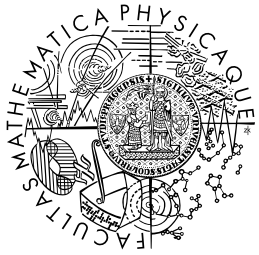


FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER'S THESIS

Bc. David Šír

Copula based models for multivariate time series

Department of Probability and Mathematical Statistics

Supervisor of the master's thesis: RNDr. Šárka Hudecová, Ph.D.

Study programme: Financial and Insurance Mathematics

Study branch: Financial and Insurance Mathematics

Prague 2023

I declare that I carried out this master's thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

I want to express my deepest gratitude to RNDr. Šárka Hudecová, Ph.D., for all the guidance, patience and support she provided me during the writing process of my thesis. Her expertise, insights and feedback were invaluable. I truly appreciate the time and effort she puts into mentoring me and helping me navigate the challenges of writing a thesis. I want to thank her for her unwavering support throughout this thesis, and I am grateful to have had her as my supervisor.

Last but not least, I would like to thank my family, who supported me during the study.

Title: Copula based models for multivariate time series

Author: Bc. David Šír

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Šárka Hudecová, Ph.D., Department of Probability and Mathematical Statistics

Abstract: The thesis deals with the modelling of multivariate time series. The SCOMDY model is described. It models individual univariate time series using an ARMA-GARCH, and their dependence structure is modelled using a copula. For copula selection goodness-of-fit test is discussed. Predictions are presented with algorithms for constructing prediction intervals. The whole theory is demonstrated with examples. Monte Carlo simulations verify the suitability and applicability of the theory. The SCOMDY model is applied to a three-dimensional time series consisting of the closing prices of stocks of Apple Inc. Microsoft Corporation and Alphabet Inc.

Keywords: time series, multivariate time series, copula, dependency, ARMA-GARCH, SCOMDY, prediction intervals

Contents

Introduction

In many fields, especially in financial mathematics and econometrics, but also in meteorology and medical research, multivariate time series are observed. In contrast to univariate time series modelling, which focuses on a single time-dependent variable, multivariate modelling considers multiple variables that may interact and have dependence between them. If we ignore the dependence structure in the modelling, we lose some information in the estimation itself, and we do not get better estimates. The modelling of time series with a dependence structure is the main objective of this thesis.

The knowledge gained can thus be applied, for example, in risk management, where the most common problem is to find the value at risk (VaR). In pricing various financial derivatives, specifically credit default swaps (CDS), collateralized debt obligation (CDO) and other options. Knowledge of the dependency structure is also helpful in portfolio construction. Based on understanding the dependency structure, we can construct an optimal portfolio. Or we can diversify the portfolio appropriately among sectors that are not too dependent. Forex, or currency pair exchanges, is another place where this theory can be applied.

In this work, we model the dependence structure using copulas. We show how to estimate such series, including estimation of the copulas, and then suggest how to make a prediction.

The paper is divided into five chapters. In the first chapter, we discuss the modelling of univariate time series. ARMA-GARCH model is introduced. Using quasi-maximum likelihood, we show how to estimate its parameters and then how to forecast the series.

The second chapter deals with copulas. The basic theory is summarised, and the most famous copulas are described, including the Clayton copula, the Gumbel copula, the Gaussian copula, and the Student's t-copula. Estimates for the unknown parameters of the copulas are derived. The theory of copulas is then concluded with a sample example.

The most important chapter of the thesis is chapter three. This chapter combines the previous two chapters and provides a detailed look at the multivariate copula-based time series model, which is the SCOMDY model. The model description and estimation process are discussed. Subsequently, the goodness-of-fit (GOF) test applicable to copulas is explained, and three different approaches to predicting multivariate time series with a dependence structure modelled by copulas are shown. The GOF test and the predictions are based on a parametric bootstrap. Finally, the whole theory is again demonstrated by a simple example.

The fourth chapter contains a Monte Carlo simulation study. The performance of the estimators, GOF test and prediction algorithms are investigated. We also address the question of the relevance of copulas to modelling a portfolio composed of modelled series.

The whole thesis concludes with a practical study where we consider a three-dimensional time series composed of stocks of Apple Inc. Microsoft Corporation and Alphabet Inc., which is the parent company of Google. The log returns of all three series are modelled using ARMA-GARCH models linked through copulas. The composition of the portfolio of these three titles is discussed.

The contribution of this thesis is to summarise the theory related to the SCOMDY model. It was necessary to mention the theory concerning ARMA-GARCH models and the copulas that are the building blocks of the SCOMDY model. The prediction confidence intervals for the SCOMDY model are not included in the literature. Therefore, three different algorithms constructing prediction intervals based on bootstrap techniques are proposed in the thesis. The first algorithm is based on simulations of the future evolution of the series. However, this algorithm does not consider the possible variability of the estimate. For this reason, two other algorithms have been proposed. The second algorithm is constructed only for the variability of the time series estimation. Finally, we present a third algorithm that considers the variability of both time series and copula estimation, which should be the correct approach. This third algorithm is the main contribution of the thesis. The subsequent simulation study based on the Monte Carlo method is also an intrinsic contribution. In the simulation section, we compare estimates of the unknown parameters obtained using the MLE and Kendall's tau-based methods. The functionality of the GOF test is investigated. Also, the performance of the three prediction algorithms is examined. The actual contribution is also a practical study in which the theory is applied. We also investigate the question of a portfolio composed of these assets.

The work includes four R scripts. The first script Copulas.R deals purely with the theory of copulas. The main part of the file presents three possible approaches to estimate the parameters of the copula. The second script Example_of_SCOMDY.R deals with an example of the SCOMDY model. The third script Simulations.R is for simulations. The second and third scripts implement prediction intervals for the AR model only. In the last, fourth script Practical_example.R, a practical problem is implemented. Here, the prediction algorithms are extended to the SCOMDY models with ARMA-GARCH models for univariate time series. This script also includes portfolio creation. This file can be applied to any multivariate time series with minimal code changes.

1. Time series

The first section of the thesis provides an overview of the essential concepts and principles related to time series analysis. The necessary definitions of time series, the ARMA-GARCH model, and the estimation procedure are introduced here. This section was inspired by Cipra [2020], Tsay [2005] and Brockwell and Davis [2002].

1.1 Introduction to time series

We begin with a general definition of time series as stochastic processes. This definition can be found, e.g., in Cipra [2020].

Definition 1. *A stochastic process is a family of random variables* $\{Y_t,\, t \in T\}$ *defined on the same probability space* $(\Omega,\, \mathcal{F},\, P)$ *indexed by* t *from the set* $T \subseteq \mathbb{R}$ *representing a time.*

A time series is defined as a stochastic process.

Remark. Our work restricts to a discrete-time, i.e. $T \subseteq \mathbb{Z}$. A time series is denoted as $\{Y_t,\, t \in \mathbb{Z}\}$.

Definition 2. *A time series* $\{Y_t,\, t \in \mathbb{Z}\}$ *is said to be stationary if it has a finite second moment and if for all* $s,\, t,\, h \in \mathbb{Z}$ *the following conditions hold*

$$\begin{aligned} E(Y_t) &= \mu, \quad \mu \in \mathbb{R}, \\ \operatorname{cov}(Y_s,\, Y_t) &= \operatorname{cov}(Y_{s+h},\, Y_{t+h}). \end{aligned}$$

A time series satisfying definition 2 is also called weakly stationary.

1.2 ARMA-GARCH

The book written by Box, Jenkins, and Reinsel [1970] summarises and algorithmises the so-far known conditional mean-based time series models and emphasises autocorrelation analysis. Therefore the method of these models is called the Box-Jenkins methodology. The simplest Box-Jenkins models for modelling time series are autoregressive (AR) models or moving-average (MA) models, or a combination of these two, which is the autoregressive-moving-average (ARMA) model. Financial time series often violate the assumption of constant conditional variance, which these models assume. Due to this fact, we need to use more sophisticated models. These models can be written in the general form

$$Y_t = \mu(\mathcal{F}_{t-1}) + \sigma(\mathcal{F}_{t-1})\, \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1.1}$$

where $\mathcal{F}_{t-1}$ is the full information about our time series until the time $t-1$. Formally, it is the smallest σ-algebra generated by all past values $Y_{t-1},\, Y_{t-2},\, \ldots$. Mathematically we write

$$\mathcal{F}_{t-1} = \sigma\{Y_{t-1},\, Y_{t-2},\, \ldots\},$$

see page 11 in Francq and Zakoian [2019].

Furthermore, $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ is a process of independent and identically distributed (abbreviated as *i.i.d.*) random variables with zero mean and unit variance. For each $t \in \mathbb{Z}$ the random variable ε_t is being independent of $\mathcal{F}_{t-1}$. Random variables ε_t are referred to as innovations.

Remark. The random variables $e_t = \sigma(\mathcal{F}_{t-1})\,\varepsilon_t$, $t \in \mathbb{Z}$, known as prediction errors, are uncorrelated, and they are not generally independent, see page 205 in Cipra [2020].

