic the dynamics of time series, including GARCH models, copulas, and vine models. Through simulation studies, we verified the performance of these instruments in a controlled environment, where outcomes could be predicted, ensuring everything was working correctly. In other words, the simulation studies were not only a tool to provide to the reader some examples to understand each step, but were also a way of testing the methodology developed, and to verify that the expected results were achieved.

The simulation studies helped us calibrate the model correctly, ensuring its proper functioning. For instance, we determined the necessity of having at least $N = 500$ observations to achieve satisfactory parameter estimation in vine copula models. But more in general, this approach allowed to gain an overall overview on the different components: from selecting the most appropriate risk measure in Chapter 2, to determining the appropriate number of points needed to fit a good model in Chapters 3 and 4, to verifying the efficacy of all instruments combined together in Chapters 6 and 7. Each of these steps was carefully executed, leading to the successful construction of the final model, which was applied to real financial data in the application of Chapter 8.

The created model also proved to be successful in the real data application, highlighting the importance of using a model able to understand and exploit the complex relationships that hold across different variables. Some of the main advantages will be discussed in the following Section.

9.1 Model's Advantages

The real-world application of Chapter 8 demonstrated the practical utility and robustness of the vine-copula-GARCH model in portfolio optimization. By applying our model to actual financial data, we could observe how well it performed compared to traditional models over different market conditions and time periods. The empirical results highlighted several key points:

1. Optimized Risk-Adjusted Performance: Our vine-copula-GARCH model consistently showed better risk-adjusted performance metrics. In particular, the model exhibited a better STARR ratio, compared to the traditional Mean-Variance and Equal Weight portfolios. This suggests that incorporating the dependence structure through vine copulas provides a significant advantage in managing risk while optimizing returns.

2. Long-term advantage: For all partitions, we observed an improvement in the STARR ratios compared to portfolios constructed using other methods that was prolonged in the entire test set, demonstrating the robustness and effectiveness of this model in capturing complex dependencies over time. This improvement resulted also in higher cumulative returns for most of the test periods considered, meaning that a potential investor would have gained more if he decided to invest according to the proposed portfolio throughout the test set considered.

3. Robustness Across Different Partitions: The model demonstrated robustness across various partitions of the dataset, consistently outperforming other models in terms of STARR ratio. This consistency underscores the model's reliability and potential for long-term investment strategies. This feature is particularly evident if compared to the Equal Weight portfolio. In fact, as we observed in Section 8.6.1, this portfolio can sometimes perform well. However, because it does not account for the information of the underlying assets, its success is inconsistent. Most of the time, understanding the relationships among the variables proves essential for guaranteeing satisfactory results. This is particularly evident in the other two partitions where the EW portfolio's performance is clearly inferior.

4. Insights into Dependencies: The consistent results in terms of Kendall's tau across different partitions reinforced the importance of understanding the underlying dependencies among assets, as well as giving insights about the stronger dependencies across different geographies, as well as in the conditional dependencies.

In conclusion, this thesis demonstrates that incorporating advanced mathematical models, particularly vine-copula-GARCH models, into portfolio optimization provides clear benefits over traditional methods. By accurately capturing the dependencies among financial instruments and adapting to changing market conditions, these models offer a robust framework for constructing optimal investment portfolios. While the model shows significant promise, there is still room for possible improvements and extensions, which will be discussed in the following sections.

9.2 Suggested Improvements

While the created model produced satisfactory results, there are several areas where improvements can be made. These improvements were not explored due to time constraints, but they offer potential for enhancing the model's accuracy and robustness.

The first improvement would be to consider longer histories of data. As discussed in Section 8.5, the division into partitions obtained from the method illustrated in Section 6.6 proved to be unsuccessful as some of the identified partitions did not have enough points to guarantee a good fit. Using a longer dataset could enable a more accurate subdivision of time based on correlation differences. This would lead to partitions characterized by consistent underlying models, enhancing the model's overall performance.

In addition, some improvements can be made in the context of applying Monte Carlo simulations. To improve the precision of the estimates, variance reduction techniques such as importance sampling, antithetic variates, or control variates could be considered. Furthermore, portfolio revaluation could be implemented to reduce the time between different scenarios, reducing the errors [79].

9.3 Possible Extensions and Future Developments

There are several potential extensions and future developments that could enhance the scope and applicability of this thesis, although they were not explored due to time and resource constraints. The main areas of focus include:

- Scalability
- Diversification

Scalability refer to the selection of indexes discussed in the real data application of Chapter 8. The model was applied to the returns of four of the main stock indexes. However, copula and vine models are capable of incorporating data from many more variables. The model developed can be applied to different stock markets, and include more geographies simultaneously. For lack of time we did not have time to expand the model to further indexes, but the flexibility of vine model enables to create much more bigger (and hence more diversified) portfolios.

Moreover, the model is not limited to stock indexes alone. Expanding the model to other types of financial assets would be a logical next step. This could include bonds, commodities, or other investment assets. The choice to focus on stock indexes in this thesis was due to data availability. In fact, the data for the stock indexes of the main geographies are available online, rather complete and most importantly they guarantee a substantial data history, allowing us to divide the dataset in partitions, and have enough data to guarantee a proper model fit.

References

[1] H. Markowitz, "Portfolio selection*," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952. DOI: https://doi.org/10.1111/j.1540-6261.1952.tb01525.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1952.tb01525.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1952.tb01525.x.

[2] E. J. Elton and M. J. Gruber, "Modern portfolio theory, 1950 to date," *Journal of banking & finance*, vol. 21, no. 11-12, pp. 1743–1759, 1997.

[3] G. P. Szegö, *Portfolio theory: with application to bank asset management*. Academic Press, 2014.

[4] M. Sklar, "Fonctions de repartition a n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229–231, 1959. [Online]. Available: http://dx.doi.org/10.2139/ssrn.4198458.

[5] T. Bedford and R. M. Cooke, "Vines–a new graphical model for dependent random variables," *The Annals of Statistics*, vol. 30, no. 4, pp. 1031–1068, 2002. DOI: 10.1214/aos/1031689016. [Online]. Available: https://doi.org/10.1214/aos/1031689016.

[6] N. Barthel, C. Czado, and Y. Okhrin, "A partial correlation vine based approach for modeling and forecasting multivariate volatility time-series," *Computational Statistics Data Analysis*, vol. 142, p. 106 810, 2020, ISSN: 0167-9473. DOI: https://doi.org/10.1016/j.csda.2019.106810. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167947319301550.

[7] M. Şerban, A. Brockwell, J. Lehoczky, and S. Srivastava, "Modelling the dynamic dependence structure in multivariate financial time series," *Journal of Time Series Analysis*, vol. 28, no. 5, pp. 763–782, 2007. DOI: https://doi.org/10.1111/j.1467-9892.2007.00543.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9892.2007.00543.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.2007.00543.x.

[8] C. Shenoy and P. P. Shenoy, "Bayesian network models of portfolio risk and return," in The MIT Press, 2000.

[9] A. Gunjan and S. Bhattacharyya, "A brief review of portfolio optimization techniques," *Artificial Intelligence Review*, vol. 56, no. 5, pp. 3847–3886, 2023.

[10] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical finance*, vol. 9, no. 3, pp. 203–228, 1999.

[11] J. Morgan *et al.*, "Creditmetrics-technical document," *JP Morgan, New York*, vol. 1, pp. 102–127, 1997.

[12] I. Kakouris and B. Rustem, "Robust portfolio optimization with copulas," *European Journal of Operational Research*, vol. 235, no. 1, pp. 28–37, 2014.

[13] G. Serraino and S. Uryasev, "Conditional value-at-risk (cvar)," *Encyclopedia of operations research and management science*, pp. 258–266, 2013.

[14] G. Szegö, "Measures of risk," *European Journal of Operational Research*, vol. 163, no. 1, pp. 5–19, 2005, Financial Modelling and Risk Management, ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2003.12.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221703009056.

[15] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[16] R. T. Rockafellar, S. Uryasev, *et al.*, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.

[17] A. Krzemienowski and S. Szymczyk, "Portfolio optimization with a copula-based extension of conditional value-at-risk," *Annals of Operations Research*, vol. 237, pp. 219–236, 2016.

[18] S. Uryasev, "Conditional value-at-risk: Optimization algorithms and applications," in *proceedings of the IEEE/IAFE/INFORMS 2000 conference on computational intelligence for financial engineering (CIFEr)(Cat. No. 00TH8520)*, IEEE, 2000, pp. 49–57.

[19] X. Lin, "Researches on the efficient frontier of mean-cvar under normal distribution condition," in *2009 International Conference on Management and Service Science*, IEEE, 2009, pp. 1–4.

[20] S. V. Stoyanov, S. T. Rachev, and F. J. Fabozzi, "Optimal financial portfolios," *Applied Mathematical Finance*, vol. 14, no. 5, pp. 401–436, 2007.

[21] S. Rachev, T. Jašić, S. Stoyanov, and F. J. Fabozzi, "Momentum strategies based on reward–risk stock selection criteria," *Journal of Banking & Finance*, vol. 31, no. 8, pp. 2325–2346, 2007.

[22] W. F. Sharpe, "Mutual fund performance," *The Journal of business*, vol. 39, no. 1, pp. 119–138, 1966.

[23] W. F. Sharpe, "The sharpe ratio," *Streetwise–the Best of the Journal of Portfolio Management*, vol. 3, pp. 169–85, 1998.

[24] C.-Y. Tsao, "Portfolio selection based on the mean–var efficient frontier," *Quantitative Finance*, vol. 10, no. 8, pp. 931–945, 2010.

[25] D. Wuertz, T. Setz, Y. Chalabi, W. Chen, and S. Theussl, *Fportfolio: Rmetrics - portfolio selection and optimization*, R package Version 4023.84, 2023. [Online]. Available: https://cran.r-project.org/web/packages/fPortfolio/index.html.

[26] E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli, "Copulas for finance-a reading guide and some applications," *Available at SSRN 1032533*, 2000.

[27] C. Czado, *Analyzing Dependent Data with Vine Copulas: A Practical Guide With R* (Lecture Notes in Statistics). Springer International Publishing, 2019, ISBN: 9783030137847. [Online]. Available: https://books.google.nl/books?id=8GgtwQEACAAJ.

[28] B. Nelsen Roger, *An introduction to copulas*. New York: Springer, 1999, ISBN: 0387986235.

[29] M. Scarsini, "On measures of concordance.," *Stochastica*, vol. 8, no. 3, pp. 201–218, 1984.

[30] L. S. Junior and I. D. P. Franca, "Correlation of financial markets in times of crisis," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 1-2, pp. 187–208, 2012.

[31] P. Jaworski, F. Durante, W. Härdle, and T. Rychlik, *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009* (Lecture Notes in Statistics). Springer Berlin Heidelberg, 2010, ISBN: 9783642124655. [Online]. Available: https://books.google.nl/books?id=vX233feaA6MC.

[32] J. Mai and M. Scherer, *Financial Engineering with Copulas Explained* (Financial Engineering Explained). Palgrave Macmillan UK, 2014, ISBN: 9781137346315. [Online]. Available: https://books.google.nl/books?id=dmSoBAAAQBAJ.

[33] X. Zeng, J. Ren, M. Sun, S. Marshall, and T. Durrani, "Copulas for statistical signal processing (part ii): Simulation, optimal selection and practical applications," *Signal Processing*, vol. 94, pp. 681–690, 2014, ISSN: 0165-1684. DOI: https://doi.org/10.1016/j.sigpro.2013.07.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168413002855.

[34] T. Nagler and *et al.*, *Vinecopula: Statistical inference of vine copulas*, 2023. [Online]. Available: https://cran.r-project.org/web/packages/VineCopula/index.html.

[35] O. Okhrin, S. Trimborn, and M. Waltz, *Gofcopula: Goodness-of-fit tests for copulae*, R package version 0.4-1, 2021. [Online]. Available: https://CRAN.R-project.org/package=GoF.

[36] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, no. 1, pp. 1–25, 1982, ISSN: 00129682, 14680262. [Online]. Available: http://www.jstor.org/stable/1912526 (visited on 06/10/2024).

[37] W. Huang and A. Prokhorov, "A goodness-of-fit test for copulas," *Econometric Reviews*, vol. 33, no. 7, pp. 751–771, 2014.

[38] A. Derumigny and J.-D. Fermanian, *Dependence Modeling*, vol. 5, no. 1, pp. 154–197, 2017. DOI: doi:10.1515/demo-2017-0011. [Online]. Available: https://doi.org/10.1515/demo-2017-0011.

[39] M. Killiches, D. Kraus, and C. Czado, *Examination and visualisation of the simplifying assumption for vine copulas in three dimensions*, 2016. arXiv: 1602.05795 [stat.AP].

[40] I. Hobæk Haff, K. Aas, and A. Frigessi, "On the simplified pair-copula construction – simply useful or too simplistic?" *Journal of Multivariate Analysis*, vol. 101, no. 5, pp. 1296–1310, 2010. [Online]. Available: https://EconPapers.repec.org/RePEc:eee:jmvana:v:101:y:2010:i:5:p:1296-1310.

[41] E. F. Acar, C. Genest, and J. Nešlehová, "Beyond simplified pair-copula constructions," *Journal of Multivariate Analysis*, vol. 110, pp. 74–90, 2012, Special Issue on Copula Modeling and Dependence, ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2012.02.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0047259X12000358.

[42] A. Derumigny, J.-D. Fermanian, A. Min, and R. van der Spek, *Condcopulas: Estimation and inference for conditional copula models*, 2023. [Online]. Available: https://github.com/AlexisDerumigny/CondCopulas.

[43] C. Czado and T. Nagler, "Vine copula based modeling," *Annual Review of Statistics and Its Application*, 2021. [Online]. Available: https://tnagler.github.io/vine-arisa.pdf.

[44] T. Bedford and R. Cooke, "Monte carlo simulation of vine dependent random variables for applications in uncertainty analysis," *ESREL 2003*, 2001.

[45] O. Morales-Napoles, "Counting vines," in *Dependence modeling: Vine copula handbook*, World Scientific, 2010, pp. 189–218.

[46] J. Dißmann, E. Brechmann, C. Czado, and D. Kurowicka, "Selecting and estimating regular vine copulae and application to financial returns," *Computational Statistics Data Analysis*, vol. 59, pp. 52–69, 2013, ISSN: 0167-9473. DOI: https://doi.org/10.1016/j.csda.2012.08.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167947312003131.

[47] I. H. Haff, "Parameter estimation for pair-copula constructions," *Bernoulli*, vol. 19, no. 2, May 2013. DOI: 10.3150/12-bej413. [Online]. Available: https://doi.org/10.3150%5C%2F12-bej413.

[48] H. Tsukahara, "Semiparametric estimation in copula models," *Canadian Journal of Statistics*, vol. 33, no. 3, pp. 357–375, 2005. DOI: https://doi.org/10.1002/cjs.5540330304. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cjs.5540330304.

[49] R. Shumway and D. Stoffer, *Time Series Analysis and Its Applications* (Springer Texts in Statistics). Springer New York, 2013, ISBN: 9781475732610. [Online]. Available: https://books.google.nl/books?id=V-fcBwAAQBAJ.

[50] P. R. Dewick and S. Liu, "Copula modelling to analyse financial data," *Journal of Risk and Financial Management*, vol. 15, no. 3, 2022, ISSN: 1911-8074. DOI: 10.3390/jrfm15030104. [Online]. Available: https://www.mdpi.com/1911-8074/15/3/104.

[51] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986, ISSN: 0304-4076. DOI: https://doi.org/10.1016/0304-4076(86)90063-1. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0304407686900631.

[52] C. Francq and J.-M. Zakoian, *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.

[53] S. Shahriari, S. Sisson, and T. Rashidi, "Copula arma-garch modelling of spatially and temporally correlated time series data for transportation planning use," *Transportation Research Part C: Emerging Technologies*, vol. 146, p. 103 969, Jan. 2023. DOI: 10.1016/j.trc.2022.103969.

[54] H. Sun, D. Yan, N. Zhao, and J. Zhou, "Empirical investigation on modeling solar radiation series with arma–garch models," *Energy Conversion and Management*, vol. 92, pp. 385–395, 2015.

[55] A. Banerjee, J. J. Dolado, J. W. Galbraith, and D. Hendry, *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford University Press, May 1993, ISBN: 9780198288107. DOI: 10.1093/0198288107.001.0001. [Online]. Available: https://doi.org/10.1093/0198288107.001.0001.

[56] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, vol. 71, no. 3, pp. 599–607, 1984, ISSN: 00063444. [Online]. Available: http://www.jstor.org/stable/2336570 (visited on 06/27/2023).

[57] E. Paparoditis and D. N. Politis, "The asymptotic size and power of the augmented dickey–fuller test for a unit root," *Econometric Reviews*, vol. 37, no. 9, pp. 955–973, 2018.

[58] A. I. McLeod and W. K. Li, "Diagnostic checking arma time series models using squared-residual autocorrelations," *Journal of time series analysis*, vol. 4, no. 4, pp. 269–273, 1983.

[59] J. D. Cryer, *Time series analysis*. Springer, 1986, vol. 286.

[60] Y. Feng, Y. Zhang, and Y. Wang, "Out-of-sample volatility prediction: Rolling window, expanding window, or both?" *Journal of Forecasting*, vol. 43, no. 3, pp. 567–582, 2024.

[61] R. Prado and M. West, *Time series: modeling, computation, and inference*. Chapman and Hall/CRC, 2010.

[62] Y. Liu and R. Luger, "Efficient estimation of copula-garch models," *Computational Statistics & Data Analysis*, vol. 53, no. 6, pp. 2284–2297, 2009.

[63] C. Czado, K. Bax, Ö. Sahin, T. Nagler, A. Min, and S. Paterlini, "Vine copula based dependence modeling in sustainable finance," *The Journal of Finance and Data Science*, vol. 8, pp. 309–330, 2022.

[64] L. Hu, "Dependence patterns across financial markets: A mixed copula approach," *Applied financial economics*, vol. 16, no. 10, pp. 717–729, 2006.

[65] E. Jondeau and M. Rockinger, "The copula-garch model of conditional dependencies: An international stock market application," *Journal of international money and finance*, vol. 25, no. 5, pp. 827–853, 2006.

[66] T. Nagler, D. Krüger, and A. Min, "Stationary vine copula models for multivariate time series," *Journal of Econometrics*, vol. 227, no. 2, pp. 305–324, 2022.

[67] R. Engle, "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business & Economic Statistics*, vol. 20, no. 3, pp. 339–350, 2002.

[68] R. Ibragimov and G. Lentzas, "Copulas and long memory," 2017.

[69] X. Chen, Z. Xiao, and B. Wang, "Copula-based time series with filtered nonstationarity," *Journal of econometrics*, vol. 228, no. 1, pp. 127–155, 2022.

[70] Z.-R. Wang, X.-H. Chen, Y.-B. Jin, and Y.-J. Zhou, "Estimating risk of foreign exchange portfolio: Using var and cvar based on garch–evt-copula model," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 21, pp. 4918–4928, 2010, ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2010.07.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437110006394.

[71] J.-J. Huang, K.-J. Lee, H. Liang, and W.-F. Lin, "Estimating value at risk of portfolio by conditional copula-garch method," *Insurance: Mathematics and economics*, vol. 45, no. 3, pp. 315–324, 2009.

[72] C. Simard and B. Remillard, "Forecasting time series with multivariate copulas," *SSRN Electronic Journal*, vol. 3, Jan. 2013. DOI: 10.2139/ssrn.2295120.

[73] A. Derumigny, J.-D. Fermanian, and A. Min, "Testing for equality between conditional copulas given discretized conditioning events," *Canadian Journal of Statistics*, vol. 51, no. 4, pp. 1084–1110, 2023.

[74] J.-J. Huang, K.-J. Lee, H. Liang, and W.-F. Lin, "Estimating value at risk of portfolio by conditional copula-garch method," *Insurance: Mathematics and Economics*, vol. 45, no. 3, pp. 315–324, 2009, ISSN: 0167-6687. DOI: https://doi.org/10.1016/j.insmatheco.2009.09.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167668709001267.

[75] M. Sahamkhadam, A. Stephan, and R. Östermark, "Portfolio optimization based on garch-evt-copula forecasting models," *International Journal of Forecasting*, vol. 34, no. 3, pp. 497–506, 2018.

[76] R. Ji, M. A. Lejeune, and Z. Fan, "Distributionally robust portfolio optimization with linearized starr performance measure," *Quantitative Finance*, vol. 22, no. 1, pp. 113–127, 2022.

[77] D. Wernli, L. Böttcher, F. Vanackere, Y. Kaspiarovich, M. Masood, and N. Levrat, "Understanding and governing global systemic crises in the 21st century: A complexity perspective," *Global Policy*, vol. 14, no. 2, pp. 207–228, 2023.

[78] D. K. Pandey, V. Kumari, A. Palma, and J. W. Goodell, "Are markets in happier countries less affected by tragic events? evidence from market reaction to the israel–hamas conflict," *Finance Research Letters*, vol. 60, p. 104 893, 2024.

[79] P. Glasserman, *Monte Carlo methods in financial engineering*. Springer, 2004, vol. 53.

[80] A. Shapiro and Y. Wardi, "Nondifferentiability of the steady-state function in discrete event dynamic systems," *IEEE transactions on Automatic Control*, vol. 39, no. 8, pp. 1707–1711, 1994.

[81] F. Dekking, C. Kraaikamp, H. Lopuhaä, and L. Meester, *A Modern Introduction to Probability and Statistics: Understanding Why and How* (Springer Texts in Statistics). Springer London, 2006, ISBN: 9781846281686. [Online]. Available: `https://books.google.nl/books?id=TEcmHJX67coC`.

[82] F. Bijma, M. Jonker, A. van der Vaart, and R. Erné, *An Introduction to Mathematical Statistics*. Amsterdam University Press, 2017, ISBN: 9789462985100. [Online]. Available: `https://books.google.nl/books?id=WpXUAQAACAAJ`.

[83] F. J. Massey, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951, ISSN: 01621459. [Online]. Available: `http://www.jstor.org/stable/2280095` (visited on 03/12/2023).

[84] W. J. Conover, *Practical nonparametric statistics*. WorldCat.org, 1971.

[85] J. Hodges Jr, "The significance probability of the smirnov two-sample test," *Arkiv för matematik*, vol. 3, no. 5, pp. 469–486, 1958.

[86] E. Zio and E. Zio, *Monte carlo simulation: The method*. Springer, 2013.

[87] R. L. Harrison, "Introduction to monte carlo simulation," in *AIP conference proceedings*, American Institute of Physics, vol. 1204, 2010, pp. 17–21.

A

Proofs

A.1 Proofs Chapter 2

A.1.1 Theorem 2

Here we discuss the proof of Theorem 2, which discusses the convexity and differentiability of the function F_β, introduced to simplify the optimization of Conditional VaR. Both the theorem and the proof are taken from [16].

But first, we need to introduce a lemma, which is derived from proposition 2.1 of [80]:

Lemma 1. *With $\boldsymbol{R}$ fixed, let $G(\alpha) = \int_{\mathbb{R}^m} g(\alpha, r) f(r) dr$, where $g(\alpha, \boldsymbol{R}) = [L(\boldsymbol{w}, \boldsymbol{R}) - \alpha]^+$. Then G is a convex, continuously differentiable function with derivative*

$$G'(\alpha) = F(\alpha) - 1, \tag{A.1}$$

where F is the cumulative distribution function of $\boldsymbol{R}$.

Proof. Immediately from 1, we get that the function $F_\beta(L, \alpha)$, as defined in 2.21, is convex, continuously differentiable function with derivative:

$$\frac{\partial F_\beta}{\partial \alpha} = 1 + \frac{1}{1-\beta}(F(\alpha) - 1) = \frac{1}{1-\beta}(F(\alpha) - \beta). \tag{A.2}$$

Being convex, there exits only one minimum of $F_\beta(L, \alpha)$, located at the point that satisfies $\frac{\partial F_\beta}{\partial \alpha}$, hence the point such that $F(\alpha) - \beta = 0$. Since we consider the CDF $F(\alpha)$ to be continuous, non decreasing with limit 1 as $\alpha \to \infty$, and limit 0 as $\alpha \to -\infty$, we obtain that the solution form a non empty closed interval, which corresponds to (2.23). And subsequently proves (2.24), defining the $VaR_\beta(L)$ as the left end point of $A_\beta(L)$.

In particular, we have that the minimum of F_β is obtained when $\alpha = VaR_\beta(L)$, hence:

$$\min_{\alpha \in \mathbb{R}} F_\beta(L, \alpha) = F_\beta(L, VaR_\beta(L)) \overset{\text{Def. (2.21)}}{=} VaR_\beta(L) + \frac{1}{1-\beta} \int_{L(\mathbf{w}, \mathbf{R}) \geq VaR_\beta(L)} (L(\mathbf{w}, r) - VaR_\beta(L)) f(r) dr.$$

Then if we apply linearity to the last integral, we can split it in:

$$\int_{L(\mathbf{w},\mathbf{R})\geq VaR_\beta}(L(\mathbf{w},r)-VaR_\beta(L))f(r)dr=\int_{L(\mathbf{w},\mathbf{R})\geq VaR_\beta}L(\mathbf{w},r)f(r)dr-VaR_\beta\int_{L(\mathbf{w},\mathbf{R})\geq VaR_\beta}f(r)dr.$$

Where $\int_{L(\mathbf{w},\mathbf{R})\geq VaR_\beta}L(\mathbf{w},r)f(r)dr$ is the conditional expectation of the loss L, conditional to $L(\mathbf{w},\mathbf{R})\geq VaR_\beta$, which corresponds to the definition of $(1-\beta)CVaR_\beta$ (see Def. 2). While, $\int_{L(\mathbf{w},\mathbf{R})\geq VaR_\beta}f(r)dr=1-F(VaR_\beta)$, from the definition of CDF.

By combining these results, we obtain that:

$$\min_{\alpha\in\mathbb{R}}F_\beta(L,\alpha)=VaR_\beta(L)+\frac{1}{1-\beta}((1-\beta)CVaR_\beta-VaR_\beta(L)(1-F(VaR_\beta)),$$

but since $F(VaR_\beta)=\beta$, we obtain that:

$$CVaR_\beta=\min_{\alpha\in\mathbb{R}}F_\beta(L,\alpha),$$

which corresponds to (2.22).

□

A.1.2 Analytical Computation of the Mean-Var Efficient Frontier

Here is the proof of the results from Theorem 1 in Section 2.3.1.1, where we computed the analytical formula of Markovitz' Mean-Variance efficient frontier. The Variance can in fact be computed as a function of the vector of target returns r_0. The final formula is the one shown in 2.15:

$$\mathrm{Var}(r_0)=\frac{cr_0^2-2br_0+a}{\Delta},$$

where,

$$\begin{cases}a=\mu^\top\Sigma^{-1}\mu\\ b=\mu^\top\Sigma^{-1}\mathbb{1}\\ c=\mathbb{1}^\top\Sigma^{-1}\mathbb{1}.\end{cases}\tag{A.3}$$

The corresponding weight vector is:

$$w\left(r_0\right)=\frac{\left(c\Sigma^{-1}\mu-b\Sigma^{-1}\mathbb{1}\right)r_0+\left(a\Sigma^{-1}1-b\Sigma^{-1}\mu\right)}{\Delta}.\tag{A.4}$$

The proof is taken from [3].

Proof. Given the Lagrangian corresponding to problem 2.10, where the Risk is the Variance:

$$\mathcal{L}\left(\omega,\lambda_1,\lambda_2\right)=\omega^\top\Sigma\omega-\lambda_1\left(\mu^\top\omega-r_0\right)-\lambda_2\left(\omega^\top\mathbb{1}-1\right),$$

we can derive the critical points by solving the following partial derivatives:

$$\frac{\partial\mathcal{L}}{\partial\omega}=2\omega^\top\Sigma-\lambda_1\mu^\top-\lambda_2\mathbb{1}=0\tag{A.5}$$

$\Rightarrow$

$$\omega^\top = \frac{1}{2}\left[\lambda_1 \Sigma^{-1}\mu - \lambda_2 \Sigma^{-1}\mathbb{1}\right], \tag{A.6}$$

and

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = -\mu^\top \omega + r_0 = 0 \tag{A.7}$$

$\Rightarrow$

$$r_0 = \mu^\top \omega \overset{(A.5)}{=} \frac{1}{2}\lambda_1 \mu^\top \Sigma^{-1}\mu + \frac{1}{2}\lambda_2 \mu^\top \Sigma^{-1}\mathbb{1} \overset{(2.15)}{=} \frac{1}{2}\lambda_1 a + \frac{1}{2}\lambda_2 b. \tag{A.8}$$

And lastly:

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = -w^T \mathbb{1} - 1 = 0 \tag{A.9}$$

$\overset{(A.5)}{\Rightarrow}$

$$1 = -w^T \mathbb{1} = \frac{1}{2}\lambda_1 \mu^\top \Sigma^{-1}\mathbb{1} + \frac{1}{2}\lambda_2 \mathbb{1}^T \Sigma^{-1}\mathbb{1} \overset{(A.3)}{=} \frac{1}{2}\lambda_1 b + \frac{1}{2}\lambda_2 c. \tag{A.10}$$

From which we obtain the system:

$$\frac{1}{2}\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} r_0 \\ 1 \end{bmatrix}$$

which, if we set $\Delta \geq 0$, has solution:

$$\frac{1}{2}\lambda_1 = \frac{r_0 c - b}{\Delta}, \quad \frac{1}{2}\lambda_2 = \frac{a - r_0 b}{\Delta},$$

from which we obtain the weights vector that we presented in 2.16, by substituting the values of λ_1 and λ_2 in equation A.6:

$$w(r_0) = \frac{\left(c\Sigma^{-1}\mu - b\Sigma^{-1}\mathbb{1}\right) r_0 + \left(a\Sigma^{-1}1 - b\Sigma^{-1}\mu\right)}{\Delta}.$$

Lastly we can compute the variance vector of the efficient frontier as:

$$Var = w^\top \Sigma w = \frac{\left(c r_0^2 - 2b r_0 + a\right)}{\Delta},$$

which represents the parabola presented in 2.15. □

A.1.3 Analytical Computation of the Mean-Var Tangent Portfolio

In Section 2.3.1.1, we proposed the formula to derive the set of weights that satisfy (2.10), which corresponds to what we refer to as the Tangency Portfolio. Here we go through the main steps to derive such a result.

Starting from problem (2.30), and setting the variance as risk, we can express the corresponding Lagrangian:

$$\mathcal{L}(w, \lambda) = \frac{w^T \mu - r_f}{\sqrt{w^T \Sigma w}} - \lambda(w^T \mathbb{1} - 1).$$

From which we set equal to zero the partial derivatives with respect to w and λ, to find the solution of the maximization problem:

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\mu\sqrt{w^T\Sigma w} - (w^T\mu - r_f)\Sigma w}{(w^T\Sigma w)^{\frac{3}{2}}} - \lambda\mathbb{1} = 0; \tag{A.11}$$

and:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = w^T\mathbb{1} - 1. \tag{A.12}$$

With a few substitutions and small algebraic computations, we finally find that the vector weight is the following:

$$w_{tp} = \frac{\Sigma^{-1}(\mu - r_f\mathbb{1})}{\mathbb{1}^T\Sigma^{-1}(\mu - r_f\mathbb{1})},$$

which corresponds to the weights of the optimal portfolio on the Markowitz efficient frontier that maximizes the Sharpe Ratio. It is the same formulation reported in Equation 1.

A.1.4 Analytical Computation of VaR

In Section 2.3.2.1 we provided the formulation of the VaR for normally distributed variables:

$$\text{VaR}_\beta(R) = \Phi^{-1}(\beta)\sigma_P + \mu_P. \tag{A.13}$$

Proof. Starting from the Definition 1 of VaR:

$$VaR_\beta = \min_z \{F_R(z) \geqslant \beta\} \overset{\text{translation}}{=} \left\{ z \text{ s.t. } \Phi\left(\frac{z-\mu_P}{\sigma_P}\right) = \beta \right\} = \left\{ z \text{ s.t. } \frac{z-\mu_P}{\sigma_P} = \Phi^{-1}(\beta) \right\} =$$

$$= \left\{ z \text{ s.t. } z = \Phi^{-1}(\beta)\sigma_P + \mu_P \right\}.$$

The VaR is therefore obtained from the following formula:

$$\text{VaR}_\beta(R) = \Phi^{-1}(\beta)\sigma_P + \mu_P.$$

□

A.1.5 Analytical Computation of CVaR

Analogously to the VaR case, we can analytically derive the $CVaR_\beta$ of a of a portfolio composed by Gaussian assets, which is the following, as already discussed in Section 2.3.3.1:

$$\text{CVaR}_\beta = \frac{1}{1-\beta}\varphi\left(\Phi^{-1}(\beta)\right)\sigma_P - \mu_P,$$

where $\varphi(\cdot)$ is the density of a standard normal, and $\Phi^{-1}(\cdot)$ the inverse of the CDF of a standard normal. The proof is taken from [19].

Proof. Starting from Definition 2 we have:

$$(1-\beta)CVaR_\beta(L) = \int_{L(\mathbf{w},\mathbf{R}) \geq VaR_\beta(L)} L(\mathbf{w}, r) f(r) dr$$

However, since we are in the Gaussian case, hence $r \sim N(\mu_P, \Sigma)$, the loss function $L(\mathbf{w}, r) = -\mathbf{w}^T r$, is

also Normally distributed $L(\mathbf{w}, r) \sim N(\mu_P, \sigma_P)$. So, making a change of variable, we obtain that:

$$\int_{L(\mathbf{w},\mathbf{R})\geq VaR_\beta(L)} L(\mathbf{w}, r) f(r) dr = \int_{VaR_\beta(L)}^{+\infty} \frac{t}{\sqrt{2\pi}\sigma_P} \exp\left\{-\frac{(t-\mu_P)^2}{2{\sigma_P}^2}\right\} dt =$$
$$\overset{\{\text{add } +\mu_P - \mu_P\}}{=} \int_{VaR_\beta(L)}^{+\infty} \frac{t+\mu_P}{\sqrt{2\pi}\sigma_P} exp\{-\frac{(t+\mu_P)^2}{2\sigma_P^2}\} dt - \int_{VaR_\beta(L)}^{\infty} \frac{\mu_P}{\sqrt{2\pi}\sigma_P} exp\{-\frac{(t+\mu_P)^2}{2\sigma_P^2}\} dt$$

From here we obtain the following:

$$\text{CVaR}_\beta = \frac{1}{1-\beta}\left[\int_{\text{VaR}_\beta(L)}^{+\infty} \frac{t+\mu_P}{\sqrt{2\pi}\sigma_P} e^{-\frac{(t+\mu_P)^2}{2\sigma_P^2}} dt - \mu_P \int_{VaR_\beta(L)}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_P} e^{-\frac{(t+\mu_P)^2}{2\sigma_P^2}} dt\right] \tag{A.14}$$

If we then consider separately the two integrals in A.14, we obtain the following results.

$$\mu_P \left[\frac{1}{1-\beta}\int_{VaR_\beta(L)}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_P} e^{-\frac{(t+\mu_P)^2}{2\sigma_P^2}} dt\right] \overset{\text{Def of VaR}}{=} \mu_P \frac{1}{1-\beta} \cdot (1-\beta) = \mu_P.$$

And:

$$\frac{1}{1-\beta}\int_{VaR_\beta(L)}^{+\infty} \frac{t+\mu_P}{\sqrt{2\pi}\sigma_P} e^{-\frac{(t+\mu_P)^2}{2\sigma_P^2}} dt = \frac{1}{1-\beta}\left[-\frac{\sigma_P}{\sqrt{2\pi}} \exp\left\{-\left(\frac{t+\mu_P}{\sqrt{2}\sigma_P}\right)^2\right\}\right]_{VaR_\beta(L)}^{+\infty} =$$
$$= \frac{1}{1-\beta}\frac{\sigma_P}{\sqrt{2\pi}} \exp\left\{-\frac{(VaR_\beta(L)+\mu_P)^2}{2\sigma_P^2}\right\} \overset{Eq.A.13}{=} \frac{1}{1-\beta}\frac{\sigma_P}{\sqrt{2\pi}} \exp\left\{-\frac{\left(\cancel{-\mu_P} - \sigma_P\Phi^{-1}(\beta) + \cancel{\mu_P}\right)^2}{2\sigma_P^2}\right\} =$$
$$= \frac{1}{1-\beta}\frac{\sigma_P}{\sqrt{2\pi}} \exp\left\{-\frac{\cancel{\sigma_P^2}(\Phi^{-1}(\beta))^2}{2\cancel{\sigma_P^2}}\right\} = \frac{\sigma_P}{1-\beta}\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\Phi^{-1}(\beta))^2}{2}\right\} \overset{\text{Density}N(0,1)}{=} \frac{1}{1-\beta}\sigma_P\varphi\left(\Phi^{-1}(\beta)\right).$$

By combining the two results, we finally obtain:

$$\text{CVaR}_\beta = \frac{1}{1-\beta}\varphi\left(\Phi^{-1}(\beta)\right)\sigma_P - \mu_P.$$

□

A.2 Proofs Chapter 3

In this Section we discuss the proof of two theorems that illustrate how to express the theoretical definition of Kendall's tau and Spearman's correlation with copulas. Both proofs are taken from [27].

A.2.1 Theorem 3.7

Proof. We want to prove that $\tau = 4\int_{[0,1]^2} C(u_1, u_2)\, dC(u_1, u_2) - 1$.

Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed pairs. Starting from definition (3.2.1) we have:

$$\tau(X, Y) = \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0).$$

For continuous variables, we can define the probability of disconcordance as:

$$\mathbb{P}\left((X_1 - X_2)(Y_1 - Y_2) < 0\right) = 1 - \mathbb{P}\left((X_1 - X_2)(Y_1 - Y_2) > 0\right),$$

$\Rightarrow$

$$\tau = 2\mathbb{P}\left((X_1 - X_2)(Y_1 - Y_2) > 0\right) - 1. \tag{A.15}$$

Furthermore, we can rewrite:

$$\mathbb{P}\left((X_1 - X_2)(Y_1 - Y_2) > 0\right) = \mathbb{P}\left(X_1 > X_2, Y_1 > Y_2\right) + \mathbb{P}\left(X_1 < X_2, Y_1 < Y_2\right), \tag{A.16}$$

Now by applying the CDF to project the variables to the unit cube, hence by applying $u_i := F_i(x_i)$, we can show that the first probability of (A.16) is equal to:

$$\begin{aligned}
\mathbb{P}\left(X_1 > X_2, Y_1 > Y_2\right) &= \mathbb{P}\left(X_2 < X_1, Y_2 < Y_1\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{P}\left(X_2 < x_1, Y_2 < x_2\right) dC\left(F_1(x_1), F_2(x_2)\right) \\
&\overset{Def.33}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x_1, x_2) dC\left(F_1(x_1), F_2(x_2)\right) \\
&\overset{Eq.(3.1)}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C\left(F_1(x_1), F_2(x_2)\right) dC\left(F_1(x_1), F_2(x_2)\right) \\
&\overset{u_i := F_i(x_i)}{=} \int_0^1\int_0^1 C(u_1, u_2)\, dC(u_1, u_2).
\end{aligned}$$

Similarly we can show that:

$$\mathbb{P}\left(X_1 < X_2, Y_1 < Y_2\right) = \int_0^1\int_0^1 C(u_1, u_2)\, dC(u_1, u_2),$$

obtaining that:

$$\mathbb{P}\left((X_1 - X_2)(Y_1 - Y_2) > 0\right) = 2\int_0^1\int_0^1 C(u_1, u_2)\, dC(u_1, u_2).$$

We then substitute the result in Equation A.15, concluding the proof.

□

A.2.2 Theorem 3.8

Proof. Starting from the definition of Spearman's ρ_s in 5, and applying the transformation $U_i = F_i(X_i)$, we obtain:

$$\rho_s = \mathrm{Cor}\left(F_1(X_1), F_2(X_2)\right) = \mathrm{Cor}\left(U_1, U_2\right),$$

We can now express the correlation as:

$$\mathrm{Cor}(X, Y) = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{Var(X)}\sqrt{Var(Y)}},$$

and since we know that if $U_i \sim U[0, 1]$, we have that $\mathbb{E}[U_i] = \frac{1}{2}$, and $Var(U_i) = \frac{1}{12}$, we obtain that:

$$\rho_s = \frac{\mathbb{E}(U_1 U_2) - \frac{1}{4}}{\frac{1}{12}} = 12\mathbb{E}(U_1 U_2) - 3 = 12\int_0^1\int_0^1 u_1 u_2 dC(u_1, u_2) - 3,$$

concluding the proof. □

B

Simulation Studies

In this Section of the Appendix we collect some figures obtained by the simulation studies performed in different Chapters of this thesis.

B.1 Chapter 2

This figures are some of the results that are discussed in Chapter 2, and more in particular in Section 2.6. In this section we simulate different set of returns from different distributions, and compare the optimization results obtained by applying three different risk measures: variance, Value-at-Risk and Conditional Value-at-Risk.

B.1.1 Gaussian Case

Results obtained in Section 2.6.4.

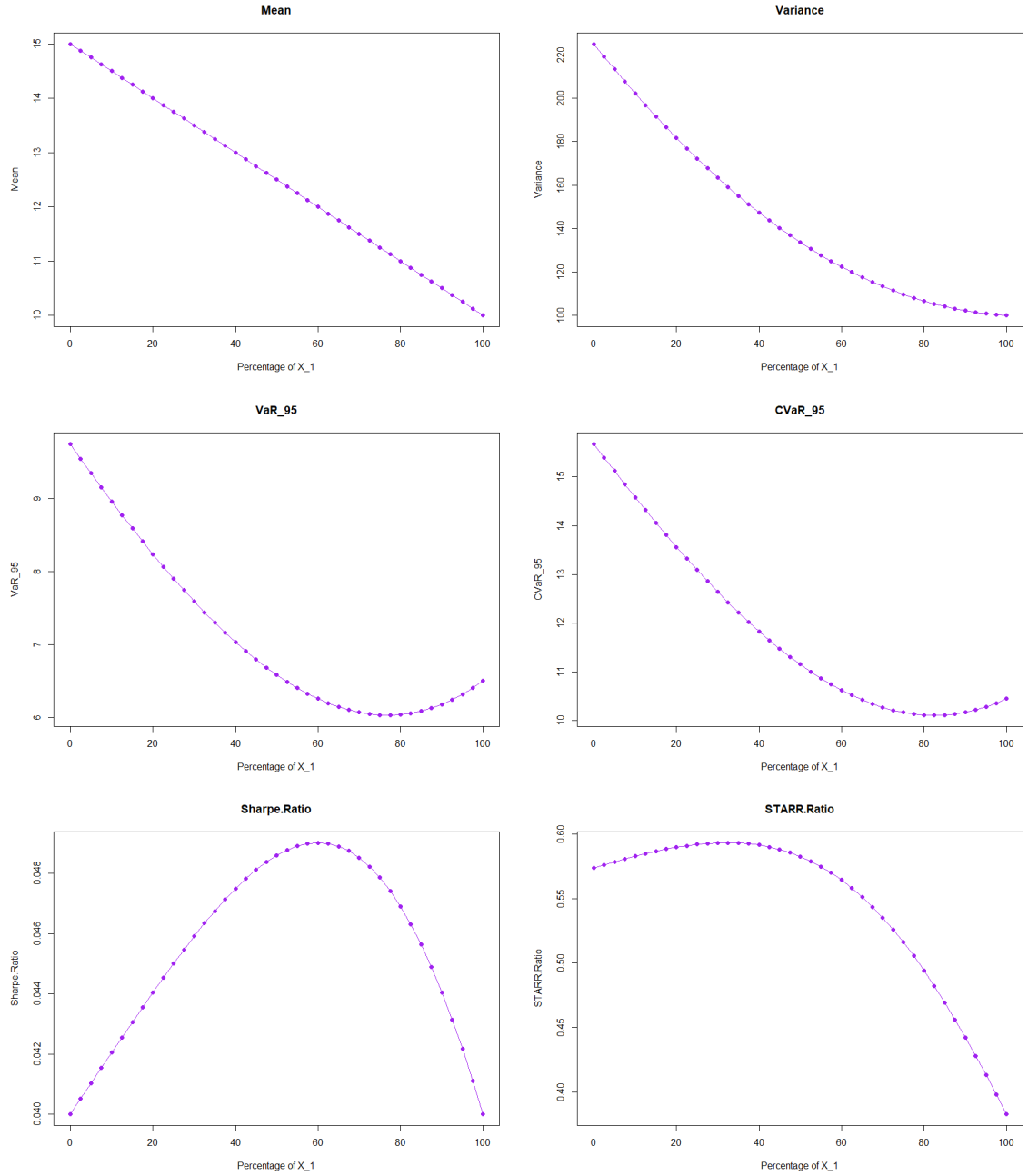


Figure B.1: *Plots showcasing the evolution of the risk measures considered, and the relative ratios in relation to the change of weights. In particular, in the X-axis we find the percentage of the first asset (less risky).*

B.1.2 The Exponential Case

Here we present the results related to Section 2.6.5.

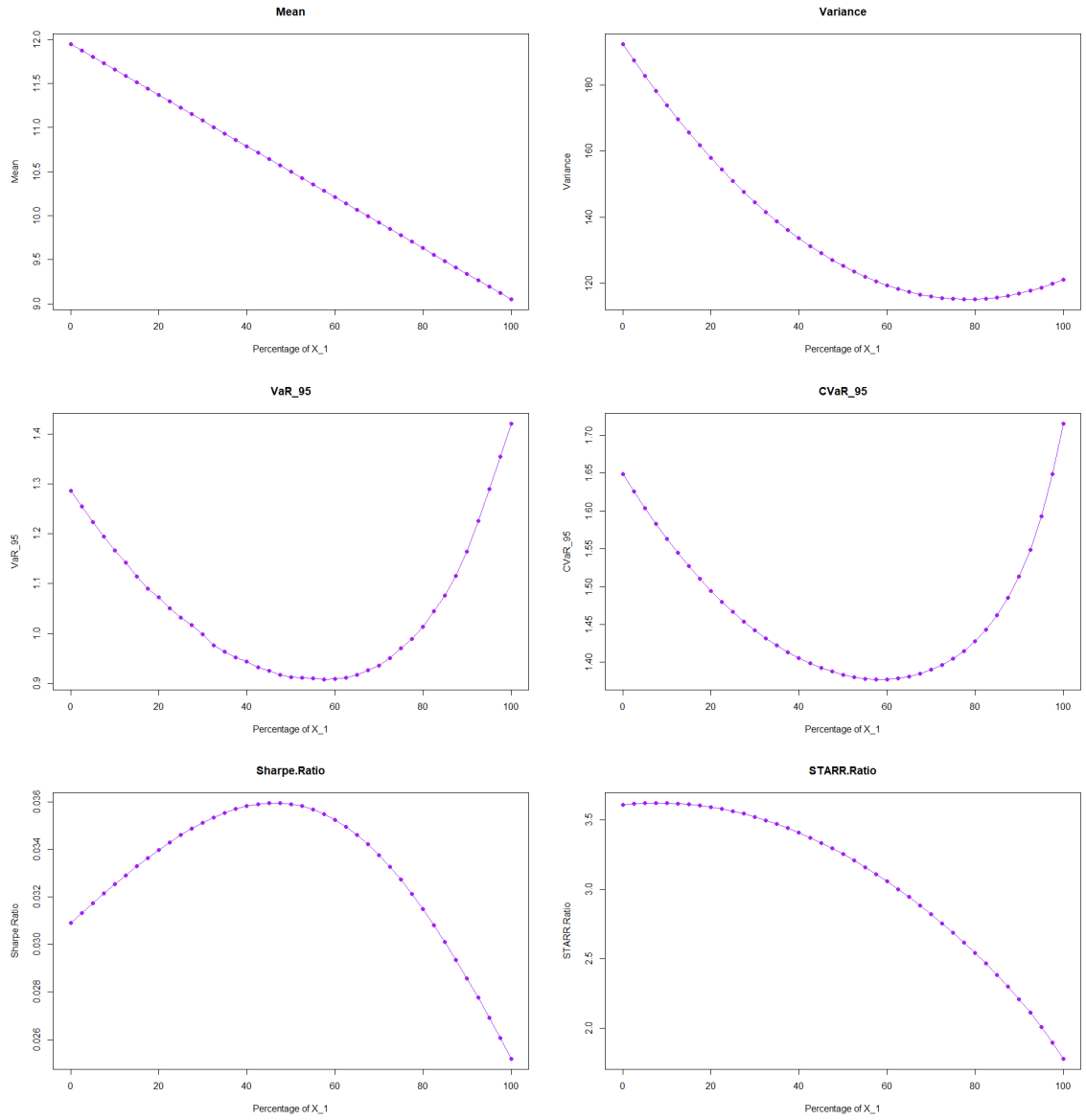


Figure B.2: *Plots showcasing the evolution of the risk measures considered, and the relative ratios in relation to the change of weights. In particular, in the X-axis we find the percentage of the first asset (less risky).*

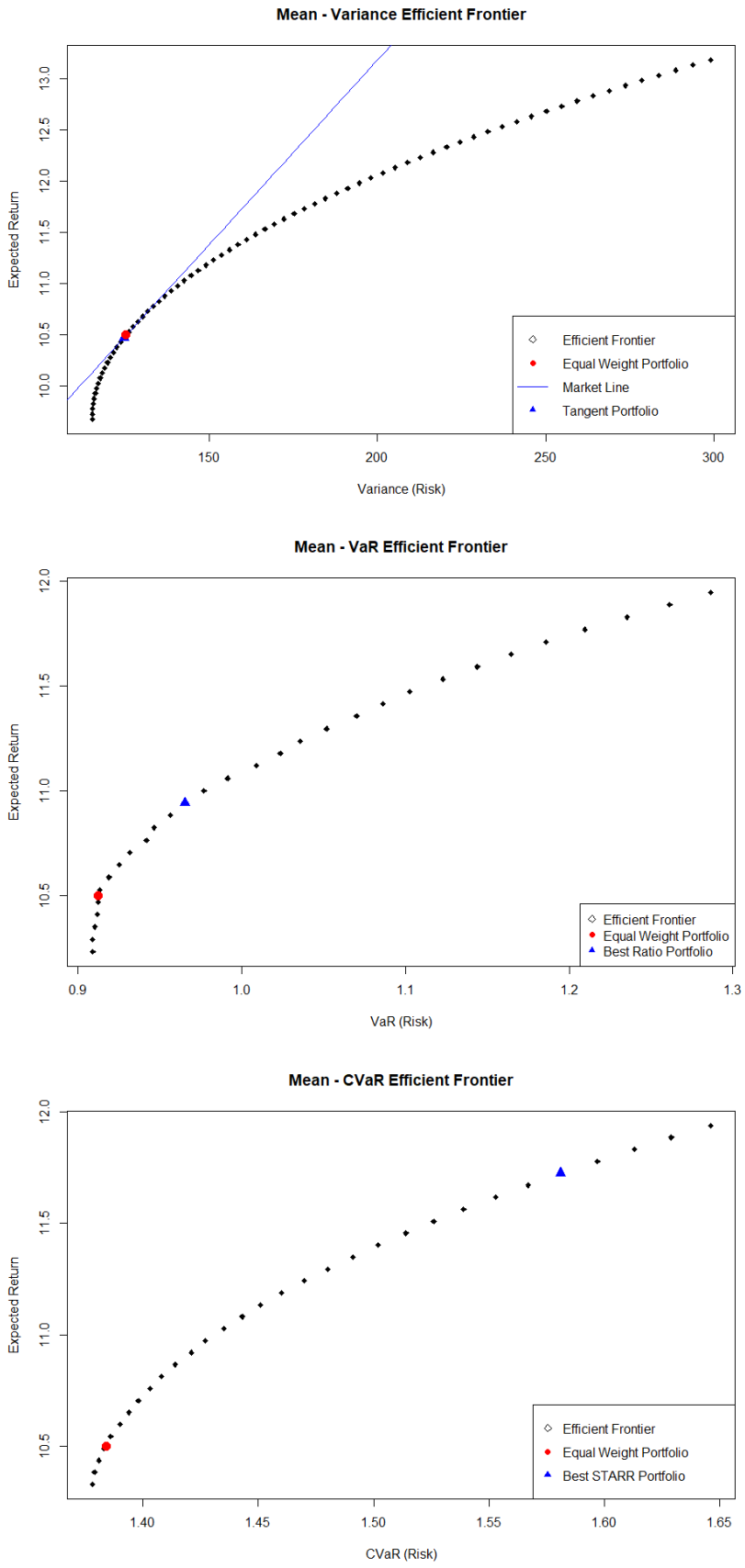
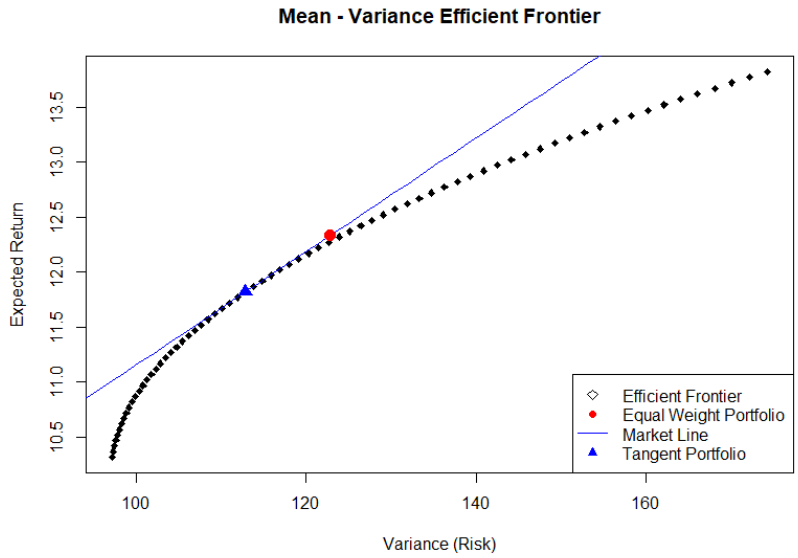
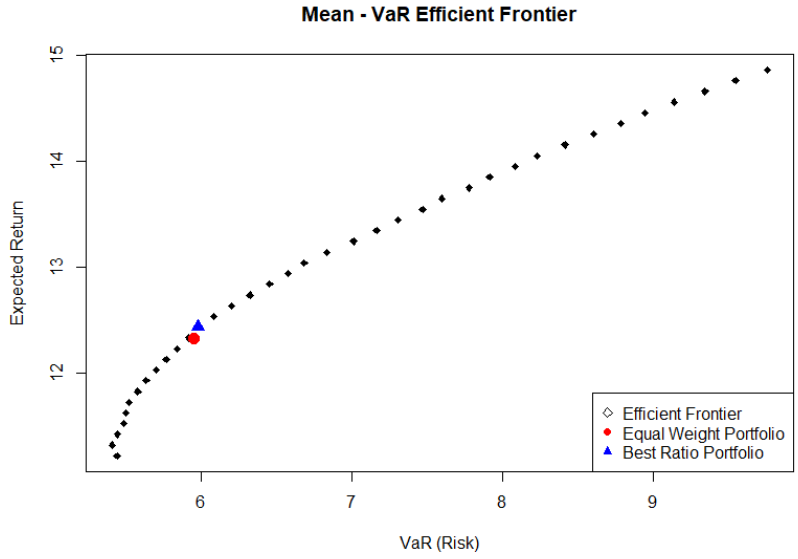


Figure B.3: *Mean-Variance, Mean-VaR_β, and Mean-$CVaR_\beta$ efficient frontiers, computed optimising from two exponentially distributed returns. In blue we observe the market line, whose intersection with the frontier identifies the optimal frontier. The results are compared with the EW portfolio (red dot).*

B.1.3 Gaussian 3-dimensional case

Results associated with Section 2.6.6 of the second Chapter, where we compare the three methods of portfolio optimization for a three dimensional portfolio whose returns are normally distributed.