The equation (1.1) for Y_t consists of two parts. The first part represents the conditional mean

$$\mu(\mathcal{F}_{t-1}) = \mathsf{E}\,(Y_t \mid \mathcal{F}_{t-1}), \quad t \in \mathbb{Z},$$

and the second part denotes the conditional variance, better known as volatility in finance

$$\sigma^2(\mathcal{F}_{t-1}) = \operatorname{var}(Y_t \mid \mathcal{F}_{t-1}), \quad t \in \mathbb{Z},$$

it is assumed $\sigma(\mathcal{F}_{t-1}) > 0$, $t \in \mathbb{Z}$.

Bollerslev [1986] introduced the generalised autoregressive conditional heteroskedasticity (GARCH) model to model conditional variance. Bollerslev developed the GARCH model as an extension of the earlier autoregressive conditional heteroskedasticity (ARCH) model introduced by Engle [1982]. The GARCH model is widely used in finance and econometrics to model time-varying volatility.

In what follows, the ARMA-GARCH model is described. In ARMA-GARCH, the conditional mean is modelled using an ARMA model, and the conditional volatility using a GARCH model. ARMA(p,q)-GARCH(m,s) model with orders of p, q of ARMA part and m, s of GARCH part is described by the following three equations. For $t \in \mathbb{Z}$ holds

$$\begin{aligned} Y_t &= \mu + \sum_{i=1}^{p} \varphi_i Y_{t-i} + e_t + \sum_{j=1}^{q} \theta_j e_{t-j}, \\ e_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i e_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2, \end{aligned} \tag{1.2}$$

where ε_t are *i.i.d.* with zero mean and unit variance and μ, φ_1, …, φ_p, θ_1, …, θ_q, α_0, α_1, …, α_m, β_1, …, β_s are the parameters described in the following.

The first equation of the ARMA-GARCH model, defined in (1.2), represents the ARMA(p,q) part. It is used for modelling the conditional mean, and the third equation of (1.2) stands for the GARCH(m,s) part, which is used for conditional variance modelling.

To make the ARMA-GARCH stationary, we impose the following constraints on the parameters of (1.2).

All roots z_1, …, z_p of the polynomial

$$\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p$$

lie outside of the unit circle in the complex plane. Formally, it means that all the roots fulfil $|z_1|$, …, $|z_p| > 1$.

On the parameters $\alpha_0, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_s$ of GARCH part we put the following restrictions

$$\begin{aligned} &\alpha_0 > 0,\, \alpha_i \geq 0,\, \beta_j \geq 0, \quad i = 1, \ldots, m,\, j = 1, \ldots, s, \\ &\sum_{i=1}^{max(m,\,s)} (\alpha_i + \beta_i) < 1, \quad \alpha_i = 0,\, i > m,\, \beta_j = 0,\, j > s. \end{aligned} \tag{1.3}$$

This condition is sufficient for the existence of variance and for the series to be stationary.

1.3 Estimation of ARMA-GARCH

Estimating the parameters of the ARMA-GARCH model is an essential part of the modelling procedure. This part is inspired by Francq and Zakoian [2019]. The unknown parameters can be estimated using the maximum likelihood (ML) method if we assume that ε_t follows a specific distribution for every $t \in \mathbb{Z}$. It is often assumed that ε_t has a standard normal distribution. However, financial time series commonly have heavier tails, which typically violates this assumption. For this reason, we perform a Gaussian quasi-maximum likelihood estimation (Gaussian QMLE) based on a quasi-likelihood function. The quasi-likelihood function is constructed as a likelihood function under the possibly invalid assumption that ε_t has a standard normal distribution for every $t \in \mathbb{Z}$.

To derive the estimation, let $\{Y_1, \ldots, Y_n\}$, $n \in \mathbb{N}$ be a time series generated from a stationary ARMA(p, q)-GARCH(m, s) model with known orders p, q, m and s.

The unknown parameters of ARMA(p, q)-GARCH(m, s) are denoted in the following way

$$\begin{aligned} \boldsymbol{\gamma} &= (\mu, \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q)^\top, \\ \boldsymbol{\delta} &= (\alpha_0, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_s)^\top, \\ \boldsymbol{\psi} &= (\boldsymbol{\gamma}^\top, \boldsymbol{\delta}^\top)^\top. \end{aligned}$$

Also, we denote $\boldsymbol{\Psi}_\gamma \subseteq \mathbb{R}^{p+q+1}$, $\boldsymbol{\Psi}_\delta \subseteq (0, \infty) \times [0, \infty)^{m+s}$ and $\boldsymbol{\Psi} = \boldsymbol{\Psi}_\gamma \times \boldsymbol{\Psi}_\delta \subseteq \mathbb{R}^{p+q+1} \times (0, \infty) \times [0, \infty)^{m+s}$ subspaces for $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ and $\boldsymbol{\psi}$. It is assumed that these subspaces are chosen so that the ARMA-GARCH is stationary for every parameter combination from this subspace. The true value of the parameter is denoted as $\boldsymbol{\psi}_0 = (\boldsymbol{\gamma}_0^\top, \boldsymbol{\delta}_0^\top)^\top \in \boldsymbol{\Psi}$.

The parametric model we are going to solve is a reformulation of equations (1.2)

$$\begin{aligned} e_t(\boldsymbol{\gamma}) &= Y_t - \mu - \sum_{i=1}^{p} \varphi_i Y_{t-i} - \sum_{j=1}^{q} \theta_j e_{t-j}(\boldsymbol{\gamma}), \\ \varepsilon_t(\boldsymbol{\psi}) &= e_t(\boldsymbol{\gamma}) / \sigma_t(\boldsymbol{\psi}), \\ \sigma_t^2(\boldsymbol{\psi}) &= \alpha_0 + \sum_{i=1}^{m} \alpha_i e_{t-i}^2(\boldsymbol{\gamma}) + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2(\boldsymbol{\psi}), \end{aligned} \tag{1.4}$$

where $\varepsilon_t(\boldsymbol{\psi}_0) = \varepsilon_t$, $e_t(\boldsymbol{\gamma}_0) = e_t$ and $\sigma_t^2(\boldsymbol{\psi}_0) = \sigma_t^2$. Note that $e_t(\boldsymbol{\gamma})$ and $\sigma_t^2(\boldsymbol{\psi})$ are computed recursively. The initial values can be set to zero, for example. However, they can be set to a different constant, see further.

To construct the Gaussian quasi-likelihood function, ε_t is assumed to have a standard normal distribution. From this assumption and from the ARMA-GARCH representation in (1.4), we can derive the distribution of $Y_t \mid \mathcal{F}_{t-1}$, $t = 1, \ldots, n$ which is also normal. From assumption $\varepsilon_t \sim \mathcal{N}(0,\, 1)$, $t = 1, \ldots, n$ it holds

$$e_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\Big(0,\, \sigma_t^2\Big),$$

which implies

$$Y_t - \mu - \sum_{i=1}^{p} \varphi_i Y_{t-i} - \sum_{j=1}^{q} \theta_j e_{t-j} \;\Big|\; \mathcal{F}_{t-1} \sim \mathcal{N}\Big(0,\, \sigma_t^2\Big),$$

from which the distribution of $Y_t \mid \mathcal{F}_{t-1}$ is obtained

$$Y_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\Big(\mu + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j},\, \sigma_t^2\Big).$$

Thus, the Gaussian quasi-likelihood function has the following form

$$\mathcal{L}_n(\boldsymbol{\psi}) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2(\boldsymbol{\psi})}} \exp\Big(-\frac{e_t^2(\boldsymbol{\gamma})}{2\sigma_t^2(\boldsymbol{\psi})}\Big).$$

The estimator of $\boldsymbol{\psi}$ maximises $\mathcal{L}_n(\boldsymbol{\psi})$. Working with the logarithmic version of the Gaussian quasi-likelihood function is more convenient. So for maximisation, we use the logarithmic Gaussian quasi-likelihood function

$$\ell_n(\boldsymbol{\psi}) = \log\Big(\mathcal{L}_n(\boldsymbol{\psi})\Big) = \sum_{t=1}^{n} \bigg(-\frac{1}{2}\log\Big(2\pi\sigma_t^2(\boldsymbol{\psi})\Big) - \frac{e_t^2(\boldsymbol{\gamma})}{2\sigma_t^2(\boldsymbol{\psi})}\bigg). \tag{1.5}$$

To calculate the logarithmic Gaussian quasi-likelihood function, initial values must be set, see Francq and Zakoian [2019, part 7.2]. Denote $\widetilde{\ell}_n(\boldsymbol{\psi})$ the logarithmic Gaussian quasi-likelihood function conditional on initial values

$$\widetilde{\ell}_n(\boldsymbol{\psi}) = \sum_{t=1}^{n} \bigg(-\frac{1}{2}\log\Big(2\pi\widetilde{\sigma}_t^2(\boldsymbol{\psi})\Big) - \frac{\widetilde{e}_t^2(\boldsymbol{\gamma})}{2\widetilde{\sigma}_t^2(\boldsymbol{\psi})}\bigg),$$

where for $t = 1, \ldots, n$ it holds

$$\begin{aligned}
\widetilde{e}_t(\boldsymbol{\gamma}) &= Y_t - \mu - \sum_{i=1}^{p} \varphi_i Y_{t-i} - \sum_{j=1}^{q} \theta_j \widetilde{e}_{t-j}(\boldsymbol{\gamma}), \\
\widetilde{\sigma}_t^2(\boldsymbol{\psi}) &= \alpha_0 + \sum_{i=1}^{m} \alpha_i \widetilde{e}_{t-i}^2(\boldsymbol{\gamma}) + \sum_{j=1}^{s} \beta_j \widetilde{\sigma}_{t-j}^2(\boldsymbol{\psi}).
\end{aligned}$$

To compute $\widetilde{\ell}_n(\boldsymbol{\psi})$, several initial values need to be specified. For the selection of initial values, see Francq and Zakoian [2019, part 7.2].