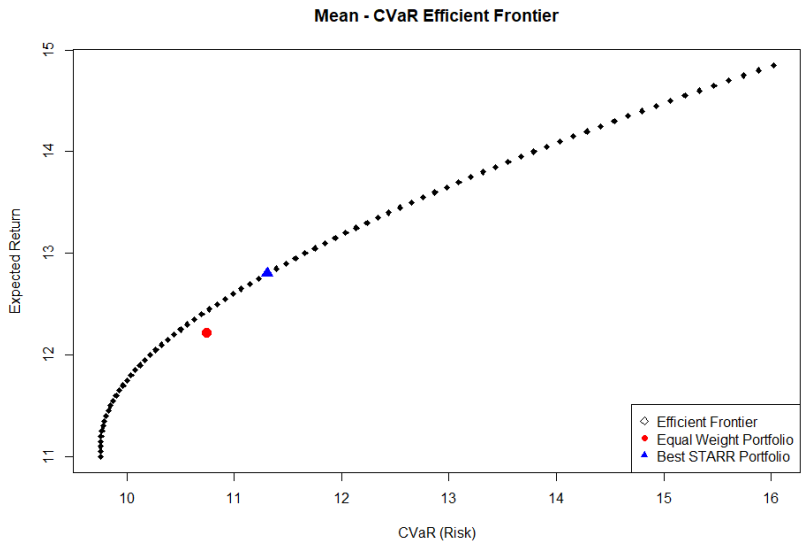
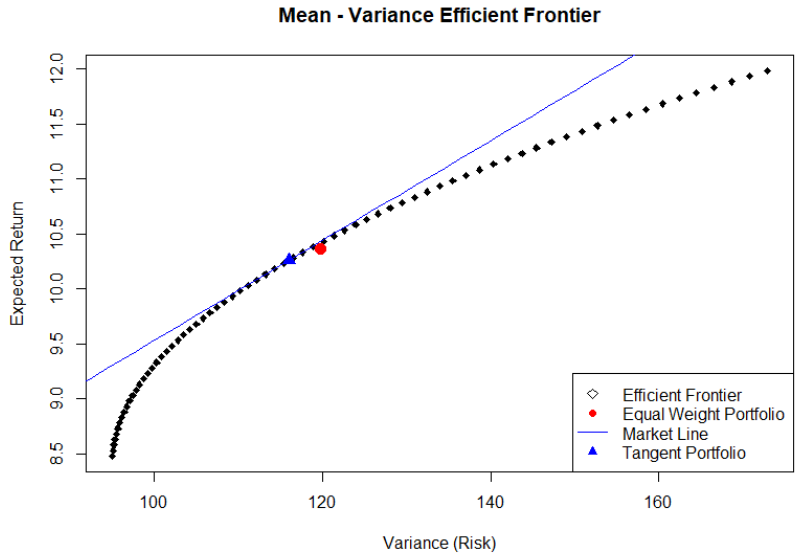


Figure B.4: *Mean-Variance, Mean-VaR_β, and Mean-$CVaR_\beta$ efficient frontiers, computed optimising the returns that follow a Gaussian distribution (3-dimensional case). In blue is presented the market line, whose intersection with the frontier identifies the optimal frontier. The results are compared with the EW portfolio (red dot).*

B.1.4 Copula case

Results associated with Section 2.6.6 of the second Chapter, where we compare the three methods of portfolio optimization for a three dimensional portfolio whose returns are simulated from a Gaussian copula, and then transformed to translated exponential by applying their inverse CDF.



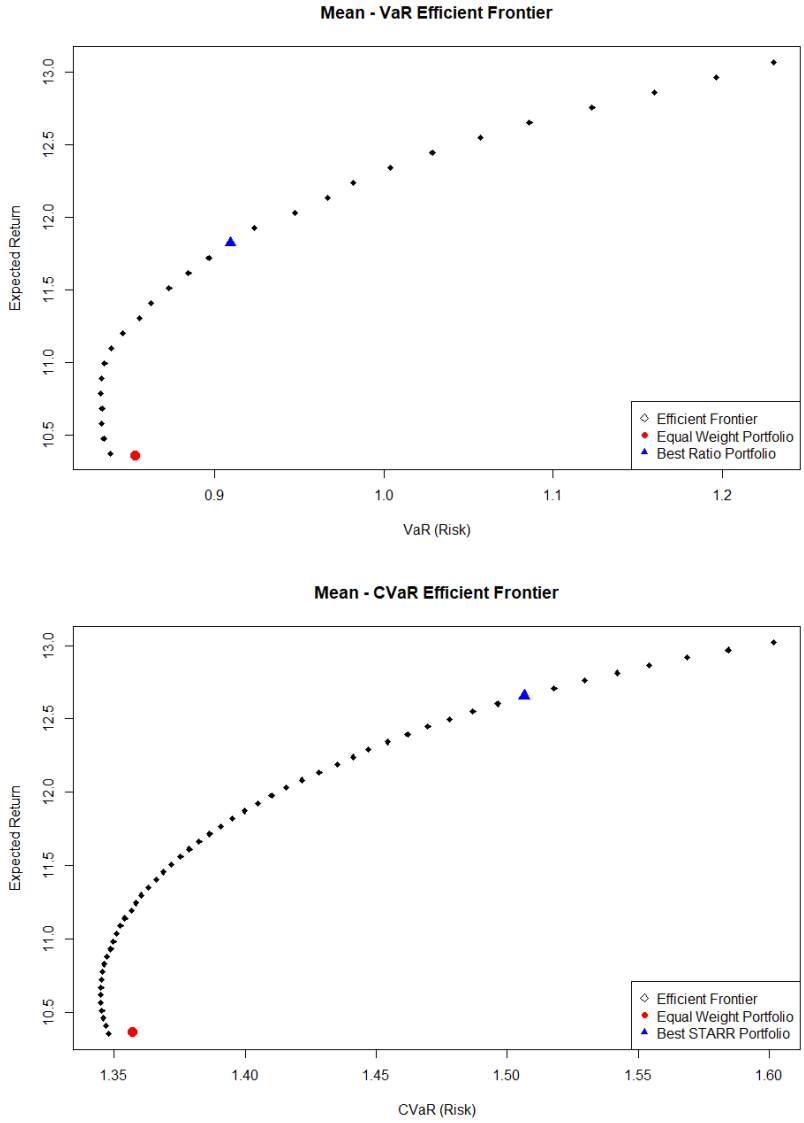


Figure B.5: *Mean-Variance, Mean-VaR_β, and Mean-$CVaR_\beta$ efficient frontiers, computed optimising the returns simulated from a Gaussian copula, and then transformed to translated exponentials by applying their inverse CDF. In blue we observe the market line, whose intersection with the frontier identifies the optimal frontier. The results are compared with the EW portfolio (red dot).*

B.2 Chapter 3

We provide some of the results discussed in Section 3.5 of Chapter 3.

B.2.1 Convergence for 2-Dimensional Copulas

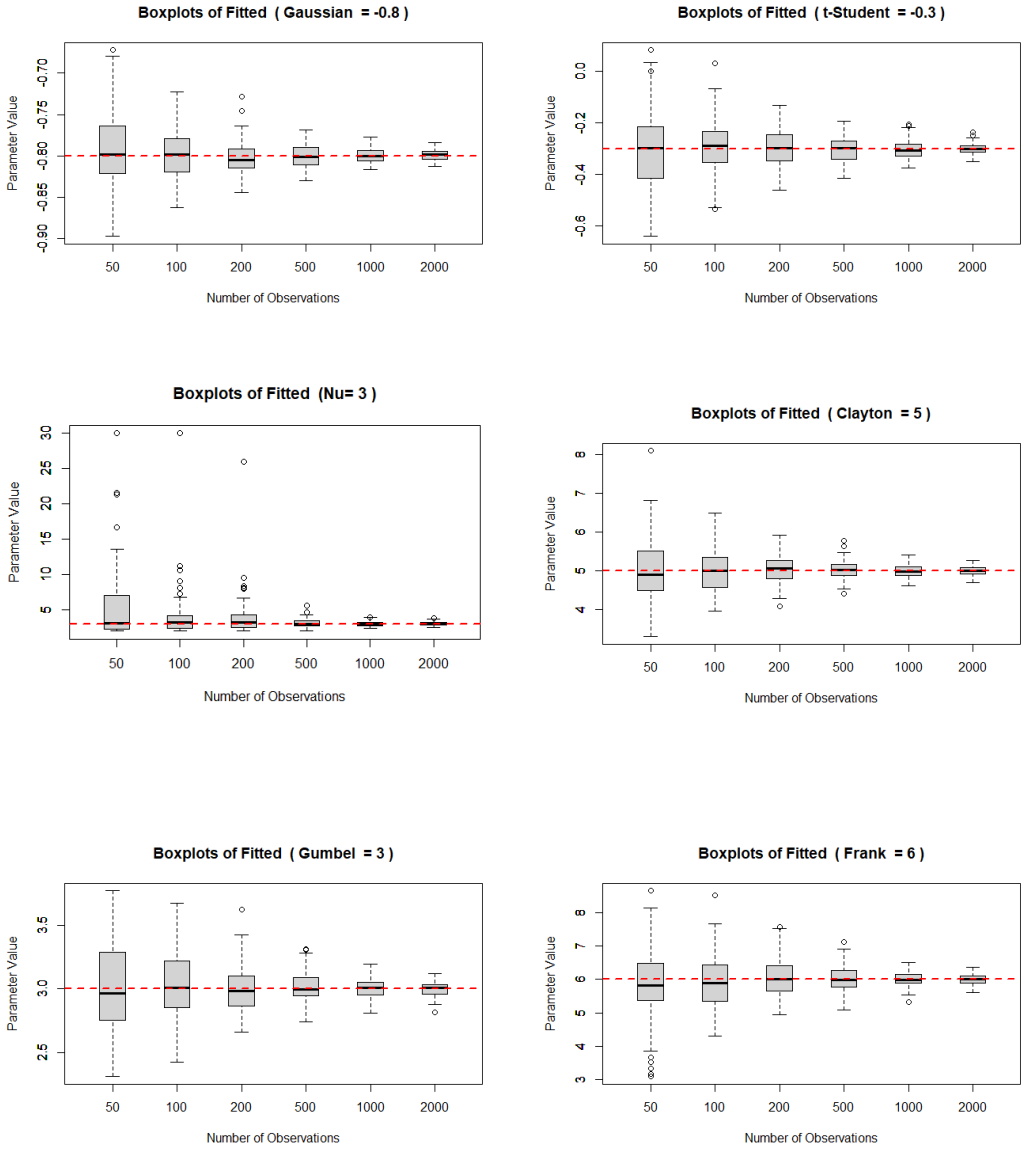


Figure B.6: *Boxplots showing the improvement of the estimated values for the parameters of each bi-dimensional copula considered as the number of observations N increases, compared with the value of the original dataset in red.*

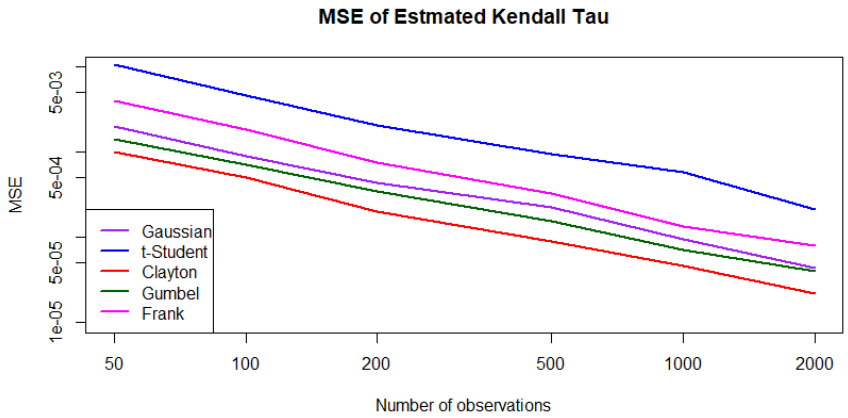


Figure B.7: *Logarithmic scaled plot studying the convergence of the estimated Kendall Tau of the fitted copulas. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error values.*

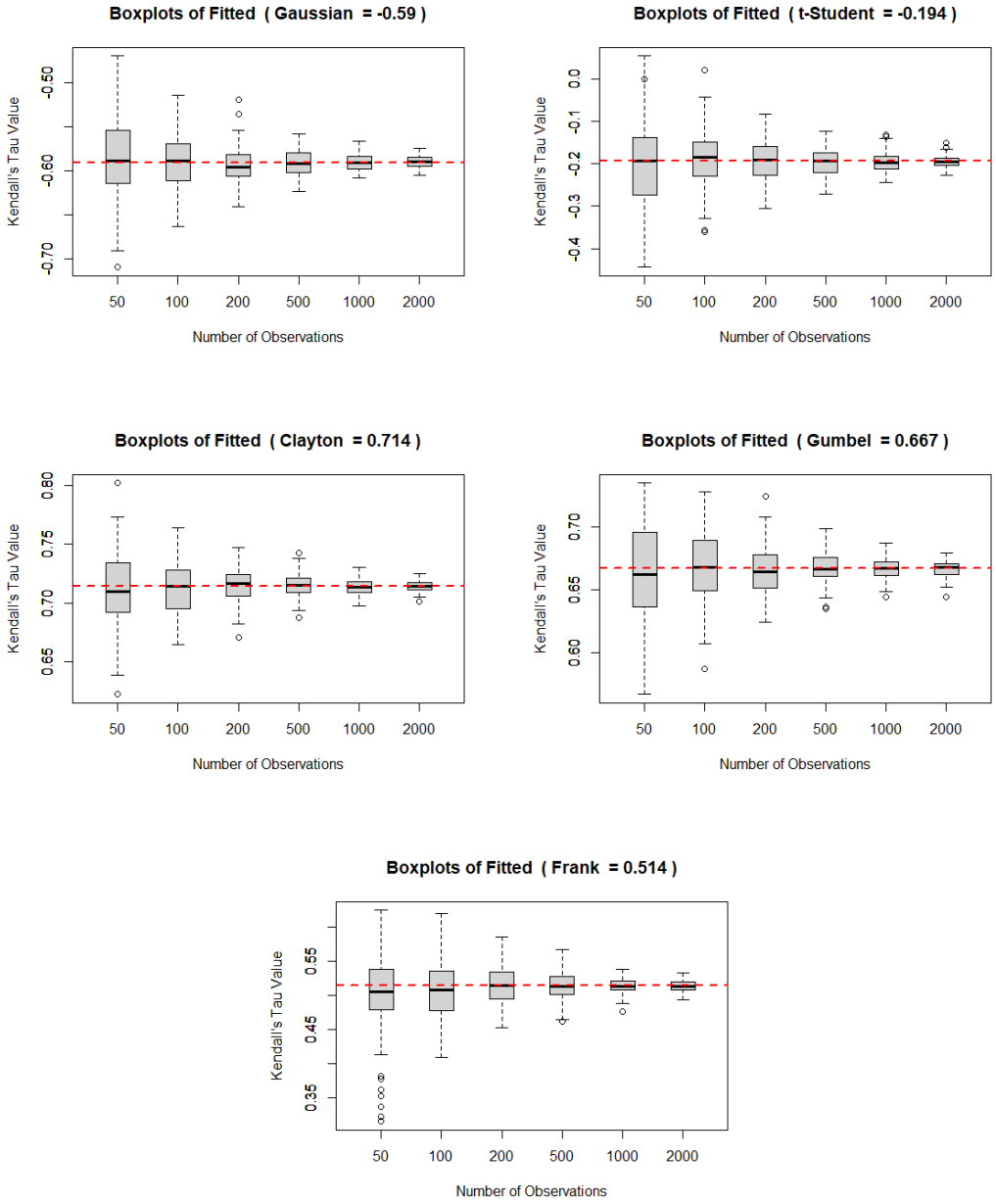


Figure B.8: *Boxplots showing the improvement of the estimated values for the Kendall's Tau of each bi-dimensional copula considered as the number of observations N increases, compared with the value of the original dataset in red.*

B.2.2 Convergence for 4-Dimensional Copulas

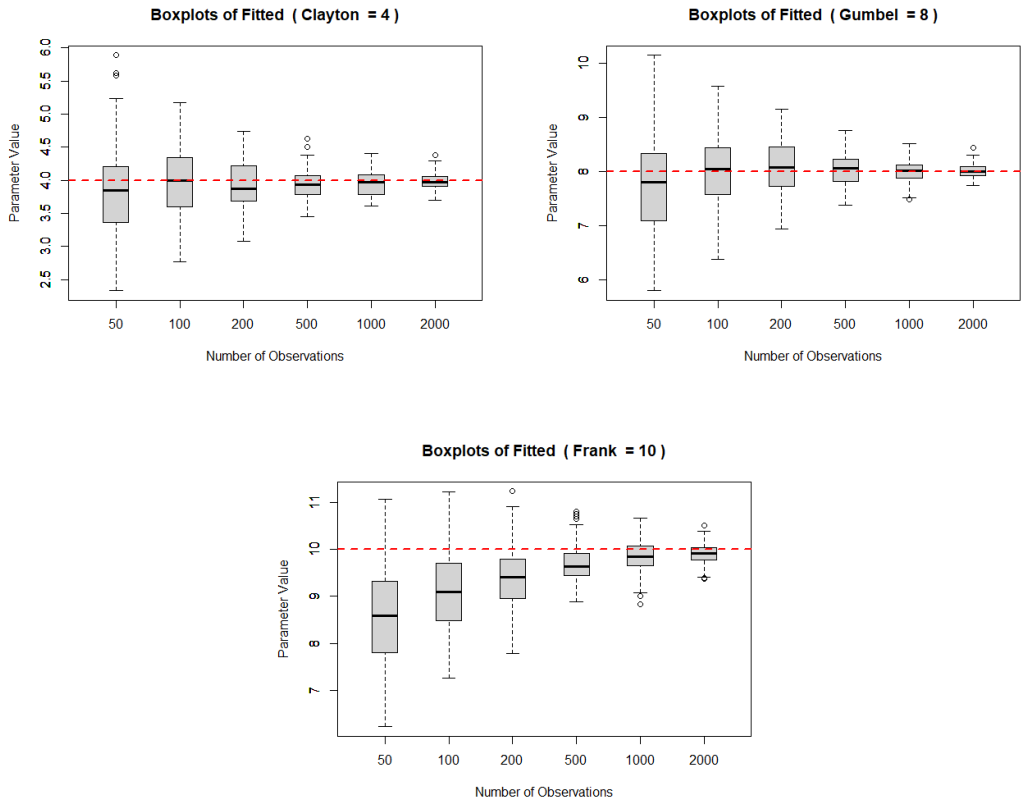


Figure B.9: *Boxplots showing the improvement of the estimated values for the parameters of each 4-dimensional copula considered as the number of observations N increases, compared with the value of the original dataset in red.*

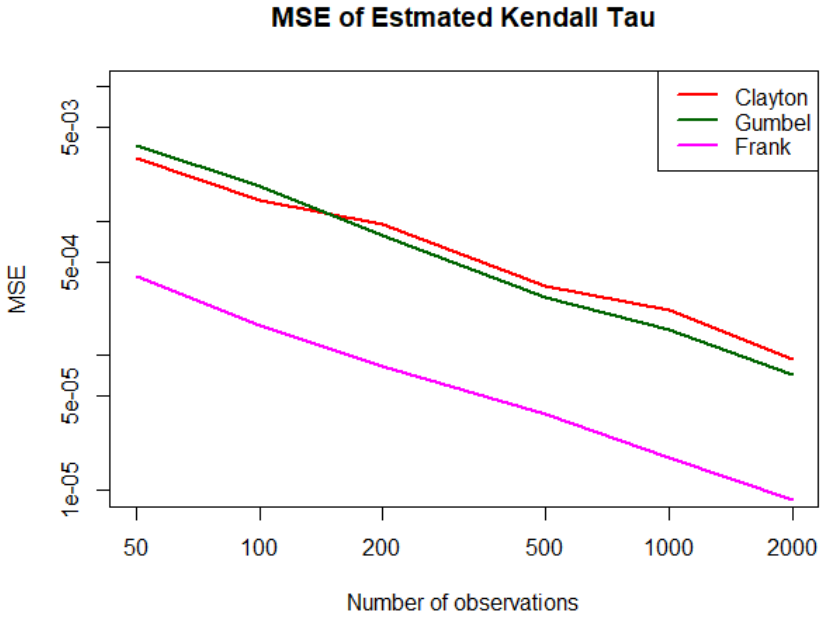


Figure B.10: *Logarithmic scaled plot studying the convergence of the estimated Kendall Tau of the fitted copulas. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error values.*

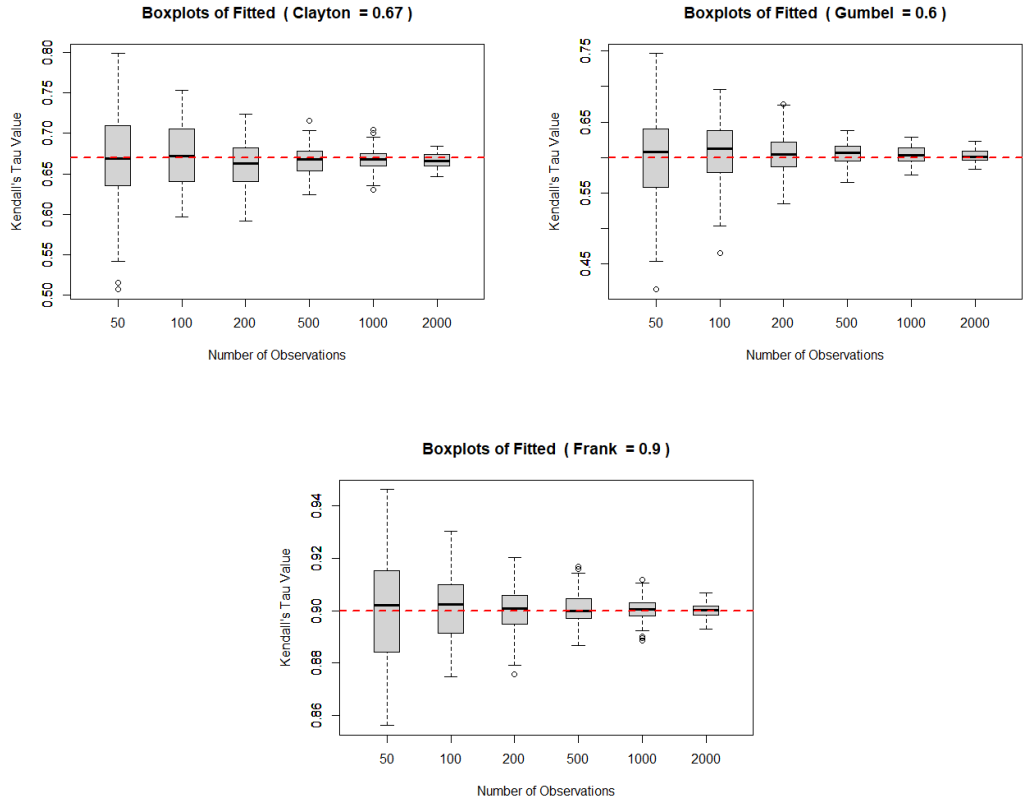


Figure B.11: *Boxplots showing the improvement of the estimated values for the Kendall's Tau of each 3-dimensional copula considered as the number of observations N increases, compared with the value of the original dataset in red.*

B.3 Chapter 4

We provide some of the results discussed in Section 4.6 of Chapter 4.

B.3.1 Estimators convergence of Vine copula

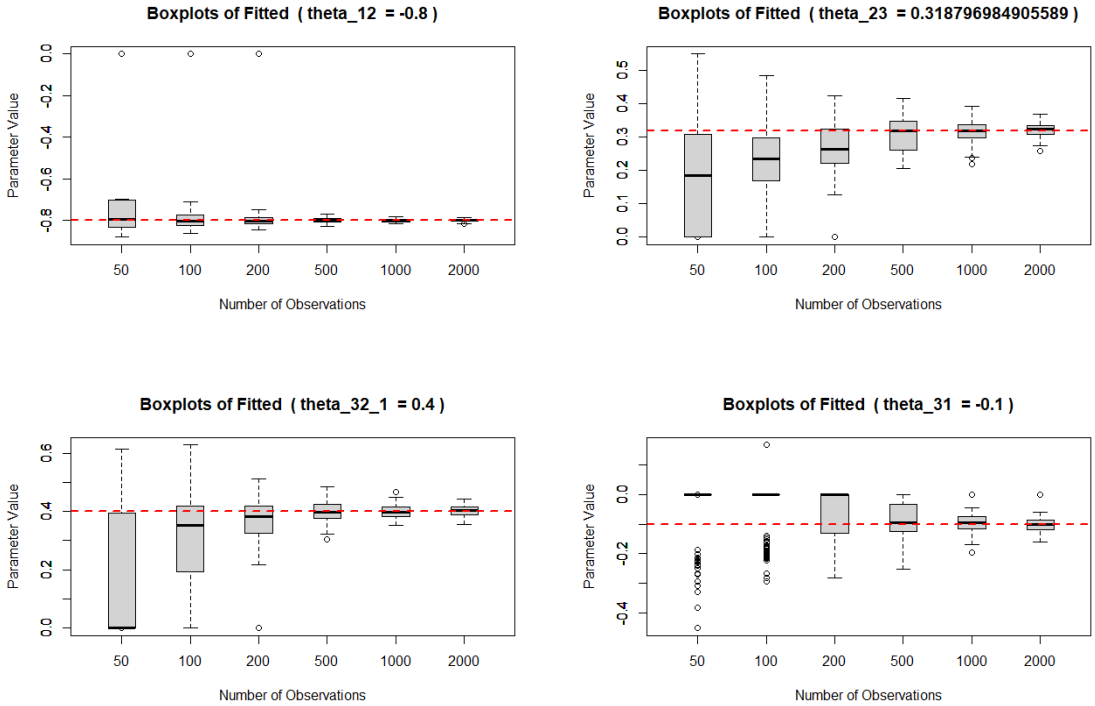


Figure B.12: *Boxplots showing the improvement of the estimated values for each parameter (and the computed θ_{23}) as the number of observations N increases.*

B.3.2 Vine structures Equivalency

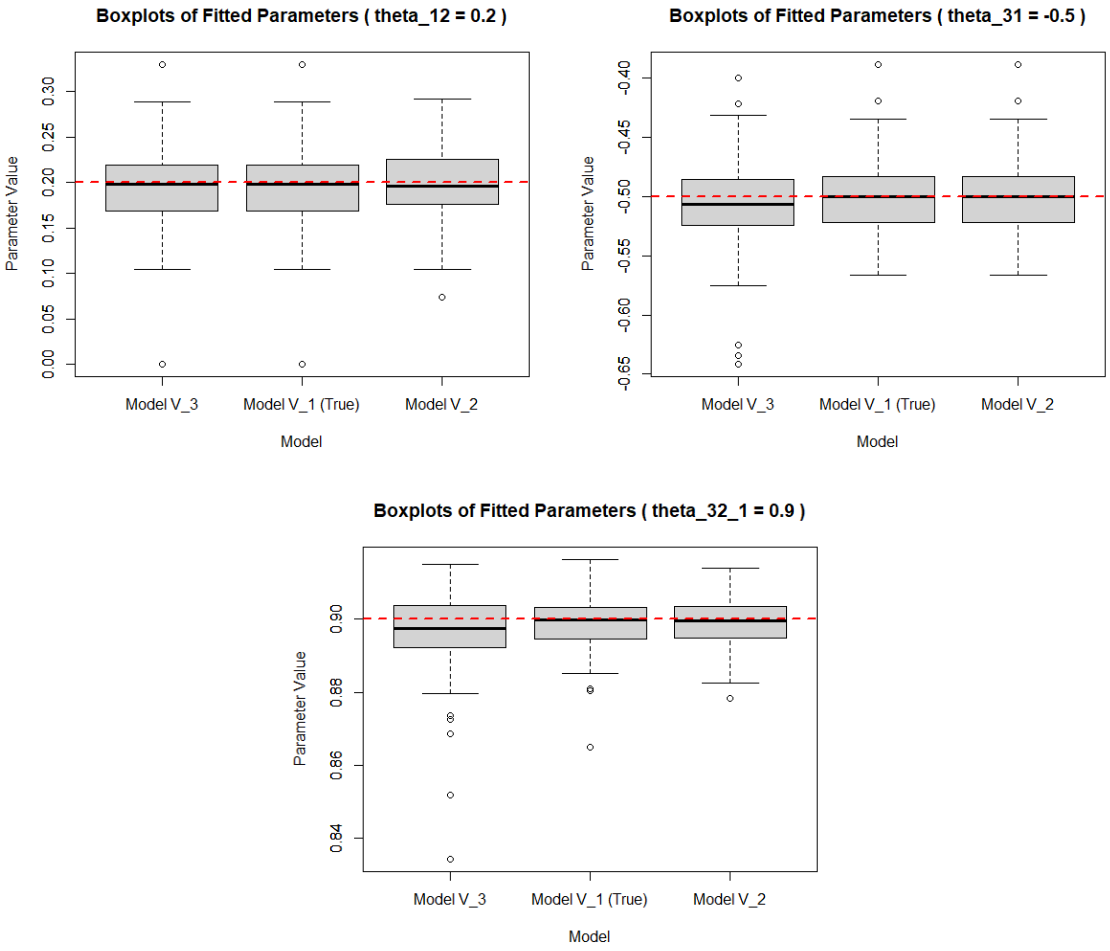


Figure B.13: *Boxplots showing the results of the simulation conducted to estimate the parameters* $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (0.2, -0.5, 0.9)$. *Each image shows the 100 fitted values by each fit generated using the three vine structures* V_3, V_1 *and* V_1, *and the true value with a red dotted line.*

B.3.3 t-Student Vine Copula

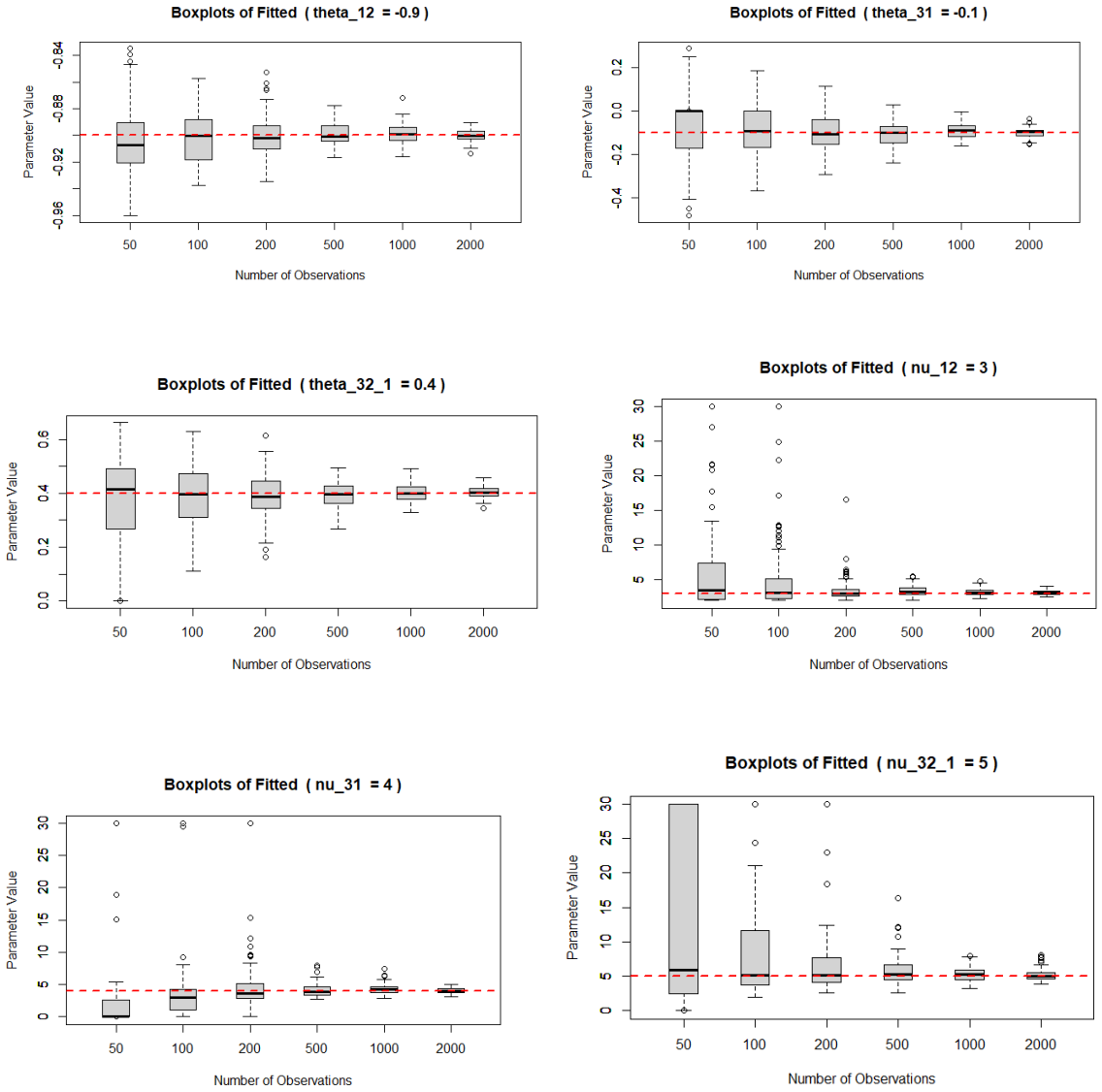


Figure B.14: *Plot of the boxplots of the improvement of the estimated values for each parameter, as the number of observations N increases.*

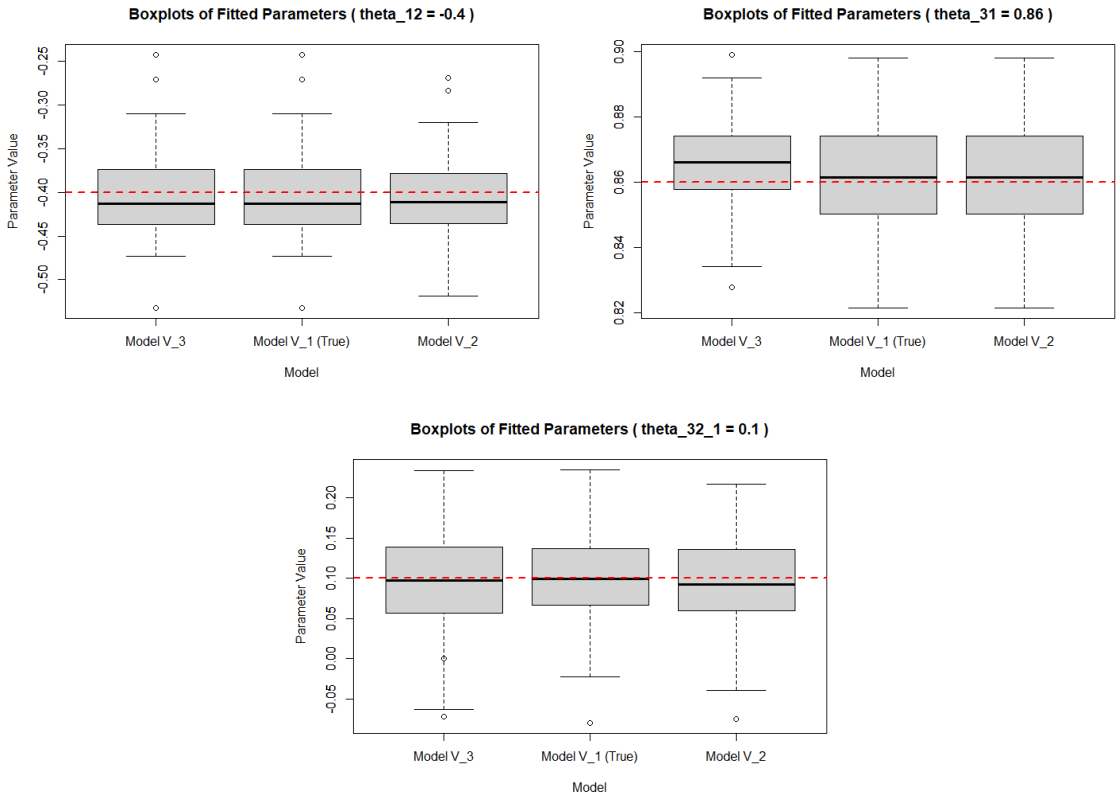


Figure B.15: *Boxplots showing the results of the simulation conducted to estimate the parameters* $(\theta_{12}, \theta_{31}, \theta_{32|1}) = (-0.4, 0.86, 0.1)$ *of a t-Student copula with* $\nu = 3$ *degrees of freedom. Each image shows the 100 fitted values, generated using the three vine structures the three vine structures* V_3*,* V_1 *and* V_1*. The true value is represented by the red dotted line.*

B.3.4 Gumbel's Vine Copula

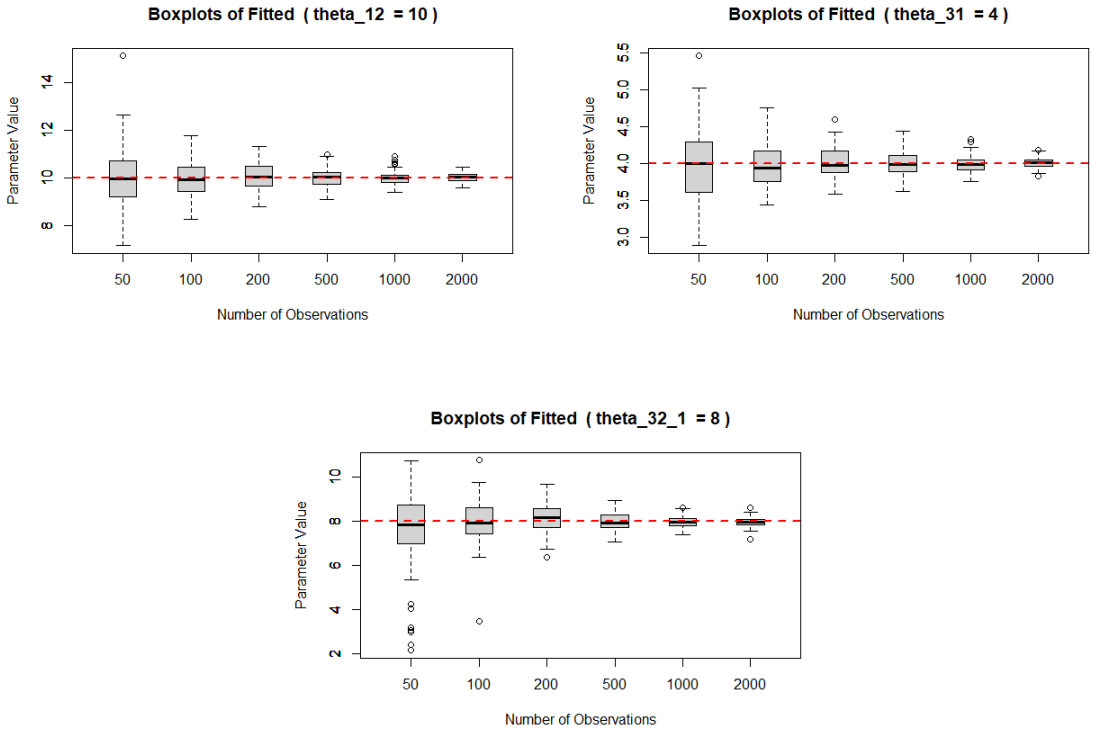


Figure B.16: *Boxplots of the improvement of the estimated values for each parameter, as the number of observations N increases.*

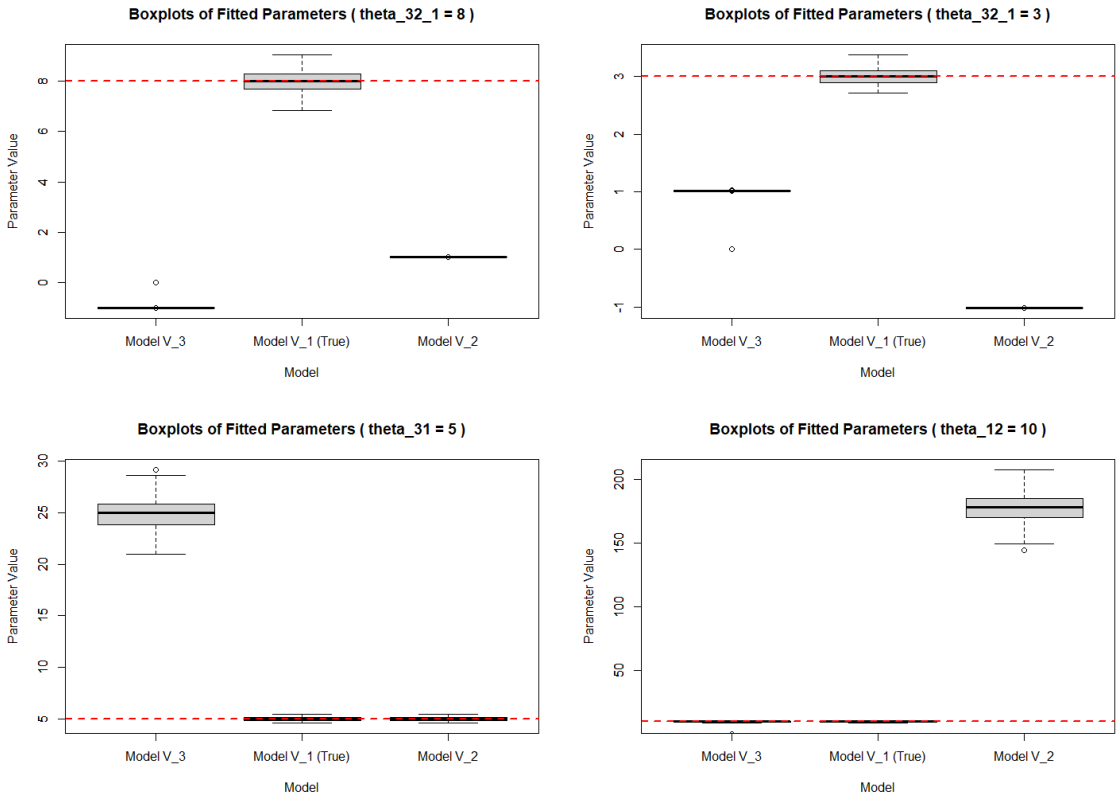


Figure B.17: *Boxplots showing the results of the simulation conducted to estimate the parameters* $\theta_{32|1} = \{1, 1, 8, 16\}$*, which are some of the most evident examples of how the correct structure influences the estimation. Each image shows the 100 fitted values by each fit generated using the three vine structures* V_3*,* V_1 *and* V_1*. The true value is represented by the red dotted line.*

C

Simulation Studies: Further Checks

In this Section of the Appendix we will run some further analysis on the procedures and functions used throughout this thesis, to make sure the code is correctly implemented and the results obtained in the practical application are obtained using a methodology that has been carefully evaluated. In particular we focus on assessing the performance of the tools employed in the simulation studies of Chapter 5 (Sect 5.8) and of Section 6.7 of Chapter 6.

C.1 Chapter 5

In this Section we analyze the steps used in the simulation study of Chapter 5, in particular we will focus on analyzing the accuracy obtained in the Estimation process, in the Selection Process, and in the Predictions.

In this case we will consider simple models that have a behaviour that can be intuitively be understood, which results in easier checks of the correctness of the procedure. In particular, we will simulate an AR(1) of parameter $\phi = 0.9$, a MA(1) with $\theta = 0.9$, an ARMA(1,1) of parameters $\phi_1 = 0.9, \theta = 0.5$, and lastly a GARCH(1,1) with $(\alpha_0, \alpha_1, \beta_1) = (1, 0.8, 0.1)$.

C.1.1 Estimation

Similarly to the study conducted for copulas and vine models in Sections 3.5 and 4.6, we verify the convergence of the MLE estimators to the original parameters of the time series considered, by changing the size of the dataset to $N = \{50, 100, 200, 500, 1000, 2000\}$. For each time series considered we provide a logarithmic plot of the MSE ((39)), and the corresponding boxplots showcasing the distribution of the estimate parameters for each size of the simulated data considered. We do not discuss each case individually, but provide general conclusions in Section C.1.1.5.

C.1.1.1 AR(1)

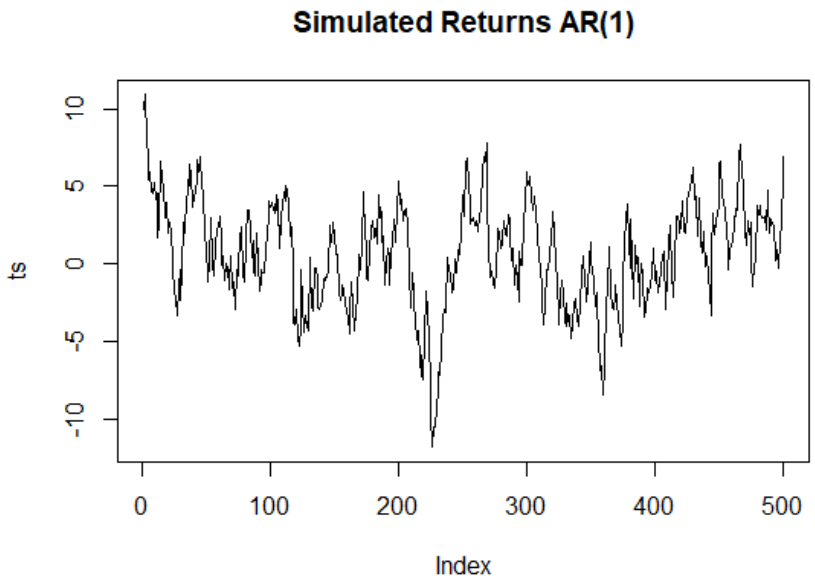


Figure C.1: *Simulated plot of an AR(1), of parameters* $\phi_1 = 0.9, \sigma = 1.5$.

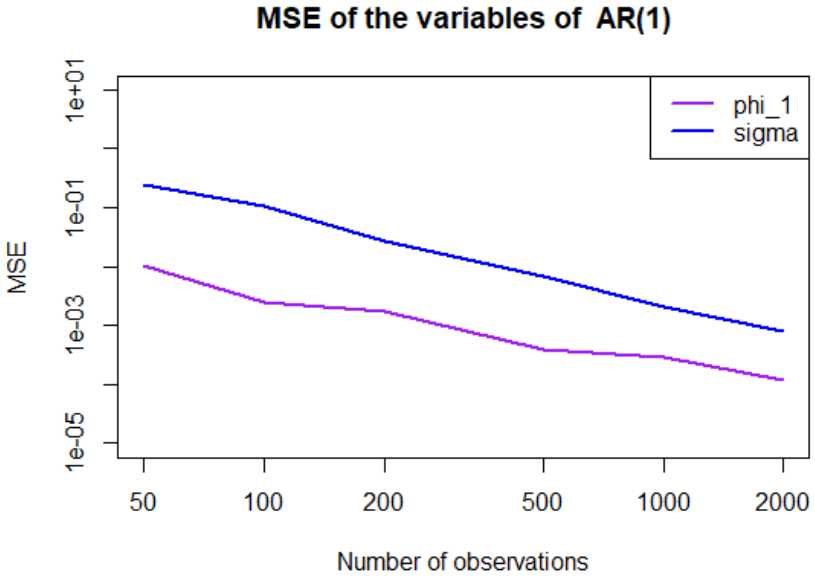


Figure C.2: *Logarithmic scaled plot studying the convergence of the estimated Parameters of the AR(1) model. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error.*

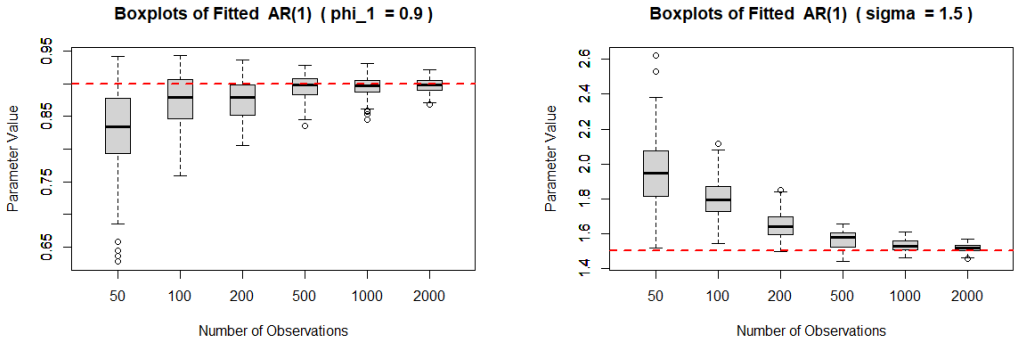


Figure C.3: *Boxplots showing the improvement of the estimated values for the parameters of the AR(1) time series considered, as the number of observations N increases, compared with the value of the original dataset in red.*

C.1.1.2 MA(1)

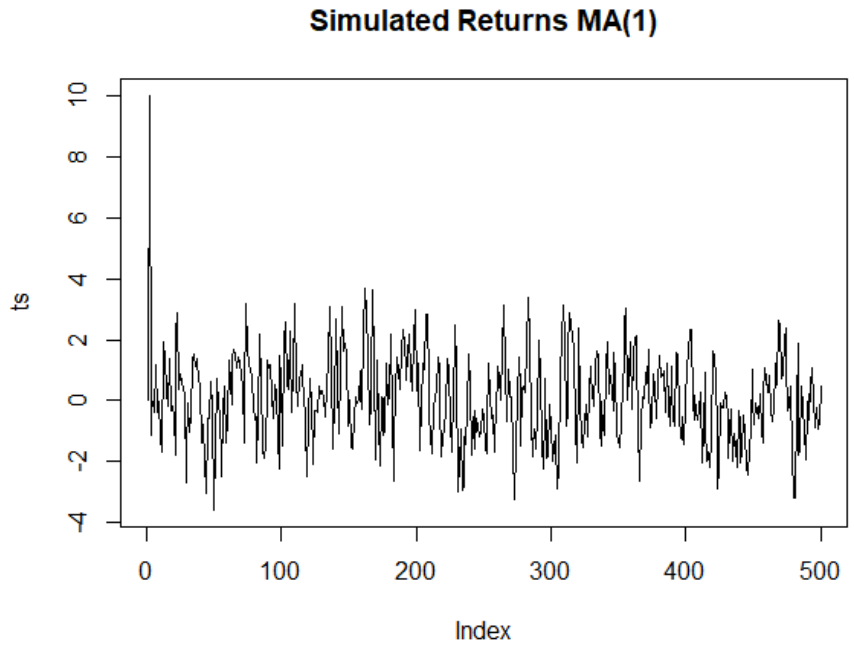


Figure C.4: *Simulated plot of an MA(1), of parameters* $\theta_1 = 0.9, \sigma = 3$.

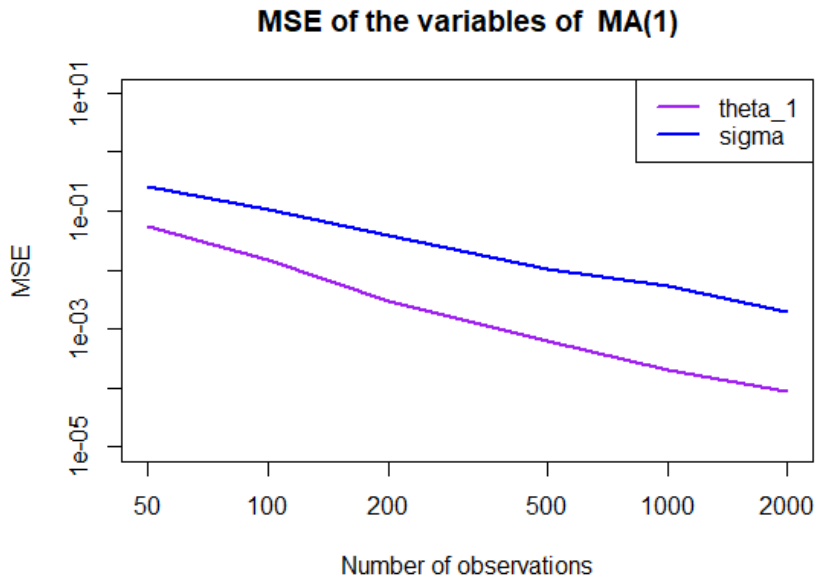


Figure C.5: *Logarithmic scaled plot studying the convergence of the estimated Parameters of the MA(1) model. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error.*

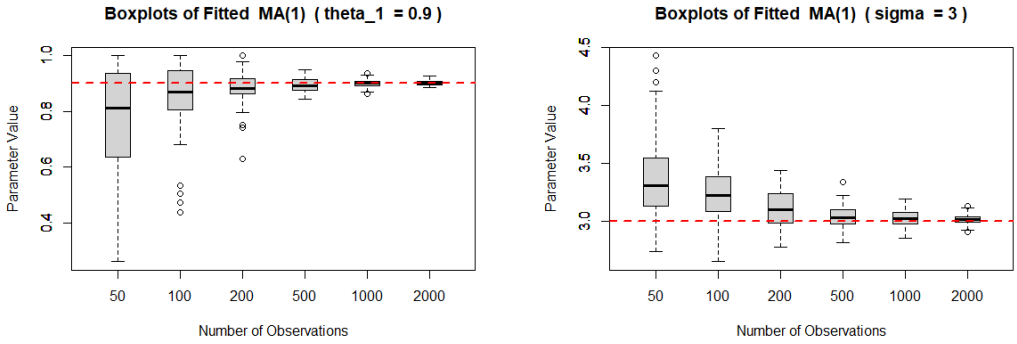


Figure C.6: *Boxplots showing the improvement of the estimated values for the parameters of the MA(1) time series considered, as the number of observations N increases, compared with the value of the original dataset in red.*

C.1.1.3 ARMA(1,1)

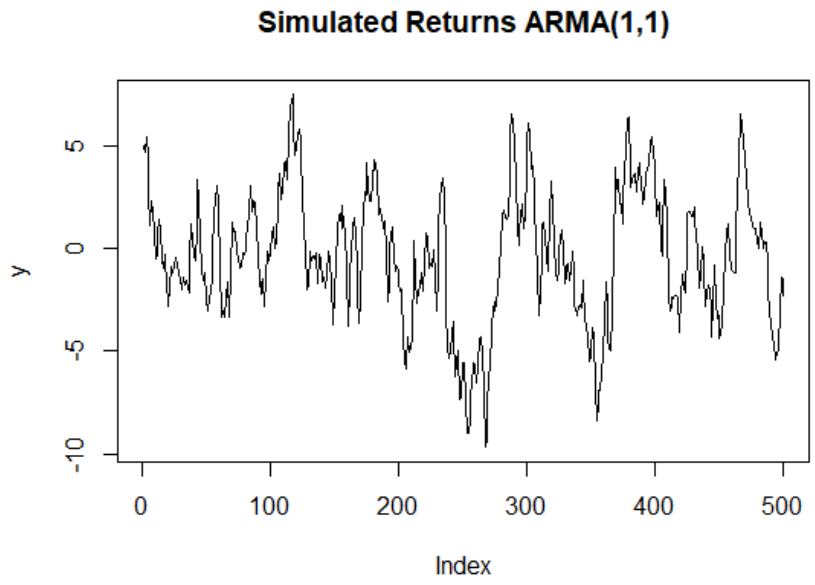


Figure C.7: *Simulated plot of an AR(1), of parameters* $\phi_1 = 0.9, \sigma = 1.5$.