The QMLE of $\boldsymbol{\psi}$ is then

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\arg\max}\, \widetilde{\ell}_n(\boldsymbol{\psi}). \tag{1.6}$$

For completeness, we denote the individual estimates of all parameters by

$$\begin{aligned}\widehat{\boldsymbol{\psi}} &= (\widehat{\boldsymbol{\gamma}}^\top, \widehat{\boldsymbol{\delta}}^\top)^\top,\\ \widehat{\boldsymbol{\gamma}} &= (\widehat{\mu}, \widehat{\varphi}_1, \ldots, \widehat{\varphi}_p, \widehat{\theta}_1, \ldots, \widehat{\theta}_q)^\top,\\ \widehat{\boldsymbol{\delta}} &= (\widehat{\alpha}_0, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_m, \widehat{\beta}_1, \ldots, \widehat{\beta}_s)^\top.\end{aligned}$$

Under quite general regularity conditions, see Francq and Zakoian [2019, part 7.2], the QMLE is consistent, and if $\mathsf{E}\,(\varepsilon_t^4)$ is finite, then the estimator is asymptotically normal.

Remark. In the previous section, we discussed an estimator that simultaneously estimates ARMA and GARCH parts. However, estimating the ARMA component first and then estimating the GARCH for the residuals from the first step is more common in practice, see Tsay [2005].

If ε_t is normally distributed, it is possible to use the maximum likelihood estimator directly based on the normal distribution. However, as mentioned earlier, this is not the case for financial time series, as they have heavy tails. But it is already possible to assume that ε_t can have a Student's t-distribution. In that case, it is already worth using maximum likelihood with the assumption of a Student's t-distribution. Another possible distribution for ε_t is the generalised error distribution (GED). See Cipra [2020] for details.

1.4 Selection of model order

In the previous section, we estimated a model where the model order was assumed, i.e., we estimated the ARMA(p, q)-GARCH(m, s) model where we assumed that the model orders p, q, m, s are known. This section suggests how to determine such an order.

The most common method to find orders is to estimate several ARMA-GARCH models with different orders and then compute the Akaike information criterion (AIC) or Bayesian information criterion (BIC) for the decision, see Tsay [2005]. Finding the most suitable ARMA model orders and adding the GARCH part as needed is also possible. The model with the lowest value of the criterion is the most appropriate. Both criteria can be found in Burnham and Anderson [2004] and are defined as follows

Definition 3. *Let us assume a model M with unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^K$. Let $\widehat{\boldsymbol{\theta}}$ is a maximum likelihood estimate of $\boldsymbol{\theta}$ and $\mathcal{L}_M(\widehat{\boldsymbol{\theta}} \mid \boldsymbol{X})$ is the maximum likelihood of model M computed from the data $\boldsymbol{X}$, where $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$, $n \in \mathbb{N}$. The* AIC *of model M estimated from data $\boldsymbol{X}$ is defined as*

$$\mathrm{AIC} = -2\log\Big(\mathcal{L}_M(\widehat{\boldsymbol{\theta}} \mid \boldsymbol{X})\Big) + 2K, \tag{1.7}$$

and the BIC *of model M estimated from data $\boldsymbol{X}$ is defined as*

$$\mathrm{BIC} = -2\log\Big(\mathcal{L}_M(\widehat{\boldsymbol{\theta}} \mid \boldsymbol{X})\Big) + K\log(n). \tag{1.8}$$

The definition is given in a general form, as we will use it in the next part of the thesis. For time series, the AIC and BIC are sometimes defined as follows

$$\widetilde{\mathrm{AIC}} = \frac{-2\log\Big(\mathcal{L}_M(\widehat{\boldsymbol{\theta}} \mid \boldsymbol{X})\Big) + 2K}{n},$$
$$\widetilde{\mathrm{BIC}} = \frac{-2\log\Big(\mathcal{L}_M(\widehat{\boldsymbol{\theta}} \mid \boldsymbol{X})\Big) + K\log(n)}{n},$$

where we make the same notation as in Definition 3 and n denotes the sample size. The log-likelihood in our notation for time series from Definition 3 corresponds to

$$\log\Big(\mathcal{L}_M(\widehat{\boldsymbol{\theta}} \mid \boldsymbol{X})\Big) = \widetilde{\ell}_n(\widehat{\boldsymbol{\psi}}).$$

Finding the best model based on $\widetilde{\mathrm{AIC}}$ or $\widetilde{\mathrm{BIC}}$ is same as on AIC or BIC.

1.5 Prediction

The model is already estimated, so it is possible to look at predictions. The analytical point prediction is constructed in the book of Cipra [2020]. Let us denote $\widehat{Y}_{t+K}(t)$ the prediction of Y_{t+K}, $K \in \mathbb{N}$, constructed in time t, known as K-step ahead prediction.

To introduce analytical predictions, we first define a general prescription for the time series along with the unknown parameters denoted as $\boldsymbol{\psi}$

$$Y_t = \mu(\mathcal{F}_{t-1}, \boldsymbol{\psi}) + \sigma(\mathcal{F}_{t-1}\, \boldsymbol{\psi})\, \varepsilon_t, \quad t \in \mathbb{Z}. \tag{1.9}$$

The K-step ahead point prediction $\widehat{Y}_{t+K}(t)$ of Y_{t+K} for the ARMA-GARCH model is iteratively derived from the first equation of (1.2)

$$\widehat{Y}_{t+K}(t) = \widehat{\mu} + \sum_{i=1}^{p} \widehat{\varphi}_i \widehat{Y}_{t+K-i}(t) + \widehat{e}_t + \sum_{j=1}^{q} \widehat{\theta}_j \widehat{e}_{t+K-j},$$

where it is assumed

$$\widehat{Y}_{t+j}(t) = Y_{t+j}, \quad j \le 0,$$
$$\widehat{e}_{t+j}(t) = \begin{cases} 0, & j > 0, \\ Y_{t+j} - \widehat{Y}_{t+j}(t+j-1) = e_{t+j}, & j \le 0. \end{cases}$$

For volatility prediction, see Cipra [2020, Section 8.3.5].

Prediction intervals are typically derived under normality assumption and will not be discussed in this thesis. It is possible to find it in the book written by Cipra [2020]. Section 3.4 discussed predictive intervals based on bootstraps for multivariate time series without assuming normality. It is possible to use these algorithms in a one-dimensional time series, see Pascual, Romo, and Ruiz [2004].

1.6 Multivariate time series

Univariate time series allows us to model the evolution of individual assets. In practice, however, we need to model the whole portfolio's evolution, consisting of several individual assets that may depend on each other. For this reason, we introduce multivariate time series that allows this modelling and deals with the possible interdependence of individual univariate time series.

Simply saying, a multivariate time series is a union of two or more univariate time series, defined in this work's first part 1.1. The formal definition of a multivariate stochastic process and multivariate time series follows.

Definition 4. *A multivariate stochastic process is a family of D-dimensional random vectors $\{\boldsymbol{Y}_t,\, t \in T\}$, where $\boldsymbol{Y}_t = (Y_{1,t},\, \ldots,\, Y_{D,t})^\top$, $t \in T$, defined on the same probability space $(\Omega,\, \mathcal{F},\, P)$ indexed by t from the set $T \subseteq \mathbb{R}$ representing a time.*

A multivariate time series is defined as a multivariate stochastic process.

Remark. We again restrict ourselves to a discrete-time on $T \subseteq \mathbb{Z}$.

A D-dimensional multivariate time series is denoted as $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$.

The following definition generalises weak stationarity for a multivariate time series, see Cipra [2020].

Definition 5. *A D-dimensional time series $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$ is said to be stationary if it has finite second moments and if for all $s,\, t,\, h \in \mathbb{Z}$ the following conditions hold*

$$\begin{aligned} E(\boldsymbol{Y}_t) &= \boldsymbol{\mu}, \quad \boldsymbol{\mu} = (\mu_1,\, \ldots,\, \mu_D)^\top \in \mathbb{R}^D, \\ \operatorname{cov}(\boldsymbol{Y}_s,\, \boldsymbol{Y}_t) &= E(\boldsymbol{Y}_s - \boldsymbol{\mu})(\boldsymbol{Y}_t - \boldsymbol{\mu})^\top = \operatorname{cov}(\boldsymbol{Y}_{s+h},\, \boldsymbol{Y}_{t+h}). \end{aligned}$$

In other words, a multivariate time series is stationary if its mean vector and covariance matrix are invariant in time. The definition is similar to the one-dimensional case, also known as weak stationarity.