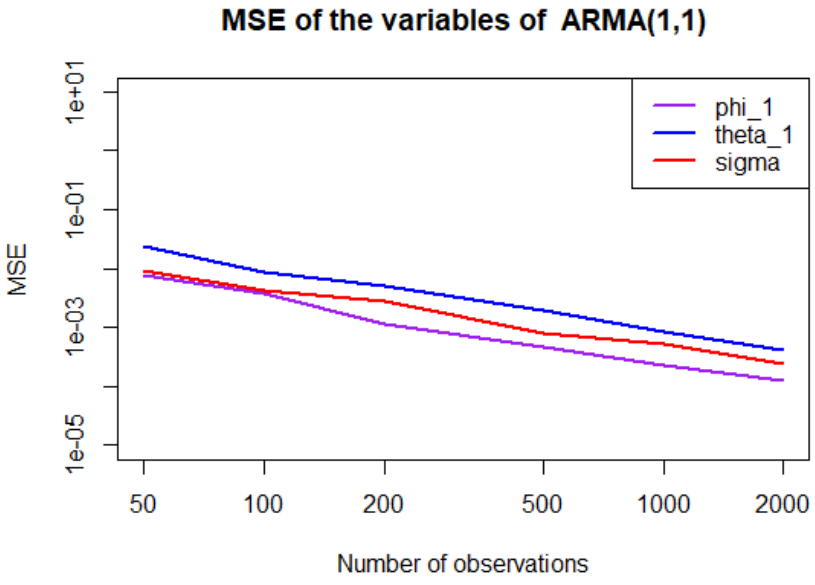


Figure C.8: *Logarithmic scaled plot studying the convergence of the estimated Parameters of the ARMA(1,1) model. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error.*

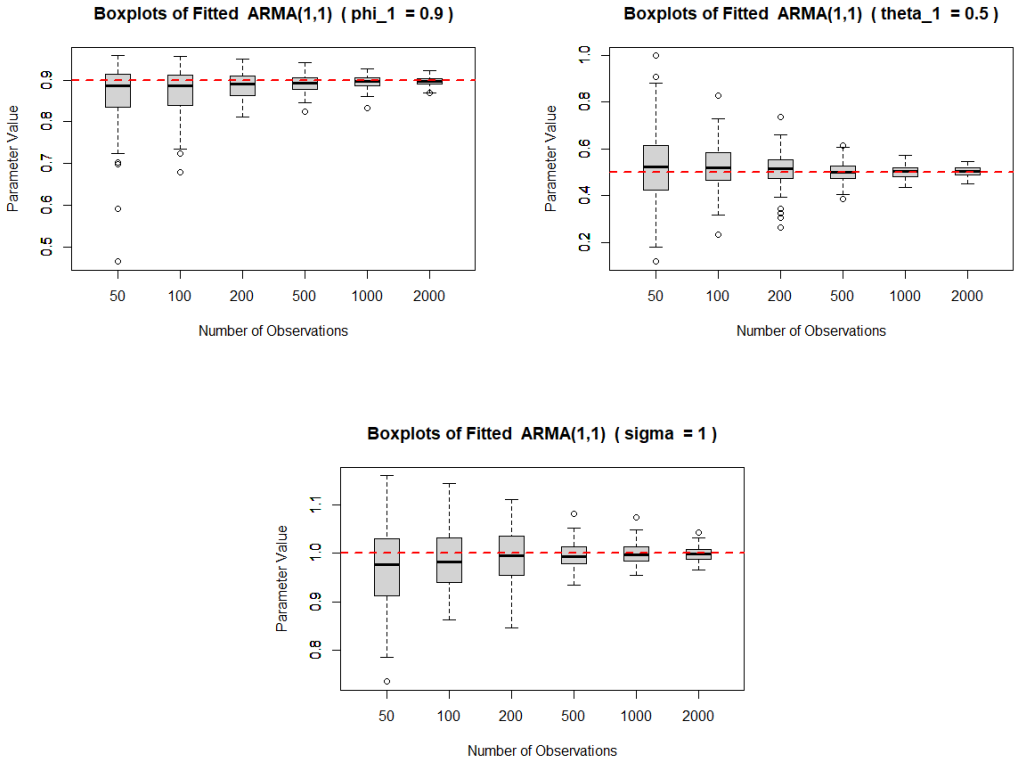


Figure C.9: *Boxplots showing the improvement of the estimated values for the parameters of the ARMA(1,1) time series considered, as the number of observations N increases, compared with the value of the original dataset in red.*

C.1.1.4 GARCH(1,1)

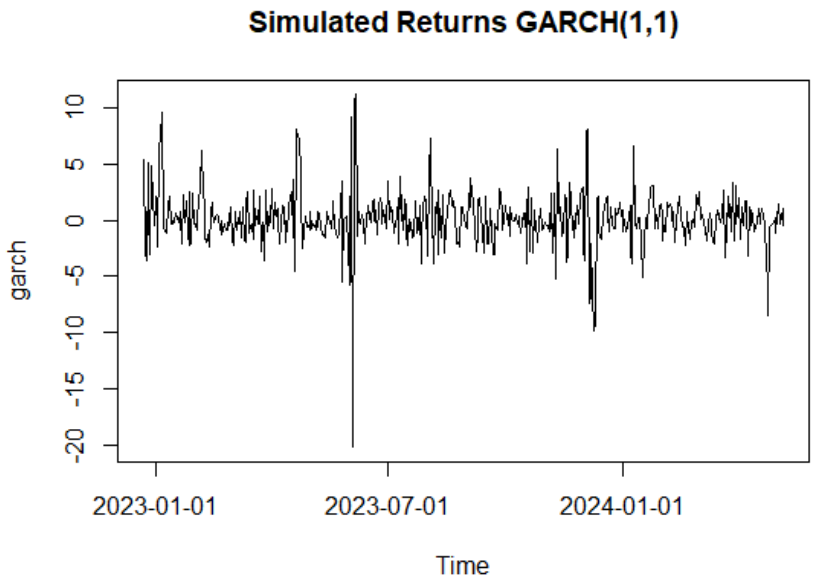


Figure C.10: *Simulated plot of an GARCH(1,1), of parameters* $\phi_1 = 0.9, \sigma = 1.5$.

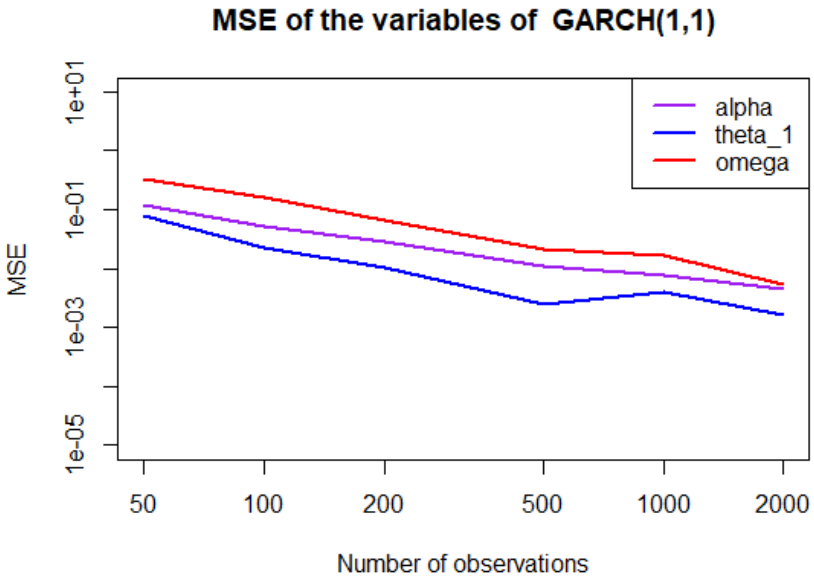


Figure C.11: *Logarithmic scaled plot studying the convergence of the estimated Parameters of the GARCH(1,1) model. On the x-axis we have the number of observations, on the y-axis the corresponding Mean Squared Error.*

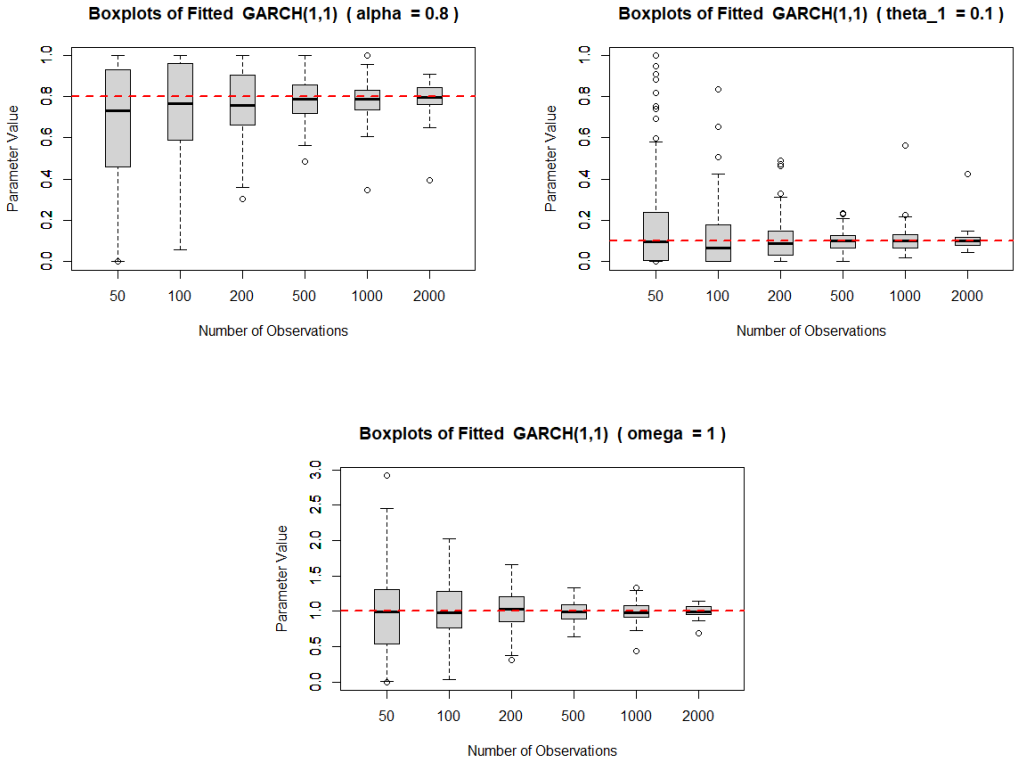


Figure C.12: *Boxplots showing the improvement of the estimated values for the parameters of the GARCH(1,1) time series considered, as the number of observations N increases, compared with the value of the original dataset in red.*

C.1.1.5 General comments on Estimation

For all models considered, we can observe how the length of the dataset significantly influences the efficacy of the estimators. In particular, it is possible to notice how $N = 500$ is the minimum size required to obtain reliable estimations of the parameters: this will be similar to the size considered in the practical application of Chapter 8. Nevertheless, we can consider this experiment as successful, as we observe a convergence of the estimators to the true parameters for all time series considered.

C.1.2 Model Selection

In this Section, we analyze the efficiency of comparing different fits through the comparison of the AIC values. Generally the selection proved to be accurate. In particular, here we provide the results for the MA(1) and the GARCH(1,1) case.

In Table C.1, the percentage of times that an MA(q) model is selected when fitting the MA(1) time series simulated with parameters $\theta_1 = 0.9, \sigma = 3$, for each dataset length N considered. We can see how, especially for longer datasets, the correct time series family is selected, however sometimes a more complex model than the original ($q > 1$) is preferred. This is one of the disadvantages of using AIC, rather than BIC. However, the results are still satisfactory and this criteria is much easily found in R's built in functions, that we decided to keep using this criteria for our implementation.

Better results are achieved for the GARCH case, whose results are reported in Table C.2, where a

GARCH model is selected 100% of times for all datasets considered with length greater than $N = 100$.

C.1.2.1 MA(1)

N	100	200	500	1000	2000
MA(1)	41	52	61	64	68
MA(q),q>1	45	40	33	32	26
TOT	86 %	92 %	94 %	97 %	95 %

Table C.1: *Percentage of times that the model selection process selected the original model (MA(1)), or correct time series, but of incorrect order.*

C.1.2.2 GARCH(1,1)

N	100	200	500	1000	2000
GARCH(1,1)	98 %	100 %	100 %	100 %	100 %

Table C.2: *Percentage of times that the model selection process selected the original model (GARCH(1,1)).*

C.1.3 Predictions

In order to check the implementation of the predictions, we proceeded to make 1 day ahead predictions using the original parameters (instead of the fitted ones used in the simulation study) of the AR(1), the MA(1) and the AR(1)-GARCH(1,1) considered also for the previous steps. The results are presented below.

C.1.3.1 AR(1)

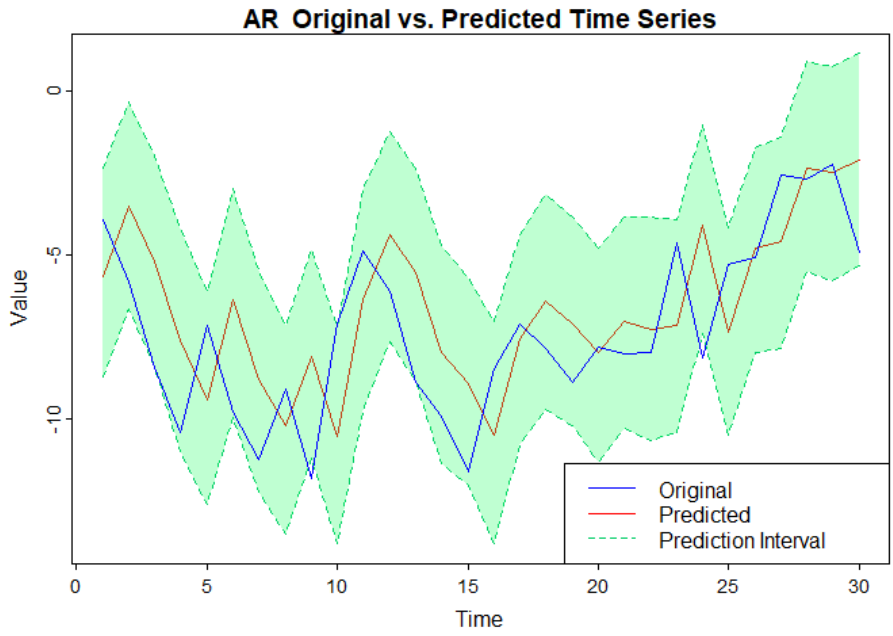


Figure C.13: *1-day-ahead predictions of a AR(1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

The prediction appears to be following the behaviour of the original time series rather well, which is coherent with the high value of the parameter ϕ_1, which suggests strong influence of the past. The predicted time series results to have an $rRMSE = 0.059$, which is an extremely good results. Moreover, 83% of the times the true value is located withing the 90% prediction interval, the result is a little lower than expected but still satisfactory.

C.1.3.2 MA(1)

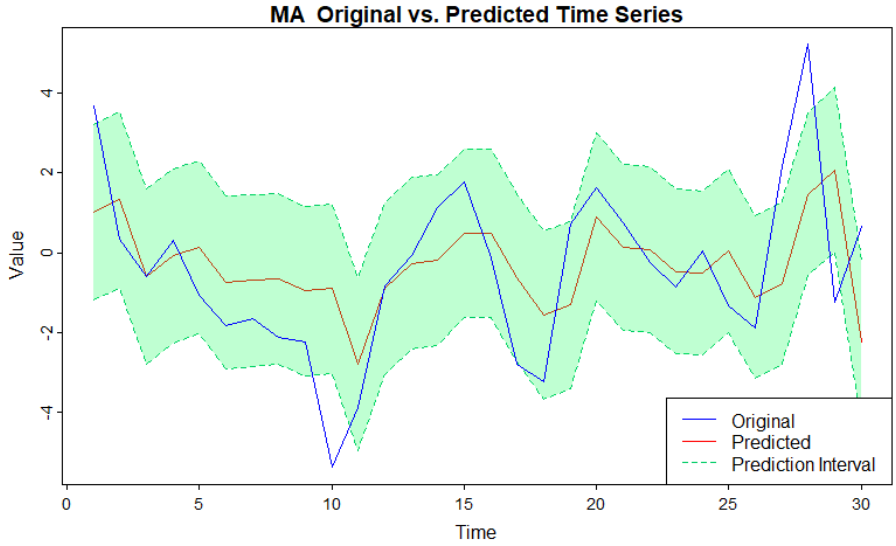


Figure C.14: *1-day-ahead predictions of a MA(1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

The $rRMSE$ for the MA(1) prediction is equal to 1.66, which is still a good result considering that a moving average is influenced by past observations only though the value of the white noise. In this case only 75.86% of times the original time series is contained within the prediction interval, value that is a little lower than the expected 90%.

C.1.3.3 AR(1)-GARCH(1,1)

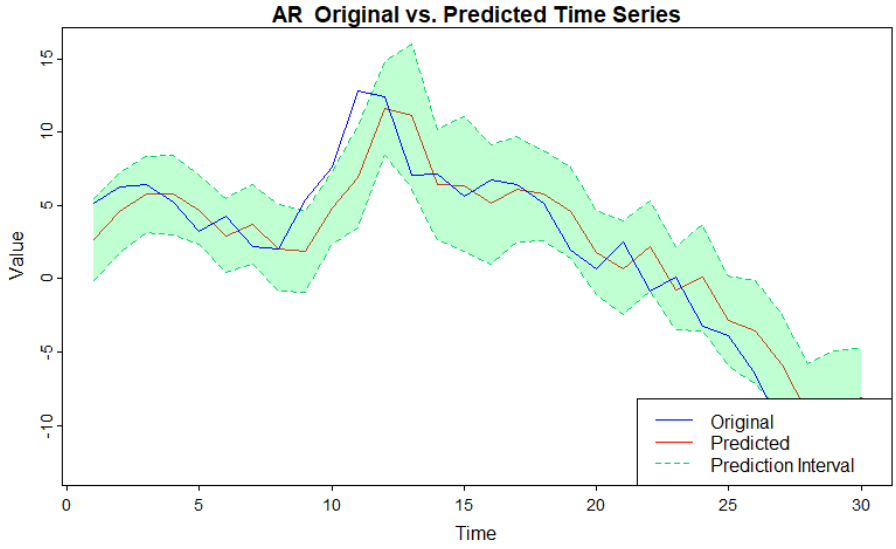


Figure C.15: *1-day-ahead predictions of a AR(1)-GARCH(1,1) time series. The predictions are compared with the true realization of the original dataset. The green area represents the* 90% *prediction interval at each time point, given the past.*

Lastly, for the AR(1)-GARCH(1,1) case, the $rRMSE = 0.065$, which is a good result. Analogously to the AR case, the influence of the past due to a high ϕ_1 value reflects positively to the predictions. Moreover, 83.3% of the times the true value is located withing the 90% prediction interval, the result is a little lower than expected but still satisfactory.

C.2 Chapter 6

Expanding on the results presented in Section 6.7.1.6 of the Simulation study of Chapter 6, we repeated the experiment for a GARCH(1,1), an AR(1)-GARCH(1,1) of different parameters than the ones considered in Chapter 6, and for an ARMA(1,1)-GARCH(1,1). The goal again is to compare the distribution of the 95% percentile of the one day ahead forecasts, made starting from the last observation, and comparing the distribution of the clouds obtained from the true model, from only fitting a time series, and from applying a copula-GARCH model. The results are presented below, and will be discussed in Section C.2.3.1.

C.2.1 GARCH(1,1)

We construct the cloud of predictions after having simulated the original dataset from two GARCH(1,1) of parameters: $(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}) = (0.6, 0.7, 0.2)$, $(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}) = (0.5, 0.8, 0.1)$. The residuals are obtained from a $Clayton(10)$.

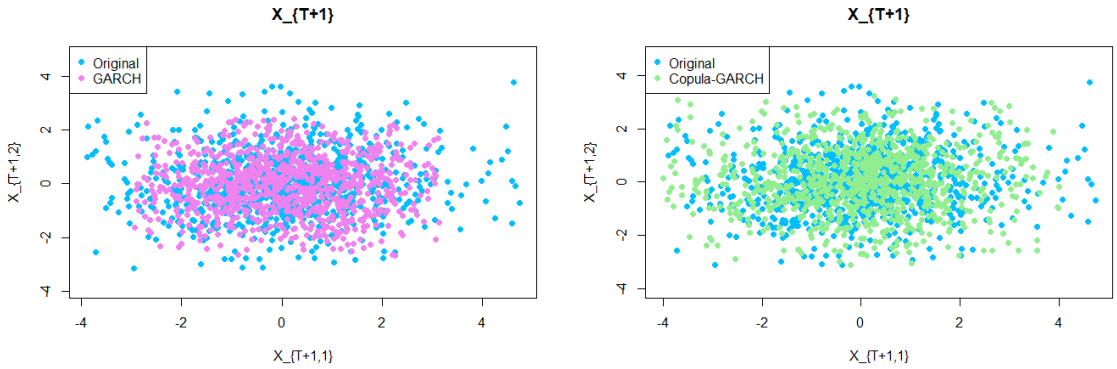


Figure C.16: *95% percentile distribution clouds from one-day-ahead predictions using* $GARCH(1,1)$ *and copula-GARCH models, compared with points sampled from the true distribution.*

C.2.2 AR(1)-GARCH(1,1)

We construct the cloud of predictions after having simulated the original dataset from two AR(1)-GARCH(1,1) of parameters: $(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}, \phi_{1,1}) = (1, 0.7, 0.2, 0.7)$, $(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}, \phi_{1,2}) = (1, 0.8, 0.1, 0.5)$. The residuals are obtained from a $Clayton(6)$.

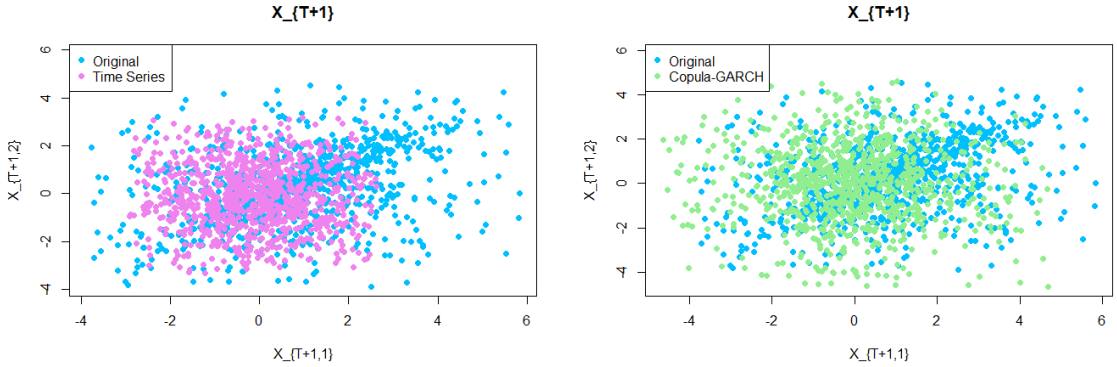


Figure C.17: *95% percentile distribution clouds from one-day-ahead predictions using* $AR(1) - GARCH(1,1)$ *and copula-GARCH models, compared with points sampled from the true distribution.*

C.2.3 ARMA(1,1)-GARCH(1,1)

Lastly, we construct the cloud of predictions after having simulated the original dataset from two ARMA(1,1)-GARCH(1,1) of parameters: $(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}, \phi_{1,1}, \theta_{1,1}) = (1, 0.7, 0.2, 0.7, 0.1)$, $(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}, \phi_{1,2}, \theta_{1,2}) = (1, 0.8, 0.1, 0.5, 0.2)$. The residuals are obtained from a $Clayton(6)$.

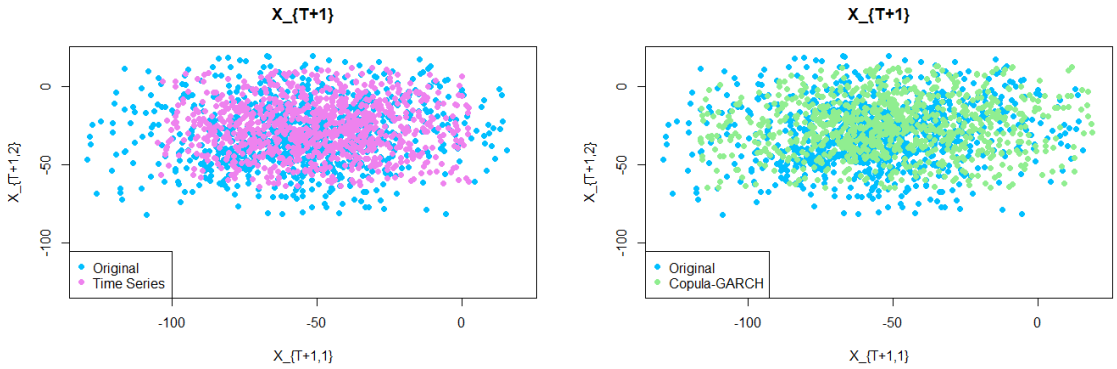


Figure C.18: *95% percentile distribution clouds from one-day-ahead predictions using* $AR(1) - GARCH(1,1)$ *and copula-GARCH models, compared with points sampled from the true distribution.*

C.2.3.1 General Comments

As for the results discussed in Sections 6.7.1.6, and 6.7.3, we can notice how fitting a proper copula model positively influences the ability of simulating more extreme values. For all time series considered, with a more evident example for the GARCH time series, we can see how the cloud of points produced by the time series fit results to be more concentrated, while fitting a proper copula to the innovations enables the cloud to spread out more, and mimic the extreme behaviour of the original model.

D

Additional Results Chapter 8

D.1 Chapter 8

In this Section of the Appendix we collect some additional tables and results obtained in Chapter 8, where we apply the model to real time data. The results of this section refer to the partitions discussed in section 8.5.1, which have been updated for the final results, as the upcoming results highlighted the need for a sufficient amount of points to guarantee a good fit on the vine model.

The partitions used are the followings:

	Dates	Length
Partition 1	24/10/2014 - 10/08/2015	187
Partition 2	11/08/2015 - 03/03/2020	1076
Partition 3	04/03/2020 - 11/10/2022	615
Partition 4	12/10/2022 - 29/06/2023	170
Partition 5	30/06/2023 - 05/07/2024	241

Table D.1: *Partitions of the dataset.*

D.1.1 Partition 1

The first partition is composed of 187 points, 158 of which in the training set, and 29 in the test set (representing 15% of the original dataset).

D.1.1.1 Cumulative Returns

We start by analyzing the cumulative returns obtained from the Mean-var portfolio, and using the copula-GARCH method on the test set, shown in Figure D.1, while the difference between the two lines is reported in Figure D.2.

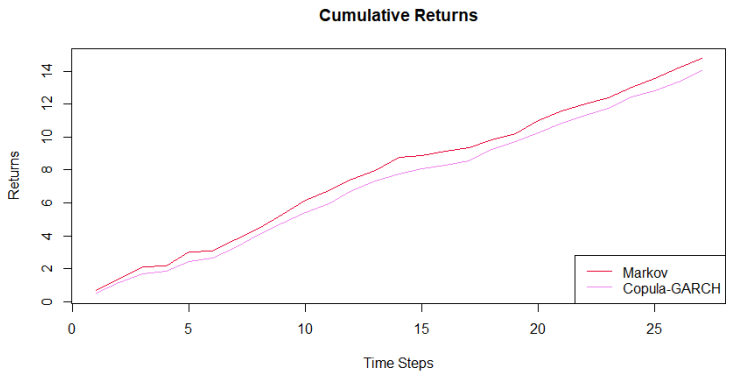


Figure D.1: *Evolution of the Cumulative Returns obtained by applying the optimal weights to the expanding test set.*

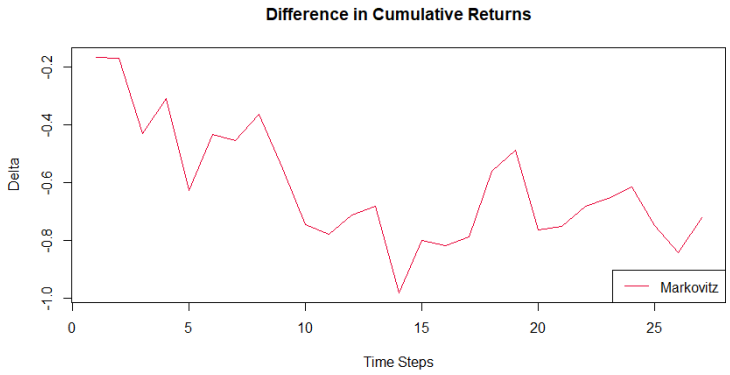


Figure D.2: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio and the ones of the Markowitz' portfolio on the tets set.*

From both pictures, we can observe how the Markowitz method is able to outperform the customized model for the entire dataset. Despite having a good p-value for the goodness of fit tests performed for each fitted model, we can conclude that for this partition the created model was not successful at optimizing the portfolio on the test period. This disadvantage might be linked to lack of data, in fact, as already anticipated, we should need around 500 days to guarantee a good convergence of the fitted parameters.

D.1.1.2 Evaluate the Portfolio

More in detail, we proceed with analysing the results obtained for the last day of the test set, $T = 187$. The corresponding fitted model is the C-vine illustrated in Figure D.3. The parameters of each bivariate copula are provided in Table D.2, where a positive dependence is clear in the bivariate copulas of the first tree, while the conditional trees appear to have rather low Kendall's tau values, which indicate rather weak conditional dependence.

The goodness of fit of the vine is assessed using the White test discussed in Section 4.5.2, obtaining a p-value of 0.805 which indicates that we cannot reject the null hypothesis, and hence that we should have a good fit. In addition, we test for the simplifying assumption (refer to Section 14), obtaining that

it cannot be rejected for any of the bivariate copulas considered.

Tree	Edge	Distribution	Parameters
T_1	DAX, SP500	Gaussian	$\rho = 0.69$, $\tau = 0.48$
	CAC, SP500	rotated BB7 (180 deg)	par 1.64, par2 $= 1.03$, $\tau = 0.46$
	N225, SP500	rotated Joe (180 deg)	par $= 1.56$, $\tau = 0.16$
T_2	DAX, N225; SP500	rotated Tawn 2 (90 deg)	par $= 1.09$, $\tau = 0.08$
	N225,CAC; SP500	Independent	$\tau = 0$
T_3	DAX,CAC; N225,SP500	rotated Tawn 2 (270 deg)	par -3.98, par2 $= 0.05$, $\tau = -0.05$

Table D.2: *Parameters of the Vine copula fitted on the first partition.*

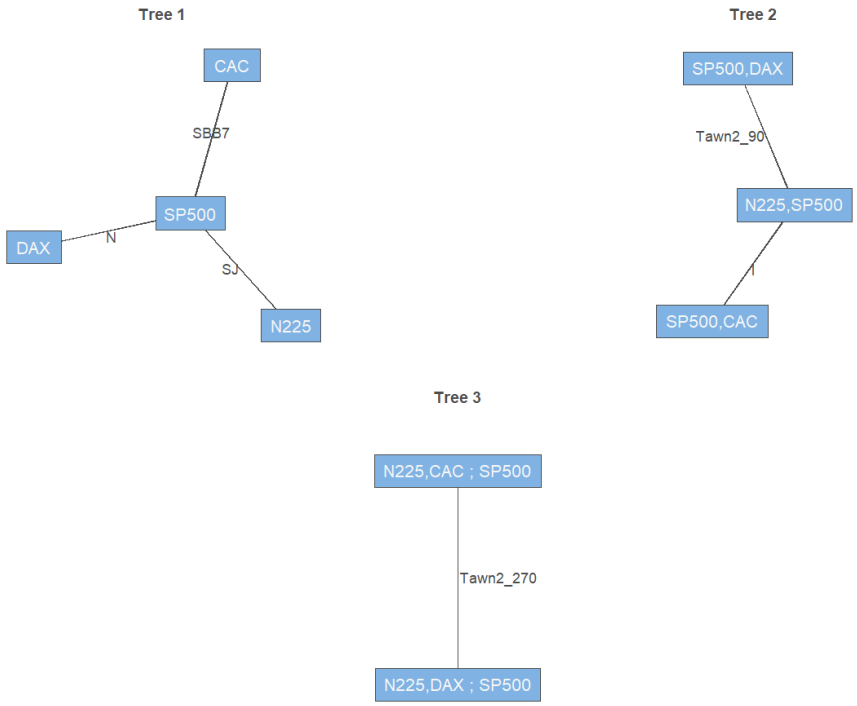


Figure D.3: *Vine structure of the fitted model on the first partition of the filtered residuals of the log-returns of the four stock indexes considered, transformed using the corresponding CDF.*

The fitted vine copula is used to obtain the optimal vine copula-GARCH portfolio; the results are compared with Markowitz's portfolio. Each optimal portfolio is obtained by finding the corresponding best ratio, with a risk free rate of $r = 0$.

The obtained weights are the following; for Markowitz:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (1.49\%, 25.52\%, 1.79\%, 71.2\%),$$

while from the vine copula Model:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (17.82\%, 0\%, 0\%, 82.18\%).$$

The corresponding metrics of the two portfolios are shown in Table D.6, and are compared to an equal weight portfolio. It is again possible to notice how Markowitz's method outperforms the other two in both the Sharpe ratio and the STARR ratio.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(25\%, 25\%, 25\%, 25\%)$	0.425	20.693	0.6%	2.1 %
Mean-Var	$(1.49\%, 25.52\%, 1.79\%, 71.2\%)$	0.826	22.205	**1.1** %	**3.7** %
COPULA	$(17.82\%, 0\%, 0\%, 82.18\%)$	0.763	23.038	0.9 %	3.3%

Table D.3: *Table comparing the portfolios obtained by optimizing the weights of two returns simulated from a copula-GARCH model, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the partition of $N = 187$ points.*

D.1.1.3 Out of Sample Evaluation

We conclude the evaluation of the model fitted on the first partition by assessing its performance on $N = 50$ days.

We apply the weights obtained, reported in Table D.3, to the four stocks, and compute the cumulative returns of the obtained portfolios. The evolution of those returns are shown in Figure D.4, while the difference of the cumulative returns between the copula-GARCH portfolio, and the other two portfolios are depicted in Figure D.5.

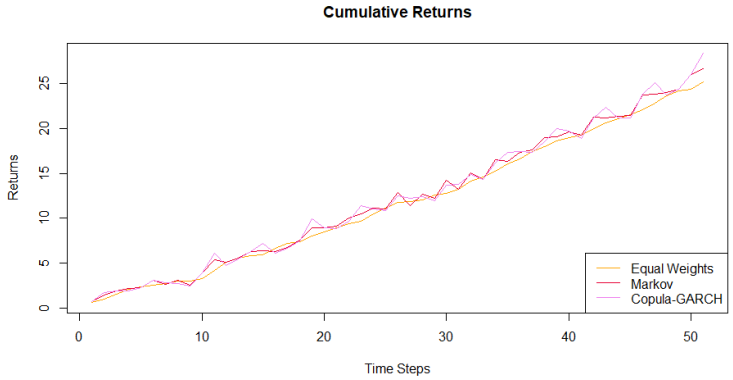


Figure D.4: *Evolution of the Cumulative Returns obtained by applying the optimal weights from different methods to $n = 50$ data points of the test set.*

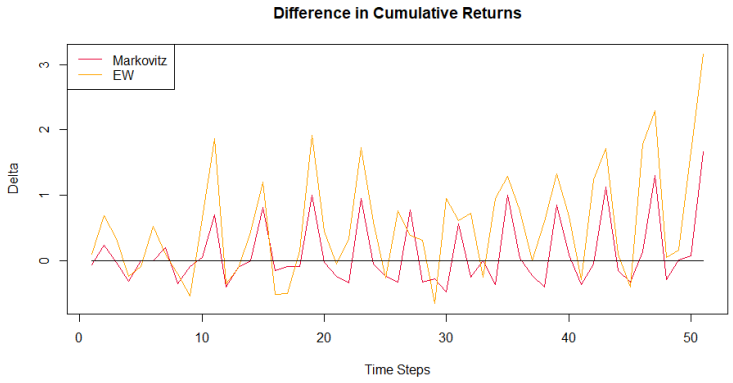


Figure D.5: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio, and the ones of the Markowitz' portfolio, or the best mean-CVaR portfolio.*

From the two images provided above, we can start to recognize an advantage in the vine copula approach compared to the other two portfolios. As already observed in the simulation study of Section 7.3, the copula-GARCH portfolio seems to perform better in the long term, as it is the only method considered that takes into account the relationship among the different financial stocks. However, this advantage is not consistent over time. Although Figure D.5 clearly shows that the positive differences highlight a significant advantage, these positive differences oscillate. This again could be connected to an inadequacy of the fitted model, due to a lack of data. In the following subsections, we will examine how an increased amount of data can improve this feature.

Lastly, in Table D.4 the cumulative returns at time $T = 50$ of the second partition are provided. Here, it is possible to notice how the vine copula-GARCH portfolio outperforms the other portfolios considered.

Method	Cumulative Return
EW	25.183
Mean-Var	26.671
COPULA	**28.339**

Table D.4: *Cumulative returns on test set, for each method at $T = 208$.*

D.1.2 Partition 2

The second partition, is characterized by a significant amount of data points, $N = 1076$, as it covers the period between $11/08/2015 - 03/03/2020$. The amount of data should be sufficient to guarantee a satisfactory fit for the vine model. For our analysis, the dataset is split into a training set and a test set, each consisting of $N = 538$ points (50% of the dataset). As we progress, the test set data will be incrementally included in the model using an expanding time window approach.

D.1.2.1 Cumulative Returns

Firstly, we analyze the overall performance of the vine copula model on the test set, constructed daily by fitting a proper vine copula model, finding the corresponding mean-CVaR efficient frontier with the Monte Carlo allocation method, and by identifying the portfolio with the optimal STARR ratio. The results are then compared the portfolio obtained by optimizing the mean-variance efficient frontier and shown in Figure D.6, while in Figure D.7 the difference between the cumulative returns of the two

portfolios are reported.

It is again clear how Markowitz's method is much better at identifying the correct set of weights on a daily basis. However, does this still hold in the long term? We will address this question in the following Subsection.

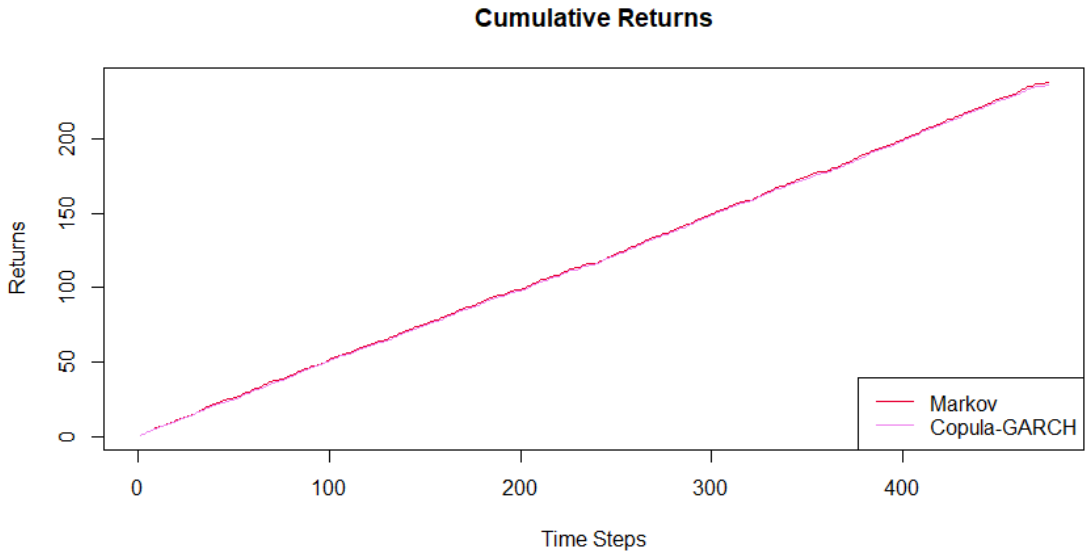


Figure D.6: *Evolution of the Cumulative Returns obtained by applying the optimal weights to the expanding test set.*

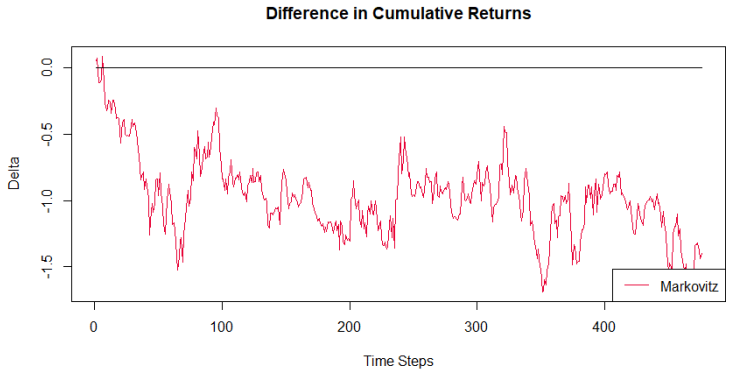


Figure D.7: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio and the ones of the Markovitz' portfolio on the tets set.*

D.1.2.2 Evaluate the Portfolio

In this Section we will discuss the portfolios obtained by fitting a model to the entire partition 2. The resulting weights will then be applied on an out of sample dataset.

The C-vine fitted to the transformed residuals of the second partition is represented in Figure D.8, and the corresponding parameters are given in Table D.5, where we can notice how the Kendall's tau of the bivariate copulas of the first tree are characterized by moderate positive relationship, which tends

to disappear in the following trees. A p-value of 0.37 of White's GoF test indicate that the fit should be a good one; moreover, testing the simplifying assumption in this partition resulted in not enough evidence to reject the assumption, which allows for a much more simple model.

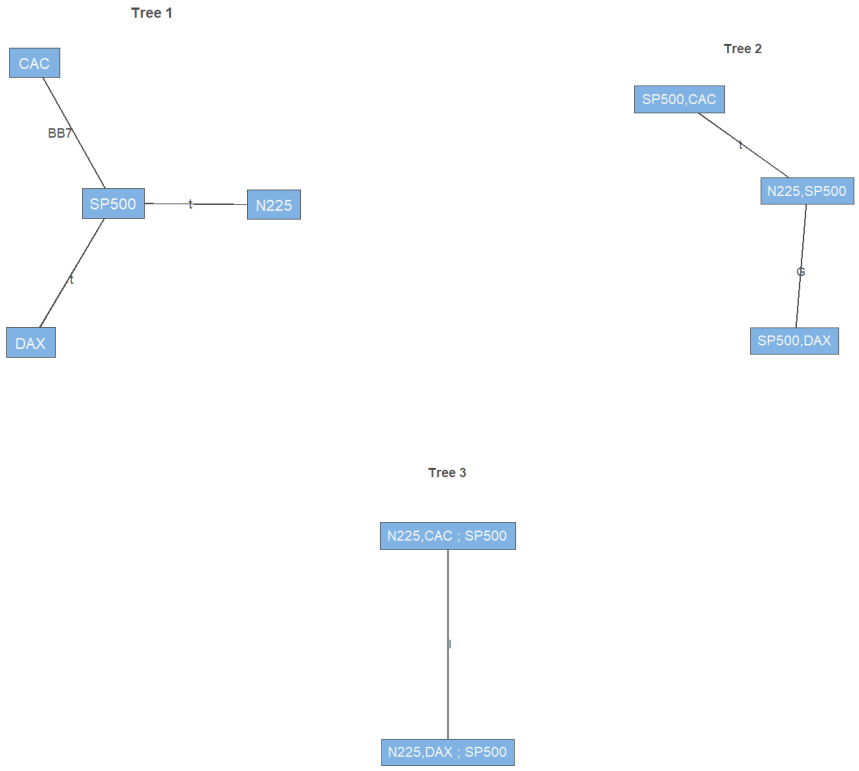


Figure D.8: *Vine structure of the fitted Vine models of the Residuals, transformed using their empirical CDF.*

Tree	Edge	Distribution	Parameters
T_1	DAX, SP500	t Student	$\rho = 0.69$, $\nu = 2.91$, $\tau = 0.49$
	CAC, SP500	BB7	par 1.56, par2 $= 0.55$, $\tau = 0.37$
	N225, SP500	t Student	$\rho = 0.28$, $\nu = 3.36$, $\tau = 0.18$
T_2	DAX, N225; SP500	Gumbel	par $= 1.09$, $\tau = 0.08$
	N225,CAC; SP500	t Student	$\rho = -0.10$, $\nu = 10.77$, $\tau = -0.06$
T_3	DAX,CAC; N225,SP500	Independent	$\tau = 0$

Table D.5: *Time Series Model Chosen for each variable.*

The fitted vine copula discussed above is used to simulate $Q = 1000$ points, which are then transformed and used to construct the efficient frontier, where the optimal portfolio has weights:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 14.8\%, 84.7\%, 0.5\%).$$

The results are compared with the optimal mean-variance portfolio obtained by selecting the weights that guarantee the best Sharpe ratio, which corresponds to:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 5.8\%, 63.3\%, 30.9\%).$$

The portfolios obtained for the different methods are compared in Table D.6. In particular, we compare the weights of the different portfolios, noting how both portfolios prefer investing in the American index (SP500), while the second index is the French index (CAC) for the vine approach, and the Japaneese one according to Markowitz's portfolio. Each set of weights is then applied to the original dataset, to compare the different measures.

The vine copula portfolio appears to be more risky than the other ones, but also characterized by greater expected value. The Mean-CVaR portfolio detains both the best Sharpe ratio (equal to Markovitz's one), the best STARR ratio. However, the copula-GARCH portfolio is still characterized by very promising results, with a Sharpe and STARR ratio very close to the best one.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(25\%, 25\%, 25\%, 25\%)$	0.295	20.259	0.5 %	1.5%
Mean-Var	$(0, 5.8\%, 63.3\%, 30.9\%)$	0.379	18.980	**0.7** %	**2.0**%
COPULA	$(0, 14.8\%, 84.7\%, 0.5\%)$	0.433	22.448	0.6	1.9%

Table D.6: *Table comparing the portfolios obtained by optimizing the weights of two returns simulated from a copula-GARCH model, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the partition of $N = 1076$ points.*

D.1.2.3 Out of Sample Evaluation

We continue with an out-of-sample efficiency evaluation of the portfolios considered. We apply the selected weights from each method to the first $N = 50$ points of the test set, computing the cumulative returns and plotting the obtained results in Figure D.9.

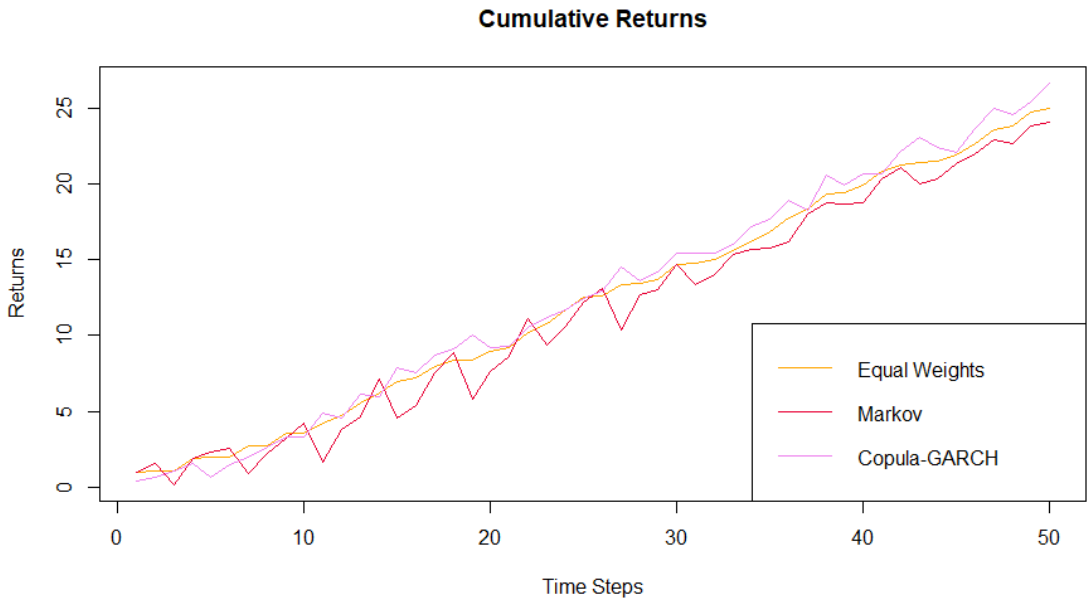


Figure D.9: *Evolution of the Cumulative Returns obtained by applying the optimal weights from different methods to $n = 50$ data points of the test set.*

In the very first period of the plot above, we can see how the cumulative returns obtained for the different methodologies are rather comparable, and the results vary at different times considered, with

a small preference for the mean-variance method, which confirms its superiority in the short term period. However, the vine copula-GARCH portfolio starts to outperform the others once the length of the dataset considered expands.

More in detail, we provide the evolution of the difference from the cumulative returns of the copula-GARCH portfolio with Markovitz's portfolio, and the classical Equal weight one in Figure D.10, where it is clear than starting from $t = 10$ of the test set, the performance of the copula-GARCH appears superior.

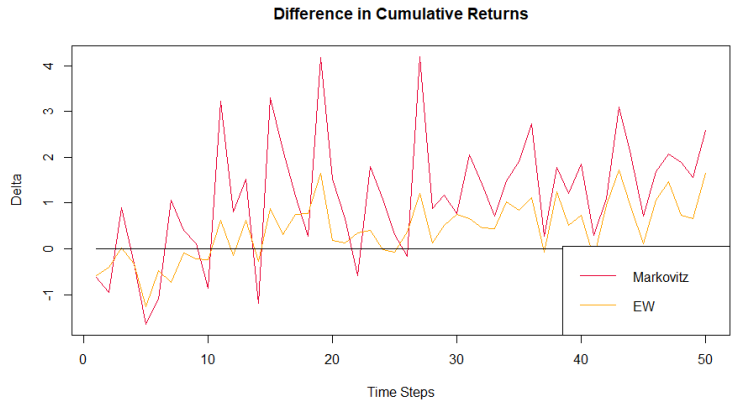


Figure D.10: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio, and the ones of the Markovitz' portfolio, or the best mean-CVaR portfolio.*

To confirm this improvement, we provide the cumulative returns on the test set at time $T_3 = 50$ in Table D.7, from which we can conclude that the vine copula-GARCH portfolio clearly outperforms the other ones.

Method	Cumulative Return
EW	25.493
Mean-Var	26.283
COPULA	**28.068**

Table D.7: *Cumulative returns on test set, for each method at $T_3 = 50$.*

D.1.3 Partition 3

The third partition, is again characterized by a sufficient amount of point to guarantee a satisfactory fit, as it has $N = 615$ observations. We split the dataset in 462 days for the training set, and 153 for the test set (25 % of the dataset).

D.1.3.1 Cumulative Returns

We start by analyzing the performance of the vine copula approach on the test set, which assigns the weights following the procedure discussed in Section 7.2. The results are presented in Figures D.11 and D.12.

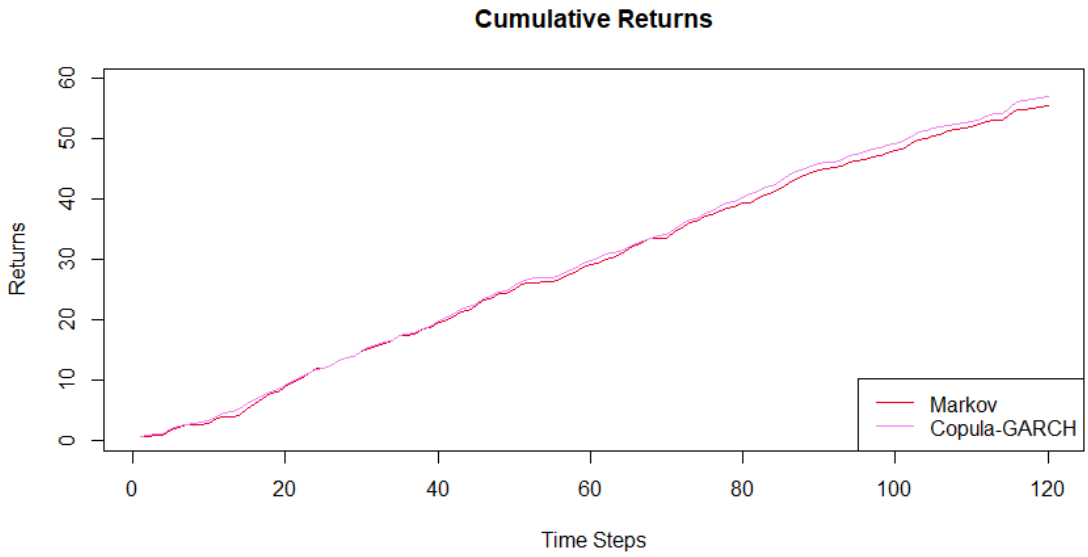


Figure D.11: *Evolution of the Cumulative Returns obtained by applying the optimal weights to the expanding test set.*

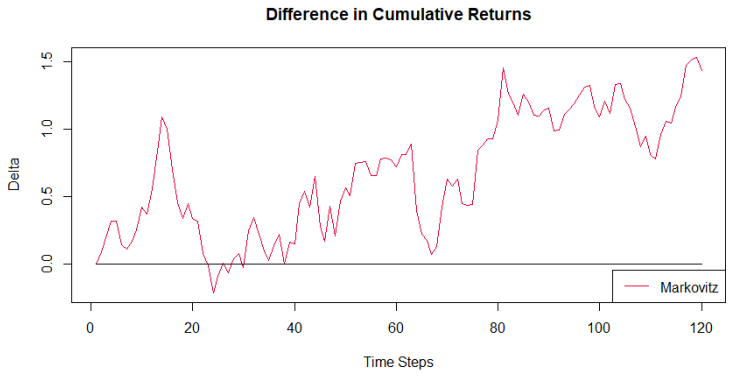


Figure D.12: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio and the ones of the Markovitz' portfolio on the tets set.*

Differently from the previous two partitions, the fitted model on the third partition is characterized by a performance that is clearly better than the one achieved by the portfolios constructed using Markowitz's method. The results are clear from the second plot, where the difference in the cumulative returns appear to be positive for all points of the test set, with the exception of a few points around $T_3 = 25$.

We proceed by analyzing the performance on the last day of the test set.

D.1.3.2 Evaluate the Portfolio

The vine fitted on the third partition is characterized by the structure shown if Figure D.13, and with the parameters of Table D.8. The fitted model obtains a p-value equal to 0.155 on the White test, which suggests a goodness of fit. Moreover, the simplifying assumption is not rejected for any copula

considered.

From the Kendall's tau we can notice that we have a moderate positive correlation among the variables that compose the first tree, while the correlation in the higher trees is almost zero, suggesting that there is not a strong conditional relationship among the different geographies.

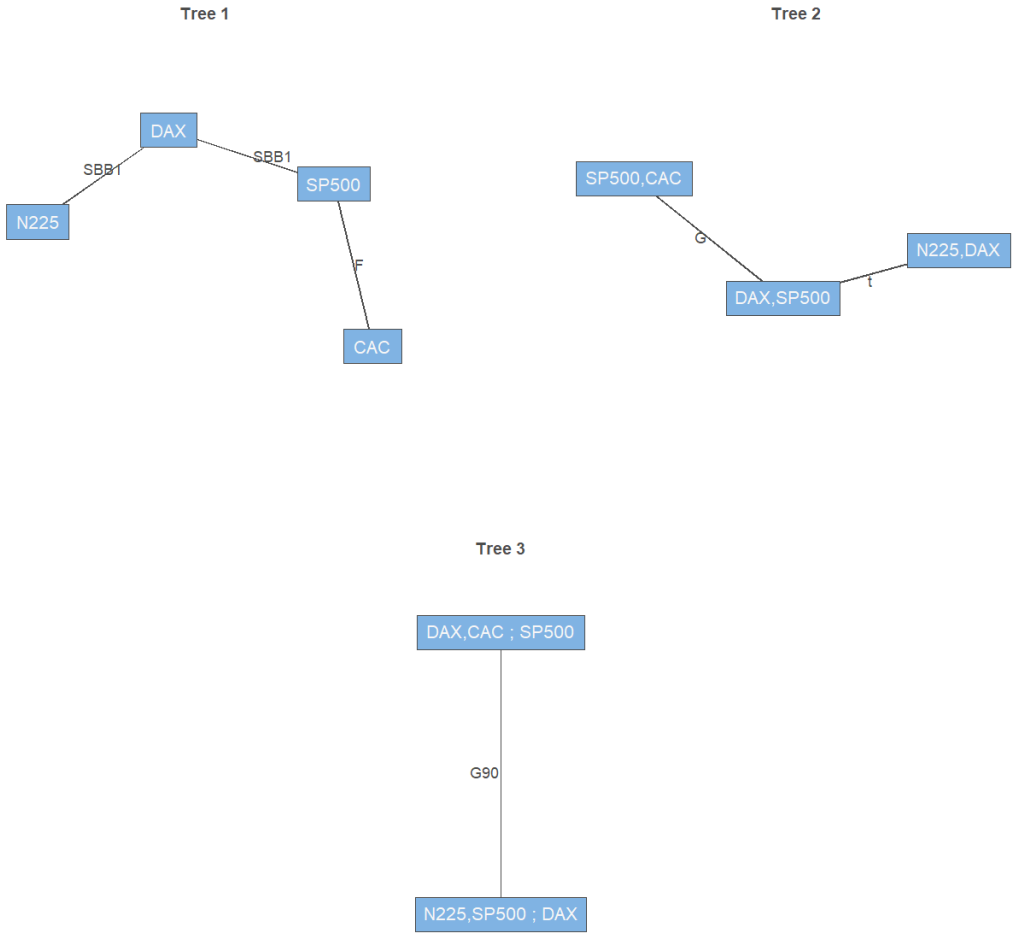


Figure D.13: *Vine structure of the fitted Vine models of the Residuals, transformed using their empirical CDF.*

Tree	Edge	Distribution	Parameters
T_1	CAC, SP500	Frank	$\rho = 2.63$, $\tau = 0.27$
	DAX, SP500	rotated BB1 (180)	par 0.62, par2 $= 1.50$, $\tau = 0.49$
	N225, DAX	rotated BB1 (180)	par $= 0.13$, par2 $= 1.12$, $\tau = 0.16$
T_2	DAX, CAC; SP500	Gumbel	par $= 1.09$, $\tau = 0.08$
	N225, SP500; DAX	t Student	$\rho = 0.07$, $\nu = 5.40$, $\tau = 0.05$
T_3	N225,CAC; DAX,SP500	rotated Gumbel (90)	par -1.05, $\tau = -0.04$

Table D.8: *Time Series Model Chosen for each variable.*

Proceeding to finding the weights of the optimal portfolios for each method, we obtain the optimal

frontiers. Markowitz' portfolio is characterized by weights:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (0, 22.4\%, 0, 77.6\%);$$

while for the vine copula approach, the optimal portfolio has weights equal to:

$$(w_{DAX}, w_{CAC}, w_{SP500}, w_{N225}) = (100\%, 0, 0, 0).$$

The results for the two method are rather different, as we can notice that the copula method suggest investing only in the German DAX index, while Markowitz's one prefers the Japanese index.