In Figure 1.1, we can see a multivariate time series created by connecting three univariate time series of closing prices of Apple Inc. (ticker AAPL), Alphabet Inc. (GOOG) and Microsoft Corporation (MSFT). The start date is January 2007, and the end is December 2022. All the data has been collected from finance.yahoo.com.

All three series are non-stationary, which is a problem for modelling them. However, we could model, for example, logarithmic returns (log returns), which might already be stationary (we will deal with this later). Subsequently, we could consider univariate ARMA-GARCH models for these individual series. Unfortunately, this does not take into account their possible dependency. A model that could take this into account is presented in the following sections of the work.

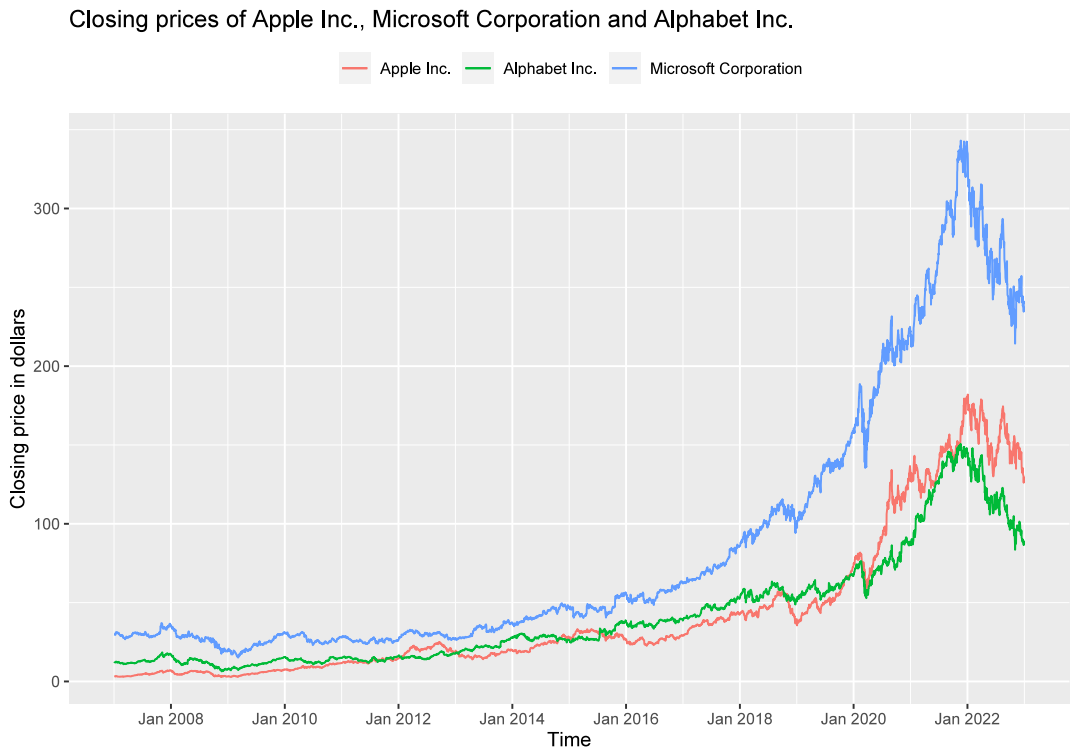


Figure 1.1: Example of multivariate time series. In the figure closing prices of Apple Inc., Microsoft Corporation and Alphabet Inc. Prices are in dollars.

2. Copulas

The second chapter briefly introduces the theory of copulas. We define copulas, describe their basic properties, introduce Sklar's theorem, and finally derive parametric and non-parametric estimations of copulas.

Simply saying, copulas are functions which connect marginal distributions to make a joint distribution. Copulas are very popular today for modelling dependent random variables. Instead of modelling the entire joint distribution, we can model the marginals and their dependence structure separately and combine these components using a copula.

This chapter is based on Trivedi, Zimmer, et al. [2007].

2.1 Introduction to copulas

The introduction to copula theory begins with its definition. The basic properties of the copulas are summarised in this section, including the famous and key Sklar's theorem.

Definition 6. *Copula is a D-dimensional mapping $C : [0, 1]^D \to [0, 1]$ satisfying the following conditions:*

(*i*) $C(u_1, \ldots, u_D) = 0$ *if* $\exists\, d \in \{1, \ldots, D\}$ *such that* $u_d = 0$,

(*ii*) $C(1, \ldots, 1, u_d, 1, \ldots, 1) = u_d, \quad \forall\, d = 1, \ldots, D,\ u_d \in [0, 1]$,

(*iii*) *for all vectors* $\left(u_1^{(1)}, \ldots, u_D^{(1)}\right)^\top$, $\left(u_1^{(2)}, \ldots, u_D^{(2)}\right)^\top$ *in* $[0, 1]^D$ *such that* $u_d^{(1)} \leq u_d^{(2)}\ \forall\, d = 1, \ldots, D$, *it holds that*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_D=1}^{2} (-1)^{i_1 + \cdots + i_D} C(u_1^{(i_1)}, \ldots, u_D^{(i_D)}) \geq 0.$$

Remark. Definition 6 implies that the copula is a cumulative distribution function, defined on a unit hypercube $[0, 1]^D$, of a multivariate distribution whose marginal distributions are uniform on an interval $[0, 1]$.

A copula can be equivalently defined as a cumulative distribution function of a D-dimensional random vector with uniform marginal distributions on $[0, 1]$, see Trivedi et al. [2007].

From the beginning of this chapter, we know that copulas are a tool for connecting multivariate distribution with its marginals. In this part, we describe this connection using the most important theorem in copula theory, Sklar's theorem. This theorem was introduced in 1959 by Abe Sklar, see Sklar [1959]. This thesis uses the formulation of this theorem from Embrechts [2009].

Theorem states that every joint distribution function can be written as a function of its marginals, and this function is a copula. The opposite implication is also true. Copulas can conduct a multivariate distribution using univariate marginals with a prescribed dependency structure.

Theorem 1. *(Sklar's theorem) Let F be a joint distribution function defined on $\mathbb{R}^D$ with marginal distribution functions $F_1, \ldots, F_D$. Then there exists a copula $C : [0, 1]^D \to [0, 1]$ such that for all $(x_1, \ldots, x_D)^\top \in \mathbb{R}^D$ it holds*

$$F(x_1, \ldots, x_D) = C\Big(F_1(x_1), \ldots, F_D(x_D)\Big). \tag{2.1}$$

Furthermore, if $F_1, \ldots, F_D$ are absolutely continuous distribution functions, the copula C is unique. Otherwise, C is uniquely defined only on $Ran(F_1) \times \ldots \times Ran(F_D)$ where $Ran(F_d)$, $d = 1, \ldots, D$, denotes the range of the function F_d.

Conversely, having a copula $C : [0, 1]^D \to [0, 1]$ and marginal distribution functions $F_1, \ldots, F_D$, the function F defined in equation (2.1) is a joint D-dimensional distribution function with marginals $F_1, \ldots, F_D$.

Proof. For the proof of Sklar's theorem, see Carley and Taylor [2002]. □

Remark. In our work, we assume $F_1, \ldots, F_D$ to be absolutely continuous, so the copula C is uniquely defined.

2.2 Modelling of dependence

There are several ways to measure the dependency of two random variables. For example, we can use Pearson's correlation coefficient, Spearman's rho or Kendall's tau. We focus on Kendall's tau, which we find to be intertwined with the copula.

Kendall's tau measures the rank correlation between two random variables. For example, the definition can be found in Embrechts, McNeil, and Straumann [2002] and is following.

Definition 7. *Let X_1 and X_2 be two random variables. Kendall's tau of X_1 and X_2 is defined as*

$$\begin{aligned}\tau(X_1, X_2) &= P\Big((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0\Big) - P\Big((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) < 0\Big) \\ &= E\Big[sign\Big((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2)\Big)\Big],\end{aligned}$$

where $\left(\widetilde{X}_1, \widetilde{X}_2\right)^\top$ is a random vector with the same distribution as $\left(X_1, X_2\right)^\top$ and both are independent.

The main properties of Kendall's tau are summarised in the following theorem.

Theorem 2. *Let X_1 and X_2 be two random variables. Then it holds*

1. *$\tau(X_1, X_2) = \tau(X_2, X_1)$,*
2. *$\tau(X_1, X_2) \in [-1, 1]$,*
3. *if X_1 and X_2 are independent then $\tau(X_1, X_2) = 0$,*
4. *$\tau(X_1, X_2) = 1$ if and only if there exists some nondecreasing transformation T such that $X_2 = T(X_1)$ almost surely (X_1, X_2 are perfectly positively dependent),*

5. $\tau(X_1, X_2) = -1$ *if and only if there exists some non-increasing transformation* T *such that* $X_2 = T(X_1)$ *almost surely* (X_1, X_2 *are perfectly negatively dependent).*

Proof. For the proof see Embrechts et al. [2002, Theorem 3]. □

Perfect positive dependency is sometimes known as a comonotonicity, and perfect negative dependency is known as a countermonotonicity.