More in detail, we compare the two portfolios with an equal weights one in Table D.9, where it is possible to notice how for the first time the portfolio constructed with the vine copula approach performs better than the other two methods considered. In fact, the Sharpe ratio is equal to Markowitz's portfolio, while its STARR ratio is greater than the other two. This improvement could be associated with the fact that the amount of data available in this partition should be sufficient to guarantee the convergence of the estimators of the parameters of the vine copula considered.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(25\%, 25\%, 25\%, 25\%)$	0.064	22.899	0.1 %	0.3%
Mean-Var	$(0, 22.4\%, 0, 77.6\%)$	0.321	22.032	**0.3** %	1.5%
COPULA	$(100\%, 0, 0, 0)$	0.588	37.878	**0.3** %	**1.6**%

Table D.9: *Table comparing the portfolios obtained by optimizing the weights of two returns simulated from a copula-GARCH model, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the partition of $N = 615$ points.*

D.1.3.3 Out of Sample Evaluation

Lastly, it is left to evaluate the performance of the vine copula fit on an out of sample dataset. We apply the weights obtained to $N = 50$ points of the test set, and compare the cumulative returns obtained with the other two method in Figure D.14.

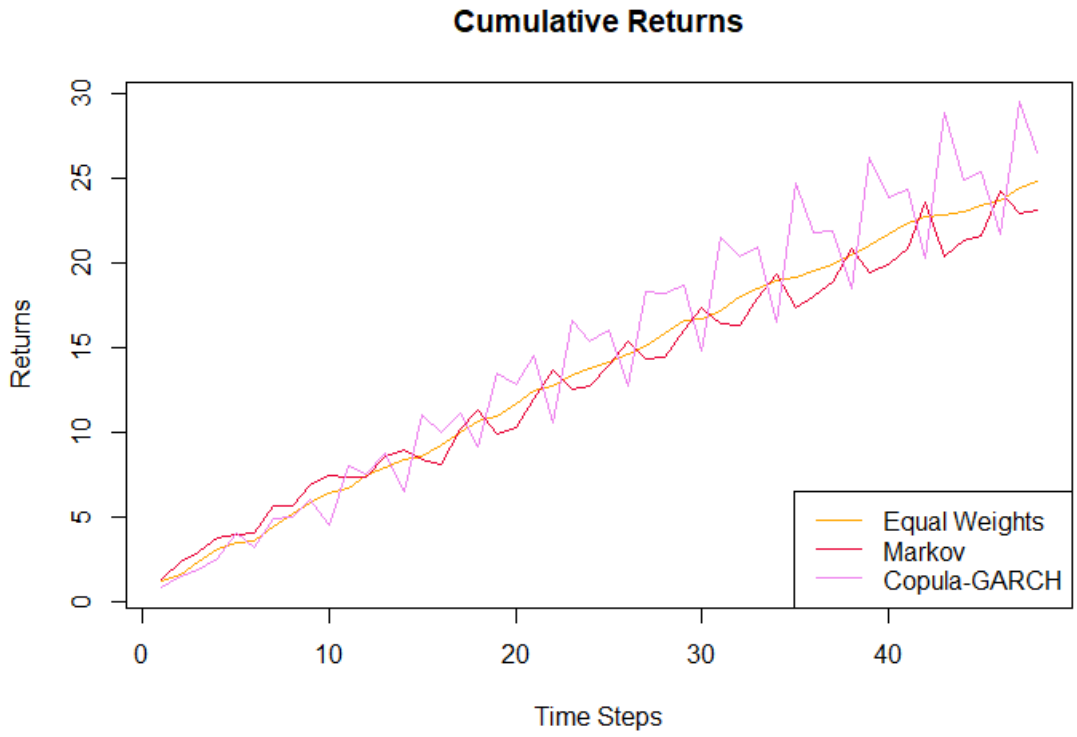


Figure D.14: *Evolution of the Cumulative Returns obtained by applying the optimal weights from different methods to $n = 50$ data points of the test set.*

The portfolio obtained from the vine copula approach seems to be the least stable among the three, however we can see how its peaks are significantly higher than the other two portfolios. The performance of the mean-variance portfolio is the worst one. More in detail, in Figure D.15 the difference between the cumulative returns is reported, and the advantage of the copula-GARCH portfolio is evident, as its positive peaks are significantly greater than the negative ones. This result is confirmed by Table D.10, where the overall cumulative return of the vine copula-GARCH portfolio appears to be better than the other portfolios considered.

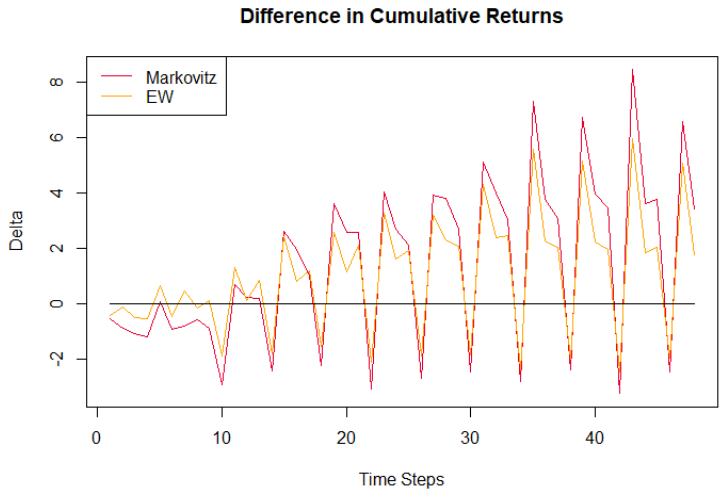


Figure D.15: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio, and the ones of the Markovitz' portfolio, or the best mean-CVaR portfolio.*

Method	Cumulative Return
EW	24.781
Mean-Var	23.134
COPULA	**26.551**

Table D.10: *Cumulative returns on test set, for each method at* $T_4 = 50$.

D.1.4 Partition 4

The fourth partition, is the one with lowest number of observations, as it has only $N = 170$ points, 144 of which compose the training set, and its 15%, hence 26 points, the test set. The test set data will be incrementally included in the model using an expanding time window approach.

D.1.4.1 Cumulative Returns

The overall performance of the vine copula-GARCH method on the test set is shown in Figure D.16, where the obtained cumulative returns are compared with those of the mean-variance optimal portfolios. The difference between the two returns is available in Figure D.17, where it is possible to notice how the vine copula approach proves to perform better in the first period of the test set, while in the last period it performs worse. This is a further confirmation that there are no apparent advantages in choosing the vine copula approach for optimizing the portfolios in short term periods, such as the daily refitting that is used in this example.

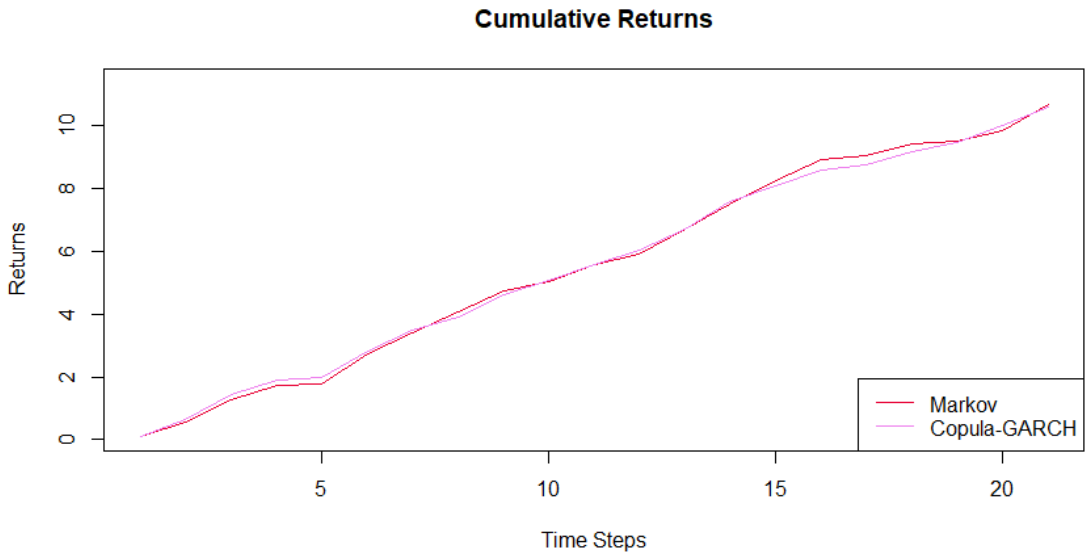


Figure D.16: *Evolution of the Cumulative Returns obtained by applying the optimal weights to the expanding test set.*

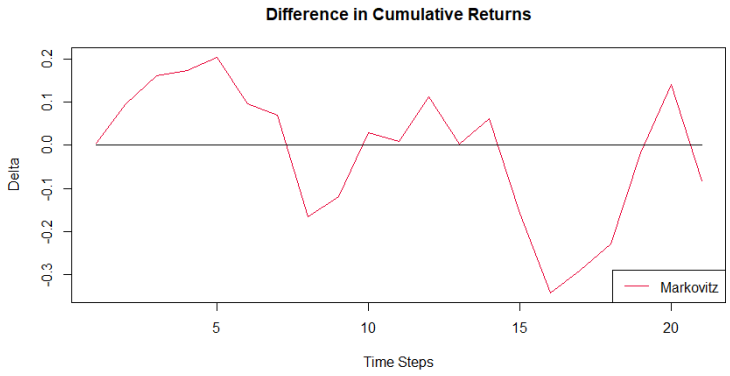


Figure D.17: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio and the ones of the Markovitz' portfolio on the tets set.*

D.1.4.2 Evaluate the Portfolio

We proceed by evaluating the portfolio obtained by fitting a vine copula to the entire fourth partition. We however have to keep in mind that, similarly to the first partition, the entire partition is composed of only $N = 170$ data points, which might not be sufficient to guarantee a fitted model that is truly able to understand and mimic the relationships of the original dataset, as ideally we should have around 500 observations.

The structure of fitted vine, which has a p-value of 0.385 for the goodness of fit test (which indicates a good fit), is provided in Figure D.18. Moreover the simplifying assumption is not rejected for any of the trees considered. The corresponding fitted parameters are discussed in Table D.11, where a rather low positive correlation can be deducted from the Kendall's tau of the copulas of the first tree, while the following trees are characterized by independence for the conditional copulas. In this case particularly,

the advantages of fitting a vine copula model to the residuals might be not effective, as it appears that there are no conditional relationship among the variables.

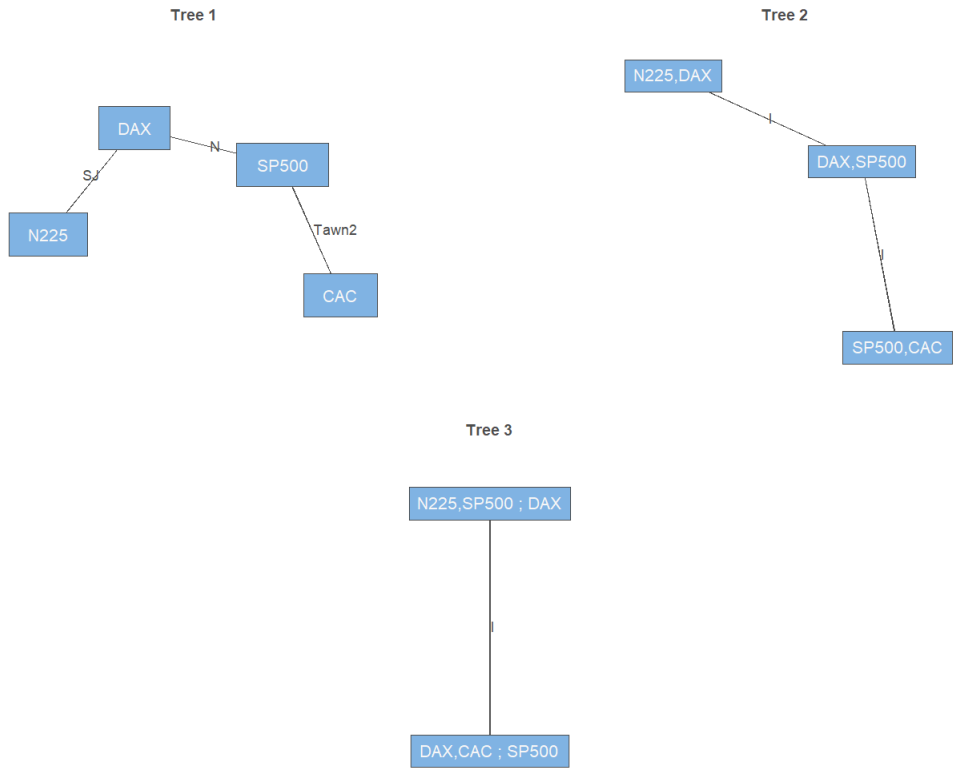


Figure D.18: *Vine structure of the fitted Vine models of the Residuals, transformed using their empirical CDF.*

Tree	Edge	Distribution	Parameters
T_1	SP500, CAC	Tawn2	par 1.66, par2 = 0.36, $\tau = 0.20$
	DAX, SP500	Gaussian	$\rho = 0.71$, $\tau = 0.50$
	N225, DAX	rotated Joe (180 deg)	par = 1.15, $\tau = 0.08$
T_2	DAX, CAC; SP500	Independent	$\tau = 0$
	N225, SP500; DAX	Independent	$\tau = 0$
T_3	N225, CAC; DAX,SP500	Independent	$\tau = 0$

Table D.11: *Time Series Model Chosen for each variable.*

Despite having no dependence in the conditional copulas of tree 2 and 3 of the vine structure, and having only few data points, we still proceed to evaluate the portfolio obtained by applying the vine-copula-GARCH approach to the fourth partition. The corresponding optimal portfolio is characterized by weights:

$$(w_1, w_2, w_3, w_4) = (63.1\%, 5.4\%, 0\%, 31.5\%).$$

The results are compared to those of the best mean-variance portfolio, which has weights equal to:

$$(w_1, w_2, w_3, w_4) = (11.7\%, 1.4\%, 36.0\%, 52.1\%).$$

Lastly, both portfolios are compared to an equal weight one in Table D.12, where we can clearly see that the mean-variance portfolio outperforms the other two, as it has the best Sharpe ratio, and the highest STARR ratio.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(25\%, 25\%, 25\%, 25\%)$	0.128	17.096	2.0 %	7 %
Mean-Var	$(11.7\%, 1.4\%, 36.0\%, 52.1\%)$	0.986	12.972	**2.1** %	**7.6** %
COPULA	$(63.1\%, 5.4\%, 0\%, 31.5\%)$	0.876	17.451	1.3 %	5 %

Table D.12: *Table comparing the portfolios obtained by optimizing the weights of two returns simulated from a copula-GARCH model, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the partition of $N = 170$ points.*

D.1.4.3 Out of Sample Evaluation

Once again, we conclude the analysis of the performance of the method studied in this thesis on an out of sample dataset, and compare the obtained cumulative results in Figure D.19.

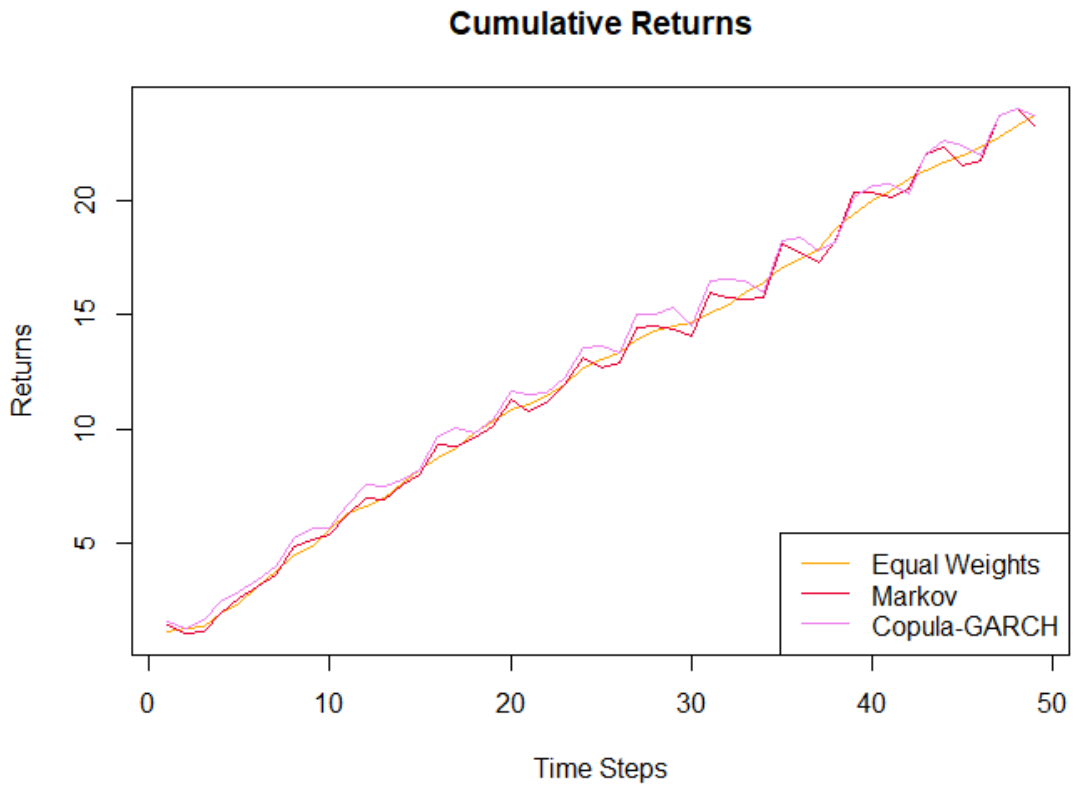


Figure D.19: *Evolution of the Cumulative Returns obtained by applying the optimal weights from different methods to $n = 50$ data points of the test set.*

In Figure D.20, the differences between the cumulative returns are illustrated, showing that the vine-copula-GARCH portfolio generally outperforms the other two portfolios for most of the dataset. This result is confirmed by Table D.13, where once again the cumulative returns of our portfolio appear grater than those of the other two portfolios considered. These results further confirm that using a vine copula to model the underlying structure provides a long-term advantage. By understanding the

relationships among the variables, the vine-copula-GARCH approach enables to achieve better results over extended periods, rather than short-term gains.

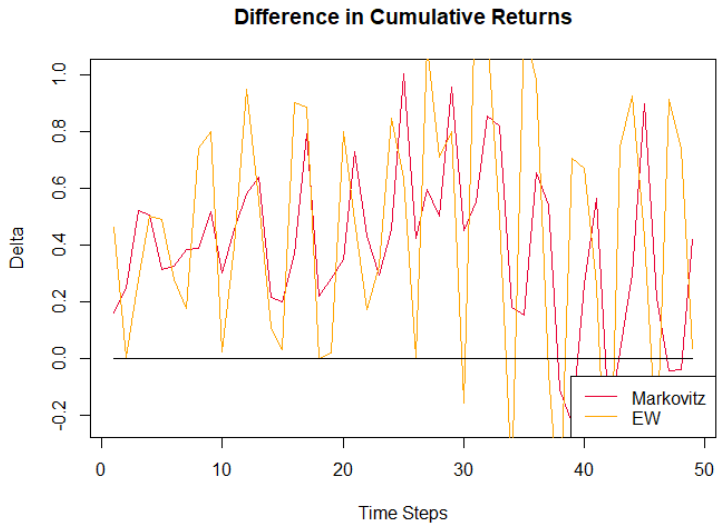


Figure D.20: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio, and the ones of the Markovitz' portfolio, or the best mean-CVaR portfolio.*

Method	Cumulative Return
EW	24.068
Mean-Var	23.703
COPULA	**24.300**

Table D.13: *Cumulative returns on test set, for each method at* $T_5 = 50$.

D.1.5 Partition 5

The last partition, is characterized by $N = 205$, as it covers the period between $30/06/2023 - 05/07/2024$. The dataset is divided in $N = 211$, the training set, and a test set of $N = 30$ points. As we progress, the test set data will be incrementally included in the model using an expanding time window approach.

D.1.5.1 Cumulative Returns

Firstly, we analyze the overall performance of the vine copula model on the test set. The results are compared the portfolio obtained by optimizing the mean-variance efficient frontier. and shown in Figure D.21, while in Figure D.22 the difference between the cumulative returns of the two portfolios are reported.

Surprisingly, despite the rather low number of points of the training set, the obtained portfolios seem to outperform the mean-variance one throughout the entire test set, as it is clear from the plot of the differences of Figure D.22. In addition, the difference between the two lines seems to progressively increase with the passing of time.

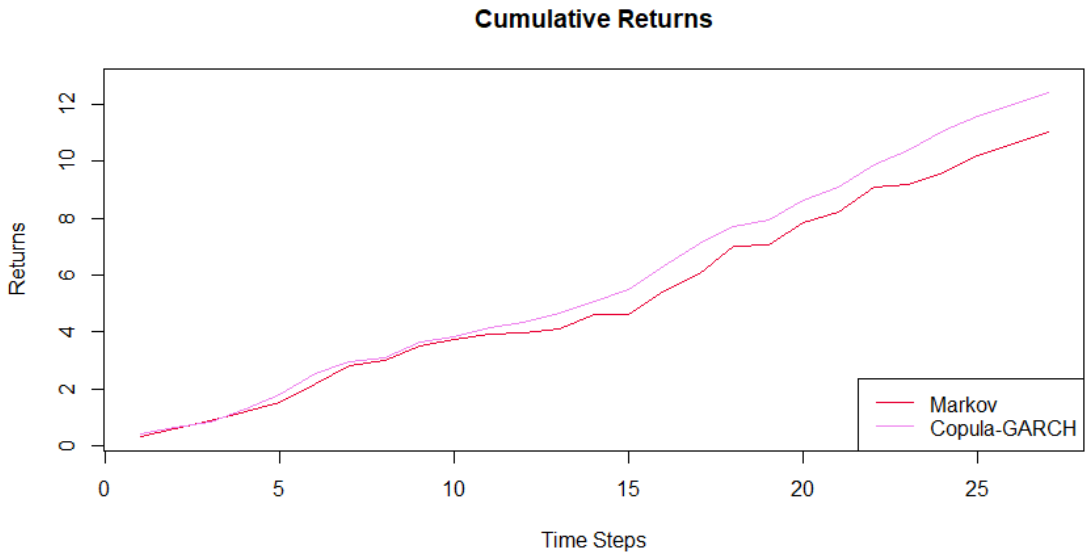


Figure D.21: *Evolution of the Cumulative Returns obtained by applying the optimal weights to the expanding test set.*

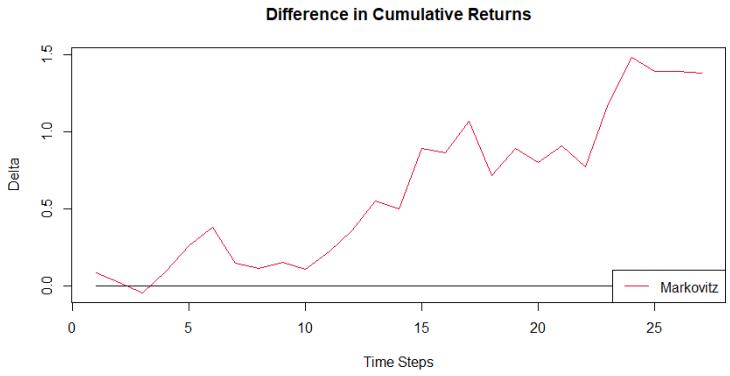


Figure D.22: *Difference between the Cumulative Returns obtained from the weights of the optimal copula-GARCH portfolio and the ones of the Markovitz' portfolio on the tets set.*

D.1.5.2 Evaluate the Portfolio

By fitting a vine model on the last partition, we obtain the structure shown in Figure D.23, and parameters provided in Table D.23. The goodness of fit test is again positive, with a p-value of 0.615, as well as the test for the simplifying assumption, which is never rejected.

Similarly to the previous partition, we can notice how the conditional copulas of the second and third trees appear to be independent, or with a Kendall's tau that is close to zero. While the bivariate copulas of the first tree are characterized by moderate positive correlations.

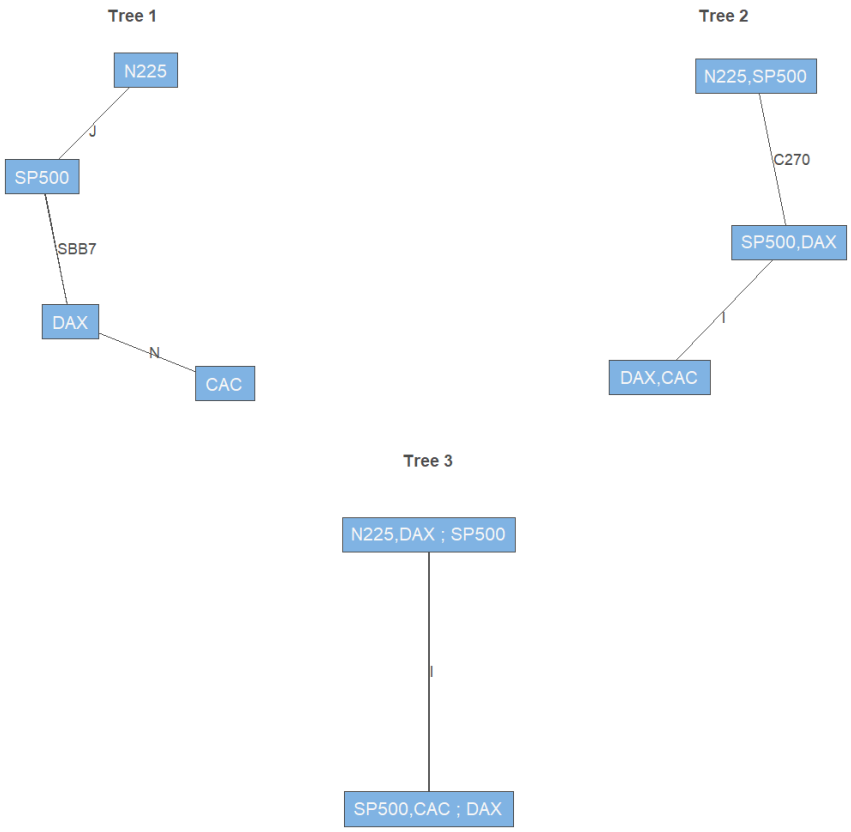


Figure D.23: *Vine structure of the fitted Vine models of the Residuals, transformed using their empirical CDF.*

Tree	Edge	Distribution	Parameters
T_1	DAX, CAC	Gaussian	$\rho = 0.36$, $\tau = 0.24$
	DAX, SP500	rotated BB7 (180 deg)	par 1.99, par2 = 1.85, $\tau = 0.57$
	N225, SP500	Joe	par 1.22, $\tau = 0.11$
T_2	CAC, SP500; DAX	Independent	$\tau = 0$
	N225,DAX; SP500	rotated Clayton (270 deg)	par -0.18, $\tau = -0.08$
T_3	N225,CAC; DAX,SP500	Independent	$\tau = 0$

Table D.14: *Time Series Model Chosen for each variable.*

The fitted vine copula is used to simulate $0Q = 1000$ points, which are then transformed and used to construct the efficient frontier. The corresponding optimal portfolio is characterized by the weights:

$$(w_1, w_2, w_3, w_4) = (46.5\%, 0.1\%, 22.9\%, 30.4\%).$$

The results are compared with the optimal mean-variance portfolio obtained by selecting the weights that guarantee the best Sharpe ratio, which corresponds to:

$$(w_1, w_2, w_3, w_4) = (11.6\%, 0.2\%, 36.1\%, 52.1\%).$$

More in detail, we compare the two portfolios with an equal weights one in Table D.15, where it is possible to notice that the portfolio constructed with the vine copula approach performs better than the

other two methods considered. In fact, both the Sharpe ratio and the STARR ratio appear to be greater than the other two, confirming the positive performance of our method for the partition considered.

Method	Weights	Mean	CVaR	Sharpe	STARR
EW	$(25\%, 25\%, 25\%, 25\%)$	0.665	15.123	0.9 %	4.4 %
Mean-Var	$(11.6\%, 0.2\%, 36.1\%, 52.1\%)$	0.954	14.044	1.8 %	6.8 %
COPULA	$(46.5\%, 0.1\%, 22.9\%, 30.4\%)$	0.977	12.744	**2.1** %	**7.7** %

Table D.15: *Table comparing the portfolios obtained by optimizing the weights of two returns simulated from a copula-GARCH model, using different methods. The table reports the weights and the results of the risk measures and ratios computed on the partition of $N = 205$ points.*

E

Probability Background

In this part of the Appendix, we will provide a brief overview of the main Probabilistic tools that will be necessary to introduce more advanced notations and definitions. In this section we will discuss the concepts as they are presented in the book [81], and in [82].

E.1 Event, Probability of an Event

When discussing Probability, it is necessary to start by introducing the concept of *sample space* Ω, which is the set of all possible outcomes of an experiment. The subsets of Ω are called *events* and denoted with a capital letter (e.g., the event $A \in \Omega$). An event has occurred if, after performing an experiment, the outcome is an element of A. Moreover, we can introduce the concept of *Probability* of an event A, $\mathbb{P}(A)$.

Definition 27 (Probability function)**.** *A Probability measure, defined on a σ-algebra $\mathcal{A}$ of the sample space Ω, is a function $\mathbb{P} : \mathcal{A} \to [0, 1]$ that has the following properties:*

1. $\mathbb{P}(\Omega) = 1$,
2. *Given two disjoint events A and B,* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

The probability expresses how much a certain event is likely to happen. We now want to introduce the concept of conditional probability, which requires the notion of intersection of two events A and B ($A \cap B$).

Definition 28 (Conditional Probability)**.** *The conditional probability of an event A given C is:*

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)},$$

where $\mathbb{P}(A \cap C)$ is the probability of the intersection of the two events, and $\mathbb{P}(C) > 0$ is the probability of the event C.

The conditional probability describes the probability of event A, given the fact that event C has occurred.

Finally, two events A and B are independent if the following property is verified:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

E.2 Random Variables

Definition 29 (Discrete Random Variable)**.** *Given a sample space Ω, a discrete random variable is a function $X : \Omega \to \mathbb{R}$ that can assume only a finite number of values $a_1, a_2, ..., a_n$ or a countable one $a_1, a_2, ...$*

Moreover, we can introduce the concept of *continuous random variables*. In this thesis, we will focus on the continuous case, so the discrete case will not be further discussed.

Definition 30 (Continuous random variable)**.** *A random variable X is continuous if for some function $f : \mathbb{R} \to \mathbb{R}$, and $\forall a, b \in \mathbb{R}$, with $a \leq b$:*

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx.$$

The function f is called the probability density function of X and has to satisfy the following properties:

1. $f(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Continuous random variables are also characterized by their *cumulative distribution function*.

Definition 31 (Cumulative distribution function)**.** *The Cumulative distribution function of a random variable X is the function $F : \mathbb{R} \to [0, 1]$ which is defined as:*

$$F(a) = \mathbb{P}(X \leq a) \ \forall a \in \mathbb{R}.$$

Let's notice that F is non-decreasing and holds the following relationship with the density function f: $F(x) = \int_{-\infty}^{x} f(u)du$.

E.2.1 Normal Distribution

A Normal distributed random variable $X_i \sim N(\mu_i, \sigma_i^2)$ is a fundamental concept in probability and statistics. The distribution is characterized by a mean, μ_i, and a standard deviation σ_i. The density function of X_i is given by:

$$f_i(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu_i)^2}{\sigma^2}\right\},$$

and its cumulative distribution function, or CDF, by:

$$F_i(x) = \mathbb{P}(z \leqslant x) = \Phi\left(\frac{x-\mu_i}{\sigma_i}\right),$$

where $\Phi(\cdot)$ denotes the CDF of a standard normal distribution. A standard normal variable is denoted as $N(0, 1)$, and is one of mean 0, and variance 1. Additionally, we introduce $\Phi^{-1}(\cdot)$ as the inverse function of the standard normal CDF.

E.3 Random Vectors

The concept of *random variable* can also be extended to the one of *random vectors*.

Definition 32 (Random Vector)**.** *A random vector* $\boldsymbol{X} = (X_1, X_2, ..., X_d)$ *is a function that maps outcomes of a random experiment to a vector in a vector space, hence* $X : \Omega \to \mathbb{R}^d$*. Each component of the vector is a random variable.*

E.3.1 Multivariate Distributions

By extending the concept of random variables to the one of random vectors, it is needed to extend the previously given definitions to the multivariate case. We can define the *joint distribution function*, which describes the relationship among the random variables and how they influence each other, and the *marginal distribution function* which describes the behavior of a single variable of the random vector.

Definition 33 (Joint distribution function)**.** *Given a random vector* $\boldsymbol{X} = (X_1, X_2, ..., X_d)$ *of d components, the joint distribution function F is a function* $F : \mathbb{R}^d \to [0, 1]$ *defined as:*

$$F(x_1, ..., x_d) = \mathbb{P}(X_1 \leq x_1, ..., X_d \leq x_d) \ \forall x_i \in \mathbb{R}, i = 1, .., d.$$

Definition 34 (Marginal distribution function)**.** *Let* $F(x_1, ..., x_d)$ *be the joint distribution function of a random vector* $\boldsymbol{X} = (X_1, X_2, ..., X_d)$*, then we can define the marginal distribution function of the component* X_i*,* $\forall x \in \mathbb{R}$*:*

$$F_i(x) = \mathbb{P}(X_i \leq x) = \lim_{x_1,...,x_{i-1},x_{i+1},...,x_d \to \infty} F(x_1, ..., x_{i-1}, x, x_{i+1}, ..., x_d).$$

In the continuous case, we can also provide the definitions of the *joint probability density function*, and of *marginal probability density function* of each variable, which is obtained by integrating over the remaining variables.

Definition 35 (Joint probability density function)**.** *Given a random vector* $\boldsymbol{X} = (X_1, X_2, ..., X_d)$*, the variables* X_i *have a joint continuous distribution if for some function* $f : \mathbb{R}^d \to \mathbb{R}$*, such that* $f(x_1, ..., x_d) \geq 0$ *and* $\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ..., x_d) dx_1 ... dx_d = 1$*, we have that :*

$$\mathbb{P}(a_1 \leq X_1 \leq b_1, ..., a_d \leq X_d \leq b_d) = \int_{a_1}^{b_1} ... \int_{a_d}^{b_d} f(x_1, ..., x_d) dx_1 ... dx_d$$

for all values $a_i \leq b_i$*.* $f(x_1, ..., x_d)$ *is the joint probability density function.*

From the previous definition, we can derive the marginal density of each component X_i of the random vector as follows:

$$f_i(x) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ..., x_d) dx_1 ... dx_{i-1} dx_{i+1} dx_d. \tag{E.1}$$

F

Statistical Background

The main definitions of statistical models and estimators are necessary to gain the background knowledge used to introduce the concept of copulas. These concepts, together with methods for comparing and evaluating different models will be discussed in the following section, based on how they are introduced in [81], and [82]. We will also introduce the concept of hypothesis testing, which is used throughout this thesis.

F.1 Statistical Models and Estimators

A *random sample* is a collection of mutually independent random variables $X_1, X_2, ..., X_d$ with the same probability distribution. We define a collection of events $(A_i)_i$ to be mutually independent if $\mathbb{P}(\cap_{i\in J}) = \prod_{i\in J} \mathbb{P}(A_i)$. The collection of all possible probability distributions of a given observation X composes a *statistical model*, which is defined as follows:

Definition 36 (Statistical Model)**.** *A statistical model $\mathcal{M}$ is a collection of probability distributions on a given sample space Ω.*

This means that a statistical model is constructed of repeated and independent measurements of the same quantity, hence the random sample. The collection of these elements is the dataset that composes the model. A random sample can have a distribution F, called the model distribution, and a density f. In other words, a dataset $x_1, ..., x_d$ is the realisation of the random sample $X_1, ..., X_d$.

We can extend Definition 36 to the one of *parametric models*, which are a particular case of statistical models.

Definition 37 (Parametric Model)**.** *A parametric model $\mathcal{M}$ is a family of probability distributions that has a finite number of parameters:*

$$\mathcal{M}_\theta = \{\mathbb{P}_\theta : \theta \in \Theta\},$$

where Θ is the set of parameters, and $\theta \in \Theta$ is one of the possible parameters of the model.

A statistical model is a collection of different probability distributions. But there is a unique distribution from which the sample originates, this distribution is called the *"true" distribution*. In the case

of parametric models, the latter distribution is characterized by its parameters that we define as the *"true" parameters* θ^*.

Finding a good distribution means estimating the distribution function F (and eventually the density f in the continuous case), which translates into estimating the parameters of the distribution. An *estimator* $\hat{\theta}$ is the random variable that corresponds to the estimates of the "true" parameters, and has the following definition:

Definition 38 (Estimator)**.** *Given the estimate* $t = h(x_1, ..., x_d)$*, a function that depends only on the dataset* $x_1, ..., x_d$*,* t *is the realisation of the random variable:*

$$\hat{\theta} = h(X_1, ..., X_d).$$

The random variable $\hat{\theta}$ *is the estimator.*

The estimator of a parameter is unbiased if its expected value is the parameter itself, otherwise, it is called biased. If we want to compare two estimators, not necessarily unbiased, we can use the *mean squared error*.

Definition 39 (Mean Squared Error)**.** *Given an estimator* $\hat{\theta}$ *of a parameter* θ*, the mean squared error of* $\hat{\theta}$ *is defined as the expected value of the squared difference of the estimator* $\hat{\theta}$ *and the parameter* θ*:*

$$MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2].$$

This definition holds when the expectation exists, hence when $\hat{\theta} \in L_2(\Omega, \mathcal{A}, \mathbb{P})$.

The MSE can also be viewed as the sum of the estimator's variance plus the squared bias. In the following subsections, we introduce the Maximum Likelihood, one of the most commonly used estimators, and two measures used to select the best estimator: the AIC and the BIC.

F.1.1 Likelihood

One of the most used estimators is the *maximum likelihood estimator*, or MLE. The idea behind the MLE is to find the parameters that are more likely to be the true parameters of the original model, this is called the Maximum Likelihood principle. In order to define the MLE we will first need to introduce the *Likelihood function*:

Definition 40 (Likelihood)**.** *Given a dataset* $x_1, ..., x_d$ *of realizations of the continuous random sample* $X_1, ..., X_d$*. Given density* f_θ *of* X*, and* θ *the parameter we want to estimate. The Likelihood function* $L(\theta)$ *can be defined as:*

$$L(\theta) = f_\theta(x_1) f_\theta(x_2) ... f_\theta(x_d).$$

We can also define the Log-Likelihood as the logarithm of L*:*

$$l(\theta) = \log L(\theta).$$

The *Maximum likelihood estimate* of the parameter θ, is the value $\hat{\theta}^{ML}$ that maximizes the Likelihood function $L(\theta)$. It is generally computed by calculating the partial derivatives of the Log-Likelihood

with respect to each parameter that we want to estimate, and by setting each partial derivative to zero.

The Likelihood proves also to be helpful to compare different models. In fact, when we have different distributions to describe a dataset, the Likelihood can be used to select which model is more likely to be the best one. Of course, this is not a quantification of the goodness of the model, but rather a way of comparing different ones. We will later discuss how to determine if a model is "good".

F.1.2 AIC and BIC

When using the Likelihood to compare and select distributions, we generally tend to prefer larger ones, since more parameters help to reach an optimal fit. But larger models are also computationally more expensive, harder to represent, and have a higher risk of overfitting. We, therefore, introduce two criteria that present a penalty for the dimension, hence they allow us to choose the best model by penalizing the most complex ones. These criteria are called the AIC and the BIC.

Definition 41 (AIC)**.** *Given a random sample X, and given a model $\mathcal{M}$, with set of parameters Θ, of length $|d|$. If $l(\theta)$ is the Log-likelihood, then the Akaike's Information Criterion is defined as:*

$$AIC := -2l(\hat{\theta}_d) + 2|d|. \tag{F.1}$$

Hence, it is minus two times the log-likelihood computed for the maximum likelihood estimator of dimension $|d|$, $\hat{\theta}_d$, to which we add two times its dimension.

The Bayesian information criterion, or BIC, on the other hand is defined as follows:

$$BIC := -2l(\hat{\theta}_d) + |d| \ln(n), \tag{F.2}$$

where n is the sample size.

For both AIC and BIC we choose the model that has the lower value. We can notice that the BIC tends to penalize complexity more.

F.2 Confidence Intervals

When we are estimating a parameter θ, instead of giving a single value, we can provide a range of acceptable values along with a level of confidence that the true value would lie in this interval: the *confidence intervals*.

Definition 42. *Given a dataset $x_1, ..., x_d$ of realisations of the random variables $X_1, ..., X_d$, and given a number $\gamma \in [0, 1]$. If we want to estimate the parameter θ, and if there exist sample statistics $L_d = g(X_1, ..., X_d)$ and $U_d = h(X_1, ..., X_d)$ for every value of θ s.t.*

$$\mathbb{P}(L_d < \theta < U_d) = \gamma \ \forall \theta$$

then (l_d, u_d) is the $100\gamma\%-$confidence interval for θ, where $l_d = g(x_1, ..., x_d)$ and $u_d = h(x_1, ..., x_d)$. The number γ is the confidence level.

F.3 Hypothesis Testing

Hypothesis testing is a statistical tool that is used when we have two alternative theories, or hypothesis indeed, and we want to choose between one of them. This is carried out by formulating a Null hypothesis and an alternative one, and by accepting the first or rejecting it in favour of the latter.

F.3.1 Null and Alternative hypothesis

As already discussed before, an hypothesis test has as its objective to accept or refuse a certain hypothesis, to which we refer to as the *Null hypothesis* H_0. We also formulate a complementary hypothesis, the *Alternative hypothesis* H_1. H_0 is assumed to be true until we can prove otherwise, and hence refuse it in favour of H_1. Generally, H_1 is what we really intend to prove.

In order to fully discuss hypothesis testing, we need to define what a *test statistic* is:

Definition 43 (Test statistics)**.** *Suppose the dataset to be the realisation of the random variables $X_1, ..., X_d$, then a test statistic is any sample statistic $T = h(X_1, ..., X_d)$ whose numerical value is used to decide whether we reject H_0.*

The realisation t of the test statistic is a function of the dataset, whose numerical value will determine whether we accept or refuse H_0.

F.3.2 P-value

One way to determine the outcome of an hypothesis testing, is to compute the *p-value*. The *p-value* is the left tail probability (or right tail depending on the direction of the values of the test statistic T in H_1), and it indicates how strong is the evidence against H_0. The p-value therefore indicates how likely T is "at least as extreme as the value t observed from the data", [81], meaning that for a small p-value we can reject H_0 in favour of the Alternative.

But how do we decide the value of the p-value, to decide when to reject the Null hypothesis? There is not a general rule, but generally, it is fixed around the value of 5%, which is called the *significance level* α. α, which has values in $[0, 1]$, is the largest probability of committing a type I error, hence the probability of rejecting H_0 when true.

Lastly, we can define the *Critical region* of a given hypothesis test with hypothesis H_0 and H_1, for a given significance level α: it is the set $K \subset \mathbb{R}$ of values of the test statistic T for which we reject H_0.

F.4 Goodness of Fit tests

A *Goodness of fit test* is a test that is aimed at evaluating whether the distribution of a sample test, and a theoretical distribution are coherent with each other. They are therefore used to discuss whether a chosen distribution is a good fit for a specific dataset. Here we discuss two tests: the Kolmogorv-Smirnov test, and the Vuong test.

F.4.1 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test is a statistical test used to investigate whether a model is a good fit for the original dataset. The idea behind it is to compute the maximum difference between the chosen

theoretical distribution F_0 that we want to test, and the empirical CDF, which is the step function $\hat{F_N}(x) = k/N$, with k the number of points with a value lower than x, and N the sample size. The test is therefore carried out by computing the distance:

$$d = \max_x |F_0(x) - \hat{F_N}(x)|.$$

For a chosen significance level α, and for a chosen hypothetical distribution F_0, we can compute the maximum vertical distance $d_{\alpha,N}$, for each specific sample size N. If the empirical CDF exceeds the given distance from F_0, then we reject the initial hypothesis that F_0 is the true distribution F, in favor of:

$$H_1 : F(x) \neq F_0(x) \text{ for some x},$$

which is the alternative hypothesis [83]. Moreover, the test can be two-sided or one-sided. The p-values can be extracted by finding the obtained maximum distance $d_{\alpha,N}$ in the table provided by [84], where further details on the method can be found. For the continuous case, the distribution of the test statistic converges to the Kolmogorov distribution D_n.

The Kolmogorov-Smirnov test is also available as a two-sample test, and is used as a Goodness of Fit test. It is a rank-based test that is aimed at discussing whether two samples of observations $x_1, \ldots, x_n$ with distribution F_x, and $y_1, \ldots, y_m$ with distribution F_y, $n \geq m$, have the same distribution, hence if $F_x = F_y$ [85]. We, therefore, use the test statistic:

$$d_n = \sup_u |F_{x,n}(u) - F_{y,m}(u)|,$$

which indicates the maximum distance between the two empirical distributions. We want this statistic to be minimal in order to be able to prove the null hypothesis:

$$H_0 : F_x(x) = F_y(x),$$

for all $x \in \mathbb{R}$. While the alternative H_1 corresponds to having the underlying distributions to be significantly different, hence to having $F_x \neq F_y$. Again, the test can also be performed in the one-sided version where the hypothesis becomes $H_0 : F_x \geq F_y$ and $H_1 : F_x < F_y$, or the opposite. A formula for the exact p-value, which corresponds to computing the probability $\mathbb{P}(d_n \leq x)$ does not exist, but the p-values can be extracted from the Table provided by [84].

F.4.2 Ljung-Box Test

Lastly, we introduce the Ljung-Box, an hypothesis test that aims to verify whether the residuals of a sample are independently distributed. The Null and Alternative hypothesis are the following:

$$H_0 : \text{The data are independently distributed.}$$

$$H_1 : \text{The data are not independently distributed; hence they exhibit serial correlation.}$$

The test statistic is given by:

$$Q = n(n+2) \sum_{h=1}^{H} \frac{\hat{ACF}^2(h)}{n-h},$$

where h represents the lag, hence the difference between two time points, H is the maximum lag considered and is normally, arbitrarily set to $H = 20$. Lastly $\hat{ACF}(h)$ is the sample autocorrelation of the residuals at a specific lag h. Q asymptotically follows a χ^2_h [49].

F.5 Other Tests

F.5.1 Q-Test

The multivariate Q-test is an hypothesis test that wants to assess weather the noise of a time series is a White Noise. Hence, if we have a multivariate time series with noise w_t, the two hypothesis are:

$$H_0 : w_t \text{ is white noise (WN).}$$

$$H_1 : w_t \text{ is not white noise (WN).}$$

Thus, not unexpectedly, the Q-test rejects the null hypothesis that the noise is white. The Q-test statistic is given by:

$$Q = n^2 \sum_{h=1}^{H} \frac{1}{n-h} Tr(\hat{\Gamma}_w(h)\hat{\Gamma}_w(0)^{-1}\hat{\Gamma}_w(h)\hat{\Gamma}_w(0)^{-1}),$$

where $\hat{\Gamma}_w(h) = n^{-1}\sum_{t=1}^{n-h} \hat{w}_{t+h}\hat{w}'_t$, h is the lag, $\hat{w}_t$ is the residual process. Let's note that Q has an asymptotic χ^2 distribution with $k^2(H-p)$ degrees of freedom [49].

F.6 Monte Carlo Simulations

Monte Carlo simulations, named after the famous casino in Monaco, refer to a set of computational methods introduced by E. Fermi, J. von Neumann, and S. Ulamn during the Manhattan Project. [79], [86] offer a broad overview of the methods, including potential extensions and applications. In this Section we will discuss the main idea behind the method and the principal characteristics.

Monte Carlo Methods are a series of methods based on random sampling, which are aimed at estimating a certain quantity. The intuition behind it relies in associating the probability and volumes, in particular we try to estimate the probability of an event as the volume of random observations that fall within a given set, compared to the universe of all possible outcomes. And its convergence to the correct value is guaranteed by the Law of Large Numbers. The method are various, but a general scheme can be outlined:

- model a system as a probability density functions f (see 30);
- repeatedly sample from the densities;
- compute the statistics of interest.

The applications of these simulation are numerous, ranging from quantum physics, to simulating stochastic processes for financial models [87]. While Monte Carlo methods are extremely useful to model complex systems that would be otherwise unobtainable analytically, they have some drawbacks, as they ca for example be computationally costly.

G

The Simplifying assumption

In Section 3.6 we introduced the concept behind the simplifying assumption for conditional copulas. Here we will discuss some hypothesis testing used to verify if a copula satisfies the simplifying assumption. The results are taken from [38], where additional tests can be found.

The main objective is to perform an hypothesis testing where the Null hypothesis is:

$$H_0 : C_{I|J}(\cdot|\mathbf{X}_J = \mathbf{x}_J) \text{ does not depend on the value of } \mathbf{x}_J,$$

while the alternative hypothesis H_1 is that this condition is not verified.

Derumigny and Fermanian discuss various ways of testing if the above hypothesis is verified, providing a wide overview of the possible methodologies that can be used to test the simplifying assumption.

Firstly, they discuss the *Brute-force* test, which computes the distance between the conditional copulas $C_{I|J}$, estimated with and without the simplifying assumption. It can be performed using either a Kolmogorov-Smirnov statistic, which implies the use of an L^∞ distance, hence it is computed using the sup over x_I and over x_J of the absolute value of the difference of the two estimated copulas; or with a Cramer von-Mises type statistics, which uses an L^2 distance between the copulas.

They also discuss some *dependence-based* type of tests. These tests are based on the observation that the Null hypothesis H_0 for the simplifying assumption is equivalent to testing if the random vector X_J is independent of the vector $Z_{I|J} = (F_1(X_1|X_J), \dots, F_p(X_p|X_J))$, the vector of the conditional marginal CDF (Proposition 4 of [38]). The latter assumption can be tested using bootstrap techniques on pseudo-samples, computing their empirical distribution, or by using tests that compute the distance between distributions, or again with tests based on a comparison between copulas.

Moreover, the two authors provide a *parametric* approach, which is carried out by testing the following hypothesis:

$$\hat{H}_0 : C_{I|J}(\cdot|\mathbf{X}_J) = C_\theta(\cdot), \text{ for some } \theta \in \Theta, \text{ for almost every } \mathbf{X}_J,$$

hence we to test if the parameters of the copula θ are constant and do not depend on the conditional variables $\mathbf{X}_J$. To perform the test, Maximum Likelihood and Kernel-Smoothing techniques are used to estimate the parameters of the conditional copula $C_{I|J}$. The estimated $\hat{\theta}$ are compared with the parameters $\hat{\theta}_0$, estimated from a simplified copula.

Finally, several Bootstrap methods for verifying the simplifying assumption were described, along with the *Boxes*-method. These two methods will be further discussed in the following subsections.

G.1 Bootstrap for the Simplifying Assumption

The idea behind the *Bootstrap method* is to approximate the limiting law of a test statistic τ of a given sample $S = (X_1, ..., X_d)$, by constructing new samples $S^* = (X_1^*, ..., X_d^*)$ from the initial one, and by extracting the test statistic τ^* of each new sample. The distribution of the new test statistics will be an approximation of the original one.

One way to extract the new sample statistics S^* is to draw each X_i^* independently with replacement from the original sample. This can be used to verify the simplifying assumption of a given copula. We can construct the sample $S^* = ((\hat{Z}^*_{1,I|J}, X^*_{1,J}), ..., (\hat{Z}^*_{n,I|J}, X^*_{n,J}))$, where each pair $(\hat{Z}^*_{i,I|J}, X^*_{i,J})$ is the pseudo-sample constructed by drawing $X^*_{i,J}$ from the components of the original random vector$X_{i,J}$, and $\hat{Z}^*_{i,I|J}$ independently from the observations $\hat{Z}_{i,I|J}$. With the new sample, we can then test the dependence-based hypothesis described in the previous subsection. This method is called the "pseudo-independent bootstrap". The paper also proposes a "conditional bootstrap", where the sample is constructed drawing independently from the estimated conditional law of X_I, a "parametric independent bootstrap" where we firstly estimate the parameters of the conditional copula, and then draw the samples independently from the estimated copula. Finally, they provide a "parametric conditional bootstrap" method where the samples $Z^*_{i,I|J,\hat{\theta}^*_i}$ are drawn independently from the copula corresponding to the previously drawn sample $X^*_{i,J}$.

G.2 Test the Simplifying Assumption with "boxes"

Paper [38] also proposes additional methods, constructed by partitioning the conditioning sub-set X_J into boxes, which allows for a simpler model, especially when dealing with dimensions higher than three. This translates into replacing a point-wise conditioning event to conditioning on the event $X_J \in A_J$, where A_J is a borelian subset. In this new setting, the copulas will be defined as:

$$\mathbb{P}(\mathbf{X}_I \leq \mathbf{x} | \mathbf{X}_J \in A_J) = C^{A_J}_{I|J}(\mathbb{P}(X_1 \leq x_1 | \mathbf{X}_J \in A_J), ..., \mathbb{P}(X_d \leq x_d | \mathbf{X}_J \in A_J) | \mathbf{X}_J \in A_J). \tag{G.1}$$

Let's also consider the following proposition:

Proposition 1. *Assume that the function $h : \mathbb{R}^d \to [0,1]$, defined by $h(y) := \mathbb{P}(X_I \leq y_I | X_J = y_J)$ is continuous everywhere. Let $x_J \in \mathbb{R}^{d-p}$ such that $F_{i|J}(\dot{|}x_J)$ is strictly increasing for every $i = 1, ..., d$. Then, for any sequence of boxes $(A_J^{(n)}(\boldsymbol{x}_J))$ such that $\cap_n(A_J^{(n)}(\boldsymbol{x}_j)) = \{\boldsymbol{x}_j\}$, we have that:*

$$\lim_n C^{A_J^{(n)}}_{I|J}(\boldsymbol{u}_I | \boldsymbol{X}_J \in A_J^{(n)}(\boldsymbol{x}_J)) = C_{I|J}(\boldsymbol{u}_I | \boldsymbol{X}_J = \boldsymbol{x}_J)$$

for every $u_I \in [0,1]^p$.

Starting from this proposition, we can reformulate a new Null hypothesis $\tilde{H}_0$, which will imply the original H_0. The new hypothesis will be:

$$\tilde{H}_0 : C_{I|J}^{A_J}(\mathbf{u}_I|\mathbf{X}_J \in A_J) \text{ does not depend on the value of } A_J \in \mathcal{A}_J, \text{for any } \mathbf{u}_I.$$

Unfortunately the opposite assumption does not hold, hence H_0 does not necessarily imply $\tilde{H}$. However, if we consider the family of disjoint subsets $\bar{\mathcal{A}}_J = \{A_{1,J}, ..., A_{m,J}\}$ we can obtain the following proposition:

Proposition 2. *Assume that, for all $A_J \in \bar{\mathcal{A}}_J$ and for all $i \in I$,*

$$F_{i|J}(x|\boldsymbol{X}_J = \boldsymbol{x}_J) = F_{i|J}(x|\boldsymbol{X}_J \in A_J), \forall \boldsymbol{x}_J \in A_J, x \in \mathbb{R}.$$

Then, H_0 implies the following null hypothesis:

$$\bar{H}_0 : A_J \to C_{I|J}^{A_J}(\cdot|\boldsymbol{X}_J \in A_J) \textit{is constant over } \bar{\mathcal{A}}_J.$$

This is a weaker assumption than the previously made one, but it has many interesting possible applications. First of all, it is important to notice that testing $\bar{H}_0$ allows also us to consider parameters that are continuous functions of $\mathbf{X}_J$. Moreover, if the latter set of variables is discrete, this Null hypothesis is equivalent to H_0. In general, testing for $\bar{H}_0$ allows to significantly simplify the model, by restricting the "the information set of the underlying conditional copula to a fixed number of conveniently chosen subsets A_J", [38].

Again testing the Simplifying Assumption using the previously described partition can be performed in different ways. The testing can be performed with a Non-parametric test, which implies estimating the conditional copulas from the empirical conditional marginal CDFs, and the empirical CDF of the conditional joint distributions: the hypothesis is then tested by computing the sum of the distance between the estimated copulas conditional to each partition considered. Alternatively, one can perform the test with a parametric approach, by estimating the parameters of the conditional copula, for each $A_J \in \bar{\mathcal{A}}_J$, using the maximum likelihood, and then by comparing them: under the Simplifying assumption the estimated $\hat{\theta}(A_J)$ should be equal. Lastly, a bootstrap method can be used.