The relationship between Kendall's tau and the two-dimensional copula is shown in the following theorem.

Theorem 3. *Let* X_1 *and* X_2 *be two random variables with continuous distribution functions* F_1 *and* F_2*, respectively. Assume* F *as their joint distribution and* C *as a copula. Then it holds*

$$\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2)\, dC(u_1, u_2) - 1.$$

Proof. For the proof see Embrechts et al. [2002, Theorem 3]. □

From Theorem 3, we see that Kendall's tau is a function of a copula C, not of the marginal distributions, which is the main difference from Pearson's correlation coefficient, which is dependent on marginals. For this reason, we find Kendall's tau more convenient.

2.3 Bivariate copulas

So far, we have defined copulas for an arbitrary dimension D. Now we restrict ourselves to $D = 2$. This section describes the commonly used bivariate copulas. Definitions of selected copulas can be found in Trivedi et al. [2007].

2.3.1 Product (independence) copula

The simplest copula is the product copula, also called the independence copula. The bivariate distribution function of two independent random variables is the product of two univariate marginal distribution functions. The product copula is therefore defined as

$$C^{prod}(u_1, u_2) = u_1\, u_2,$$

where $u_1, u_2 \in [0, 1]$.

Theorem 4. *Two random variables are independent if and only if their joint distribution function is a product copula.*

Proof. The theorem can be proven using the well-known relationship between independent random variables X_1, X_2 and their joint distribution function

$F(x_1, x_2)$. It holds that two random variables with distribution functions $F_1(x_1)$ and $F_2(x_2)$ are independent if and only if for all $(x_1, x_2)^\top$ holds

$$F(x_1, x_2) = F_1(x_1)\, F_2(x_2),$$

which proves the theorem.

□

The product copula can be generalised to a D-dimension, $D > 2$ in a straightforward way

$$C^{prod}(u_1, \ldots, u_D) = \prod_{d=1}^{D} u_d,$$

where $u_d \in [0, 1]$ for $d \in 1, \ldots, D$.

2.3.2 Comonotonicity copula

Another simple copula is a comonotonicity copula

$$C^{com}(u_1, u_2) = \min(u_1, u_2),$$

where $u_1, u_2 \in [0, 1]$.

This copula describes a perfect positive dependence. Its Kendall's tau equals one, see following Theorem 5. Comonotonicity copula can be generalised to a D-dimensional $D > 2$ random vector like a product copula.

2.3.3 Countermonotonicity copula

Copula describing perfect negative dependence is known as countermonotonicity copula and is defined as follows

$$C^{count}(u_1, u_2) = \max(u_1 + u_2 - 1, 0),$$

where $u_1, u_2 \in [0, 1]$.

Its Kendall's tau equals minus one, see following Theorem 5. Countermonotonicity copula cannot be generalised to higher dimensions.

The following theorem describes the relationship between Kendall's tau and comonotonicity, resp. countermonotonicity copula.

Theorem 5. *Let X_1 and X_2 be two random variables with continuous distribution functions and copula C. Then it holds*

1. *$\tau(X_1, X_2) = 1$ if and only if $C = C^{com}$,*
2. *$\tau(X_1, X_2) = -1$ if and only if $C = C^{count}$.*

Proof. For the proof see Embrechts et al. [2002, Theorem 3].

□

2.3.4 Clayton copula

We continue with a Clayton copula which is defined as

$$C_\theta^{Cl}(u_1,\, u_2) = \max\Big((u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta},\, 0\Big), \tag{2.2}$$

where $u_1,\, u_2 \in [0,\, 1]$ and $\theta \in [-1,\, \infty) \setminus \{0\}$ is a dependence parameter. If θ approaches minus one, we obtain a countermonotonicity copula, and if θ tends to infinity, we obtain a comonotonicity copula. We get a product copula if θ approaches zero from the right.

Kendall's tau for the Clayton copula can be found in Ghalibaf [2020], and it takes the following form

$$\tau = \frac{\theta}{\theta + 2}, \tag{2.3}$$

where θ is the copula parameter.

To sum up, we can model both negative and positive dependency between variables according to the parameter θ.

Figures 2.1(a), 2.1(b) and 2.1(c) compare densities, defined in (2.5), of the Clayton copula for different parameters θ which correspond to Kendall's tau equal to 0.25, 0.5, 0.75. With the same parameters of θ in Figures 2.2(a), 2.2(b), and 2.2(c) distribution functions are compared. All figures assume uniform marginals. For more information about figures, see Section 2.4.

2.3.5 Gumbel copula

A copula that models only a positive dependence is a Gumbel copula. This copula has the following shape

$$C_\theta^{Gu}(u_1,\, u_2) = \exp\Big(-\Big((-\log u_1)^\theta + (-\log u_2)^\theta\Big)^{1/\theta}\Big),$$

where $u_1,\, u_2 \in [0,\, 1]$ and $\theta \in [1,\, \infty)$ is a dependence parameter. If $\theta = 1$, the Gumbel copula refers to a product copula, and if θ tends to infinity, we obtain a comonotonicity copula.

Kendall's tau for the Gumbel copula can also be found in Ghalibaf [2020], and it is of the following form

$$\tau = 1 - \frac{1}{\theta},$$

where θ is a copula parameter. The Gumbel copula cannot be used for modelling negative dependency.

Figures 2.1(d), 2.1(e) and 2.1(f) compare densities, defined in (2.5), of the Gumbel copula for different parameters θ which correspond to Kendall's tau equal to 0.25, 0.5, 0.75. With the same parameters in Figures 2.2(d), 2.2(e) and 2.2(f) distribution functions are compared. All figures assume uniform marginals. For more information about figures, see Section 2.4.

The Clayton and the Gumbel copula belong to the so-called Archimedean copulas. The definition of Archimedean copulas follows.

Definition 8. *Bivariate Archimedean copula is copula defined as*

$$C(u_1,\, u_2) = \phi^{[-1]}\Big(\phi(u_1) + \phi(u_2)\Big),$$

where $\phi : [0,\, 1] \to [0,\, \infty]$ *is a continuous, decreasing and convex function satisfying* $\phi(1) = 0$. *Further* $\phi^{[-1]}$ *is pseudo-inverse function of* ϕ *defined as*

$$\phi^{[-1]}(x) = \begin{cases} \phi^{-1}(x), & x \in [0,\, \phi(0)], \\ 0, & x \in (\phi(0),\, \infty]. \end{cases}$$

The function $\phi(x)$ *is the Archimedean copula's generator function.*

Clayton copula can be obtained using $\phi(x) = \frac{1}{\theta}(x^{-\theta} - 1)$ and the Gumbel copula using $\phi(x) = (-\log x)^{\theta}$.

2.3.6 Gaussian (normal) copula

A Gaussian copula, sometimes called a normal copula, is derived from the multivariate normal distribution

$$C_{\rho}^{Gauss}(u_1,\, u_2) = \Phi_R^2\Big(\Phi^{-1}(u_1),\, \Phi^{-1}(u_2)\Big),$$

where $u_1,\, u_2 \in (0,\, 1)$, Φ_R^2 is the joint cumulative distribution function of a bivariate normal distribution with zero mean vector and covariance matrix

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where $\rho \in [-1,\, 1]$. Φ^{-1} is the quantile function of the standard normal distribution.

Kendall's tau for the Gaussian copula is

$$\tau = \frac{2}{\pi} \arcsin(\rho),$$

see Ghalibaf [2020].

The Gaussian copula allows modelling both positive dependence (for positive ρ) and negative dependence (for negative ρ). If $\rho = 1$, the random variables are comonotonic, if $\rho = -1$, the random variables are countermonotonic.

Figures 2.1(g), 2.1(h) and 2.1(i) compare densities, defined in (2.5), of the Gaussian copula for different parameters ρ which correspond to Kendall's tau equal to 0.25, 0.5, 0.75. With the same parameters in Figures 2.2(g), 2.2(h) and 2.2(i) distribution functions are compared. All figures assume uniform marginals. For more information about figures, see Section 2.4.

Gaussian copula can be generalised to a D-dimension $D > 2$ in the same way as the product copula.

Remark. A bivariate distribution with a distribution function defined by the Gaussian copula

$$F(x_1,\, x_2) = C_R^{Gauss}\Big(F_1(x_1),\, F_2(x_2)\Big)$$

for arbitrary marginal distribution functions $F_1(x)$ and $F_2(x)$ is not generally a bivariate normal distribution. However, it can be easily derived that if $F_1(x)$ and $F_2(x)$ are normal distribution functions, then the bivariate distribution given by the above copula is bivariate normal.

2.3.7 Student's t-copula

The last copula we describe is a Student's t-copula or simply t-copula. This copula is based on multivariate t-distribution, and it is defined similarly to the normal copula as

$$C^t_{\nu,\rho}(u_1, u_2) = T_{\nu,R}\Big(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)\Big), \tag{2.4}$$

where $u_1, u_2 \in (0, 1)$, $T_{\nu,R}$ is the joint cumulative distribution function of a bivariate t-distribution with covariance matrix

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where $\rho \in [-1, 1]$ and with ν degree of freedom, t_ν^{-1} is the quantile function of the Student t-distribution with ν degree of freedoms. For the definition of multivariate t-distribution, you can see, for example, Kibria and Joarder [2006].

Kendall's tau for Student's t-copula is the same as for the Gaussian copula and can be found Ghalibaf [2020]

$$\tau = \frac{2}{\pi} \arcsin(\rho),$$

Similarly to the normal copula, we can model both positive and negative dependence.

Figures 2.1(j), 2.1(k) and 2.1(l) compare densities, defined in (2.5), of the Student's t-copula for different parameters ρ which correspond to Kendall's tau equal to 0.25, 0.5, 0.75. The degree of freedom is fixed as $\nu = 5$, with the same parameters in Figures 2.2(j), 2.2(k) and 2.2(l) distribution functions are compared. All figures assume uniform marginals. For more information about figures, see Section 2.4.

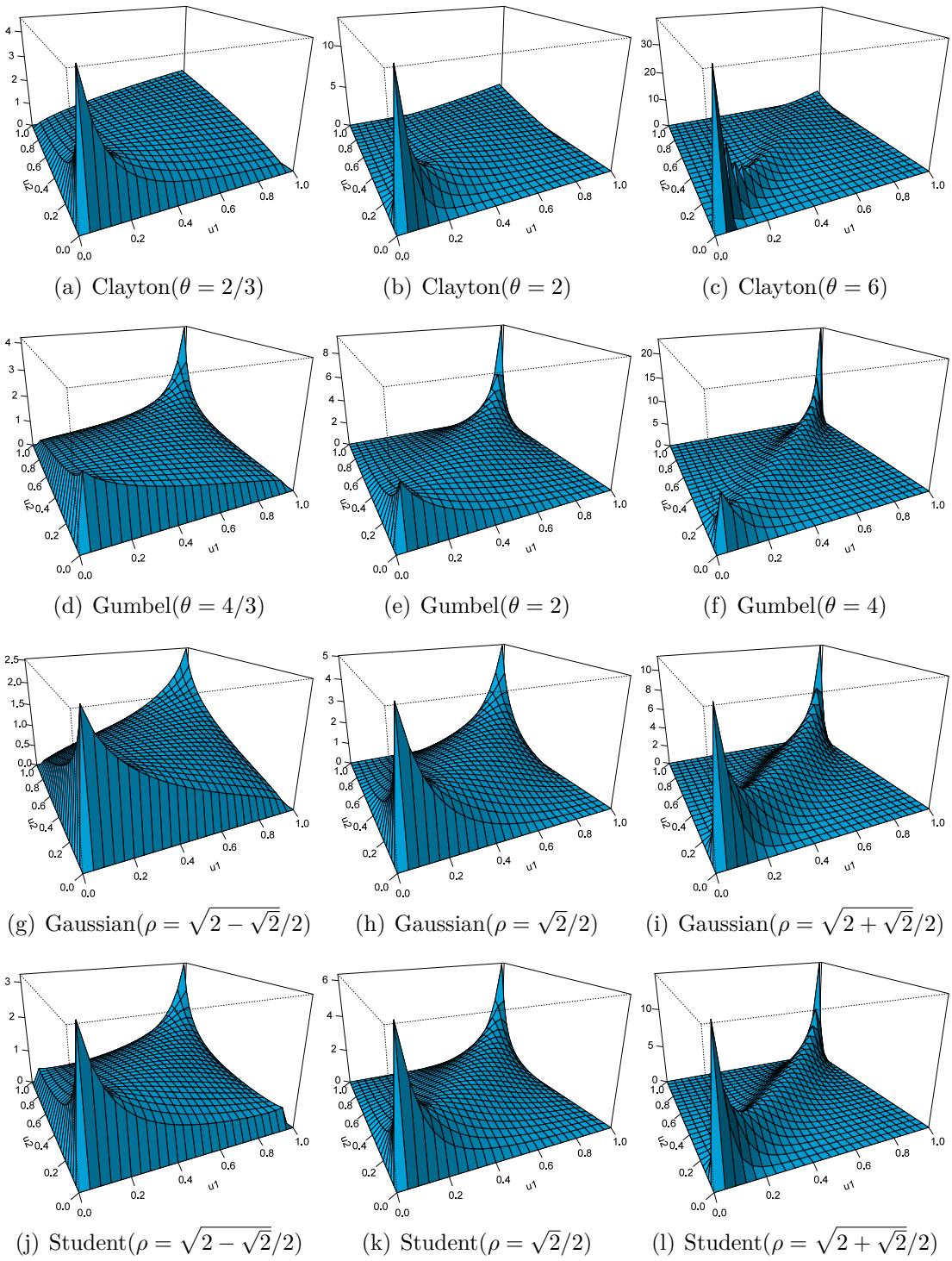
Student's t-copula can be generalised to a D-dimension $D > 2$ in the same way as the product copula.

2.4 Copula visualization

Perspective plots of density and cumulative distribution functions of Clayton, Gumbel, Gaussian and Student's t ($\nu = 5$) copulas are visualised in this section. The density of a given copula is written in the following equation (2.5). Copulas are compared with different parameters (θ for Clayton and Gumbel copulas or ρ for Gaussian and Student's t-copulas) that correspond to the same Kendall's tau $\tau \in \{0.25, 0.50, 0.75\}$. Copula parameters corresponding to prescribed Kendall's tau values can be found in Table 2.1.

In Figure 2.1, densities are compared, and in 2.2, cumulative distribution functions of copulas are compared. In both figures, each row contains three figures for the same copula with different parameters that correspond to different Kendall's tau. Uniform marginals are assumed for all the figures.

Finally, we show the distribution functions and densities of the Clayton, see Figure 2.3, and Gumbel, see 2.4, copulas with non-uniform marginals, specifically with different combinations of normal distribution and exponential distribution.



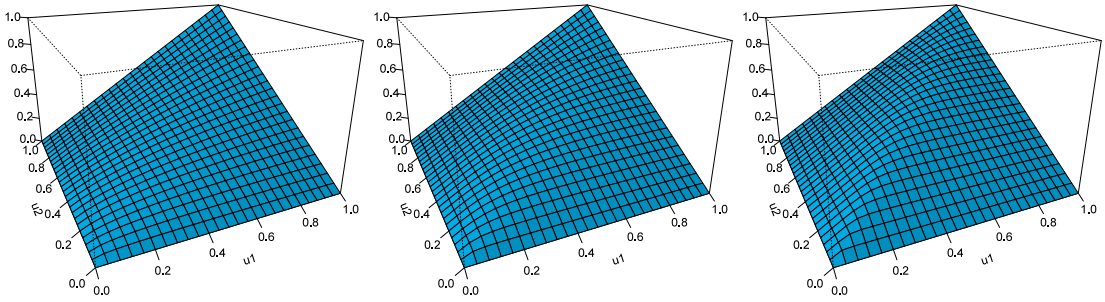
(a) Clayton($\theta = 2/3$) (b) Clayton($\theta = 2$) (c) Clayton($\theta = 6$)

(d) Gumbel($\theta = 4/3$) (e) Gumbel($\theta = 2$) (f) Gumbel($\theta = 4$)

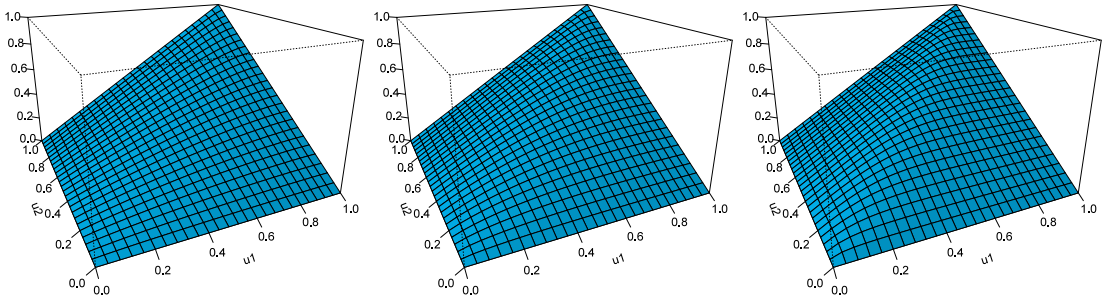
(g) Gaussian($\rho = \sqrt{2 - \sqrt{2}}/2$) (h) Gaussian($\rho = \sqrt{2}/2$) (i) Gaussian($\rho = \sqrt{2 + \sqrt{2}}/2$)

(j) Student($\rho = \sqrt{2 - \sqrt{2}}/2$) (k) Student($\rho = \sqrt{2}/2$) (l) Student($\rho = \sqrt{2 + \sqrt{2}}/2$)

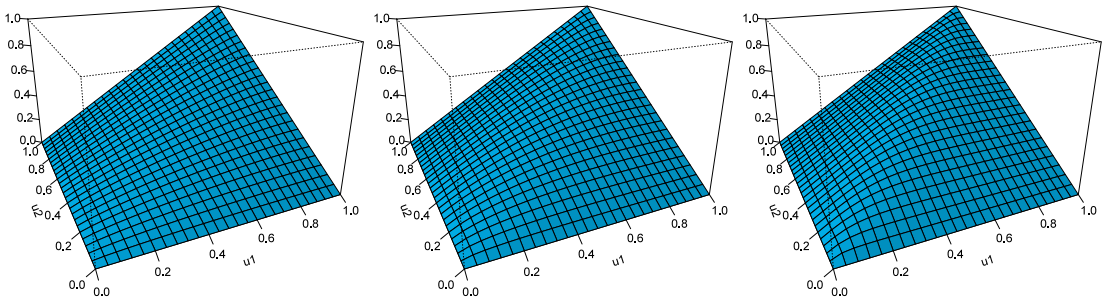
Figure 2.1: Density of different copulas with different parameters. The first column corresponds to Kendall's tau equal to 0.25, second to 0.50 and third to 0.75.



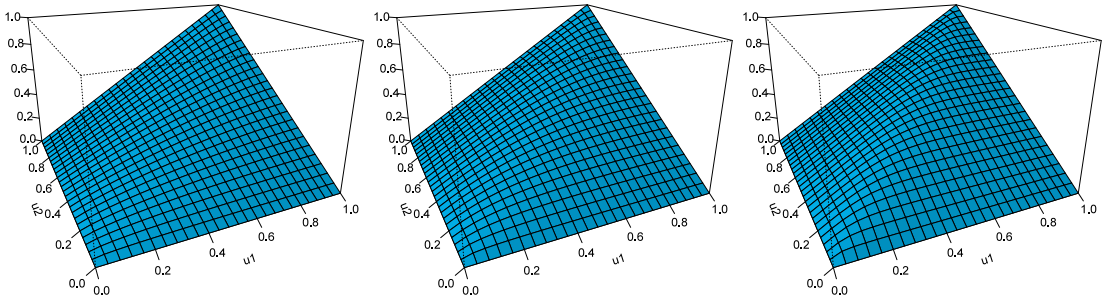
(a) Clayton($\theta = 2/3$) (b) Clayton($\theta = 2$) (c) Clayton($\theta = 6$)



(d) Gumbel($\theta = 4/3$) (e) Gumbel($\theta = 2$) (f) Gumbel($\theta = 4$)

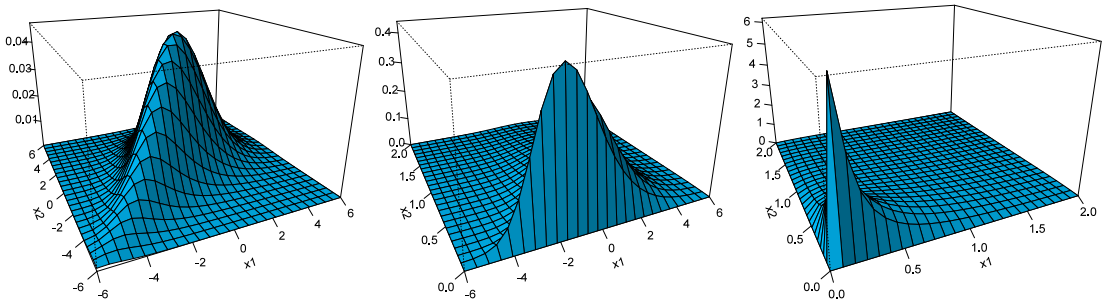


(g) Gaussian($\rho = \sqrt{2-\sqrt{2}}/2$) (h) Gaussian($\rho = \sqrt{2}/2$) (i) Gaussian($\rho = \sqrt{2+\sqrt{2}}/2$)

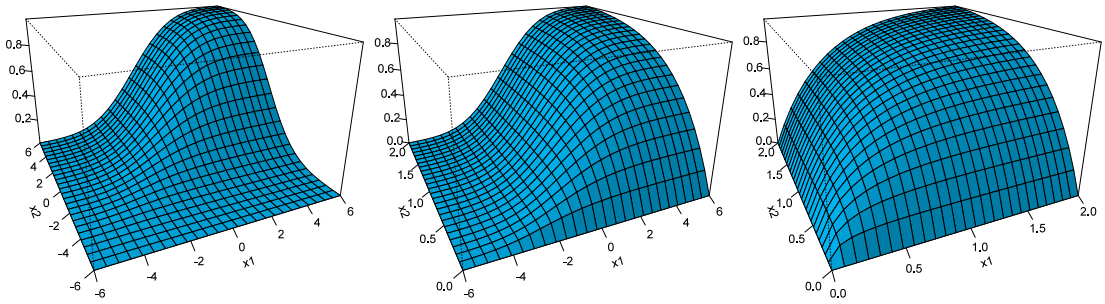


(j) Student($\rho = \sqrt{2-\sqrt{2}}/2$) (k) Student($\rho = \sqrt{2}/2$) (l) Student($\rho = \sqrt{2+\sqrt{2}}/2$)

Figure 2.2: Cumulative distribution function of different copulas with different parameters. The first column corresponds to Kendall's tau equal to 0.25, second to 0.50 and third to 0.75.

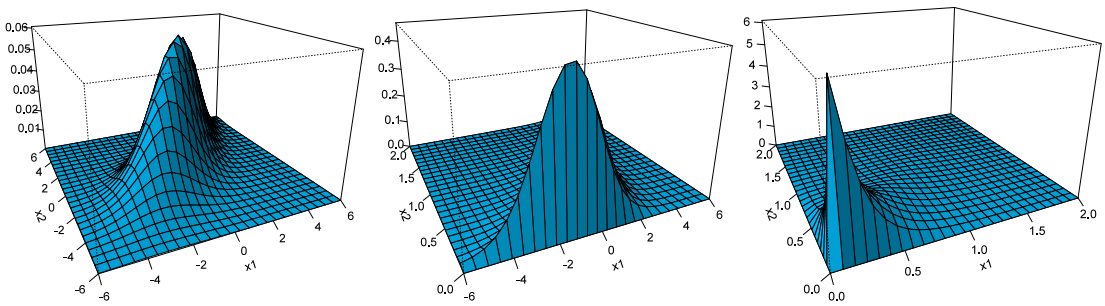


(a) Clayton(Norm,Norm) (b) Clayton(Norm,Exp) (c) Clayton(Exp,Exp)

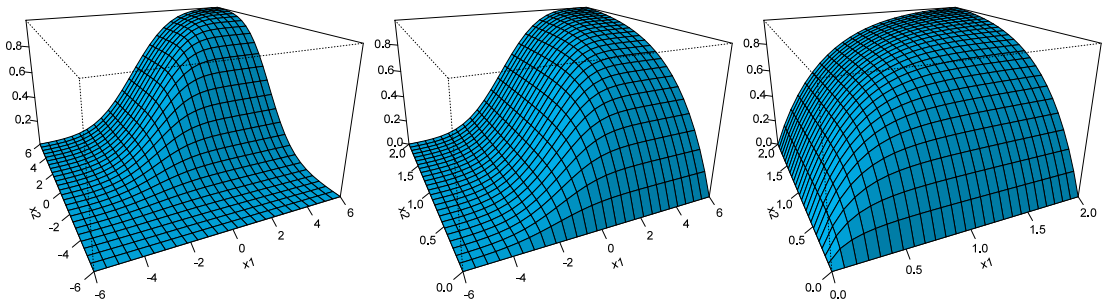


(d) Clayton(Norm,Norm) (e) Clayton(Norm,Exp) (f) Clayton(Exp,Exp)

Figure 2.3: Density on the first row and distribution function on the second row of Clayton copula with $\theta = 1$ and with different margins which are normal distribution with zero mean, variance equal to two and exponential distribution with mean 1/2.



(a) Gumbel(Norm,Norm) (b) Gumbel(Norm,Exp) (c) Gumbel(Exp,Exp)



(d) Gumbel(Norm,Norm) (e) Gumbel(Norm,Exp) (f) Gumbel(Exp,Exp)

Figure 2.4: Density on the first row and distribution function on the second row of Gumbel copula with $\theta = 2$ and with different margins which are normal distribution with zero mean, variance equal to two and exponential distribution with mean 1/2.

Copula	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
Clayton	$\theta = 2/3$	$\theta = 2$	$\theta = 6$
Gumbel	$\theta = 4/3$	$\theta = 2$	$\theta = 4$
Gaussian	$\rho = 1/2\sqrt{2-\sqrt{2}}$	$\rho = \sqrt{2}/2$	$\rho = 1/2\sqrt{2+\sqrt{2}}$
Student's t	$\rho = 1/2\sqrt{2-\sqrt{2}}$	$\rho = \sqrt{2}/2$	$\rho = 1/2\sqrt{2+\sqrt{2}}$

Table 2.1: Copula parameters corresponding to the specified Kendall's tau.

2.5 Copula estimation

The estimation section is one of the key parts of the copula chapter. Various procedures for estimating the copula together with its marginals are presented. We restrict ourselves to bivariate copulas. However, the procedure can be easily generalised to a higher dimension $D > 2$. This chapter is based on Choroś, Ibragimov, and Permiakova [2010].

We describe three main approaches for estimation: fully parametric, semi-parametric and fully non-parametric. The fully parametric approach estimates marginal distributions and copula parameters, usually using the maximum likelihood. In the semi-parametric approach, we first non-parametrically estimate marginal distributions, and then the copula parameter is estimated parametrically. The maximum likelihood approach or estimation based on Kendall's tau can be used to estimate the copula parameter. Finally, we present a fully non-parametric approach.

Let us assume that we have bivariate *i.i.d.* data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, where $\boldsymbol{X}_i = (X_{1,i}, X_{2,i})^\top$ for $i = 1, \ldots, n$, from a bivariate continuous distribution with a cumulative distribution function $F(x_1, x_2)$ with marginals $F_1(x_1)$ and $F_2(x_2)$. Due to Sklar's theorem, see Theorem 1, there exists a unique copula C such that

$$F(x_1, x_2) = C\Big(F_1(x_1), F_2(x_2)\Big).$$

Our main aim is to estimate marginal distributions and a copula C.

In the parametric and semi-parametric approach, it is further assumed that $C \in \mathcal{C} = \{C_\theta : \theta \in \Theta\}$, where Θ is a set of all possible values for copula parameter.

2.5.1 Parametric estimation

In the first case, the model is described by unknown parameters. The maximum likelihood method is the most commonly used method for estimation. The density of data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is needed for likelihood.

Theorem 6. *Assume the copula C_θ to be absolutely continuous with density*

$$c_\theta(u_1, u_2) = \frac{\partial^2 C_\theta(u_1, u_2)}{\partial u_1 \, \partial u_2}. \tag{2.5}$$

Denote f the joint density of the distribution given by cumulative distribution function F and f_1, f_2 marginal densities of distributions given by the cumulative

functions F_1, F_2. The density f can be represented as the following multiplication

$$f(x_1,\, x_2) = c_\theta\Big(F_1(x_1),\, F_2(x_2)\Big) \prod_{d=1}^{2} f_d(x_d). \tag{2.6}$$

Proof. From Sklar's theorem, see Theorem 1, and chain rule, we have

$$f(x_1,\, x_2) = \frac{\partial^2 F(x_1,\, x_2)}{\partial x_1\, \partial x_2} = \frac{\partial^2 C_\theta\Big(F_1(x_1),\, F_2(x_2)\Big)}{\partial x_1\, \partial x_2} = \\ \frac{\partial^2 C_\theta\Big(F_1(x_1),\, F_2(x_2)\Big)}{\partial F_1(x_1)\, \partial F_2(x_2)}\, \frac{\partial F_1(x_1)\, \partial F_2(x_2)}{\partial x_1\, \partial x_2}.$$

The first term can be simplified to the copula density due to (2.5), and the second can be divided into the product of the marginal densities. To sum up, the joint density can be written as

$$f(x_1,\, x_2) = c_\theta\Big(F_1(x_1),\, F_2(x_2)\Big) \prod_{d=1}^{2} f_d(x_d).$$

□

Theorem 6 can be generalised to a higher dimension $D > 2$

$$f(x_1,\, \ldots,\, x_D) = c_\theta\Big(F_1(x_1),\, \ldots,\, F_D(x_D)\Big) \prod_{d=1}^{D} f_d(x_d),$$

which can be proved similarly to Theorem 6.

Let us denote $\boldsymbol{\alpha}_1 \in \boldsymbol{A}_1$ and $\boldsymbol{\alpha}_2 \in \boldsymbol{A}_2$ vectors of the unknown parameters of marginal distributions F_1, F_2, respectively, and its spaces $\boldsymbol{A}_1$, $\boldsymbol{A}_2$. Using (2.6) we can derive the log-likelihood

$$\ell(\boldsymbol{\alpha}_1,\, \boldsymbol{\alpha}_2,\, \theta) = \sum_{i=1}^{n} \log f(X_{1,i},\, X_{2,i}).$$

We divide this sum into two parts

$$\ell(\boldsymbol{\alpha}_1,\, \boldsymbol{\alpha}_2,\, \theta) = \sum_{i=1}^{n} \log f(X_{1,i},\, X_{2,i}) = \sum_{i=1}^{n} \log c_\theta\Big(F_1(X_{1,i}),\, F_2(X_{2,i})\Big) + \sum_{d=1}^{2} \ell_d(\boldsymbol{\alpha}_d), \tag{2.7}$$

where

$$\ell_d(\boldsymbol{\alpha}_d) = \sum_{i=1}^{n} \log f_d(X_{d,i}), \quad d = 1,\, 2,$$

are marginal log-likelihoods. The MLE of the parameters are then derived as arguments which maximise the log-likelihood (2.7)

$$\Big(\widehat{\boldsymbol{\alpha}}_1,\, \widehat{\boldsymbol{\alpha}}_2,\, \widehat{\theta}\Big)^\top = \underset{\boldsymbol{\alpha}_1 \in \boldsymbol{A}_1,\, \boldsymbol{\alpha}_2 \in \boldsymbol{A}_2,\, \theta \in \Theta}{\arg\max}\ \ell(\boldsymbol{\alpha}_1,\, \boldsymbol{\alpha}_2,\, \theta).$$

An estimator based on the method of maximum likelihood is consistent and asymptotically normal under regularity conditions, which can be found, for example, in Akahira and Takeuchi [2012].

It is possible to find a maximum likelihood estimator for all parameters simultaneously, as described above. However, sometimes it is complicated due to a large number of parameters. In practice, it is more common first to estimate the marginal distribution parameters $\boldsymbol{\alpha}_1 \in \boldsymbol{A}_1$ and $\boldsymbol{\alpha}_2 \in \boldsymbol{A}_2$ and then estimate the copula parameter $\theta \in \Theta$ with the marginal parameters fixed from the first stage. It is seen that the log-likelihood in (2.7) consists of two main parts, so it is possible to maximise one part first and then the second. For more information on the two-stage estimation procedure, see Joe [2005].

2.5.2 Semi-parametric estimation

Semi-parametric estimation is the most commonly used estimation approach in practice. To begin with, we estimate marginal distributions F_1 and F_2 first. We can use different estimation methods. However, the most popular is the empirical distribution function which is defined as

$$\widehat{F}_d(t) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}\{X_{d,i} \leq t\}, \quad t \in \mathbb{R},\, d = 1,\, 2. \tag{2.8}$$

Here we divide by $n+1$ to avoid possible ambiguities on the unit edges of the copula. Due to this divider, we make the value $\widehat{F}_d(t) < 1$ for all possible $t \in \mathbb{R}$. From now on, this function will be called the adjusted empirical distribution function.

For the estimation of the copula parameter, we have two options. We can estimate it using maximum likelihood or using Kendall's tau. The unknown copula parameter is denoted as θ.

Estimation based on maximum likelihood

As a first method, we can use estimated distribution functions $\widehat{F}_1$ and $\widehat{F}_2$ for creating pseudo-sample $\widehat{\boldsymbol{U}}_1, \ldots, \widehat{\boldsymbol{U}}_n$ where for $i = 1, \ldots, n$ it holds

$$\widehat{\boldsymbol{U}}_i = \Big(\widehat{F}_1(X_{1,i}),\, \widehat{F}_2(X_{2,i})\Big)^\top = \Big(\widehat{U}_{1,i},\, \widehat{U}_{2,i}\Big)^\top.$$

Finally, we can estimate the copula parameter θ by maximising the likelihood

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} c_\theta(\widehat{\boldsymbol{U}}_i),$$

where c_θ is the density of C_θ from (2.5).

Usually, it is more convenient to maximise log-likelihood

$$\ell(\theta) = \log \mathcal{L}(\theta) = \sum_{i=1}^{n} \log\Big(c_\theta(\widehat{\boldsymbol{U}}_i)\Big). \tag{2.9}$$

Finally, the estimator is found as an argument of the maxima

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \ell(\theta),$$

where Θ is the set of all possible values of the copula parameter θ.

The resulting estimator $\widehat{\theta}$ is consistent and asymptotically normal under suitable regularity conditions, for more information, see Choroś et al. [2010].

Estimation based on Kendall's tau

The second method is based on Kendall's tau. This method requires the expression for the relationship between the copula's parameter and Kendall's tau. Let us assume that

$$\tau = f(\theta),$$

where f is an invertible function. These functions can be found for Clayton copula, Gumbel copula, Gaussian copula and Student's t-copula in Section 2.3. Subsequently, we estimate Kendall's tau from the data as

$$\widehat{\tau} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \operatorname{sign}\Big((X_{1,i} - X_{1,j})(X_{2,i} - X_{2,j})\Big) \tag{2.10}$$

and we solve an equation

$$\widehat{\tau} = f(\widehat{\theta}),$$

which is equivalent to

$$\widehat{\theta} = f^{-1}(\widehat{\tau}).$$

The properties of the estimator $\widehat{\theta}$ are summarised in the following theorem.

Theorem 7. *The estimator $\widehat{\theta}$ estimated by a procedure based on Kendall's tau is consistent and asymptotically normal.*

Proof. The consistency of $\widehat{\theta}$ follows from the consistency of $\widehat{\tau}$ shown in Gibbons and Chakraborti [2011, Chapter 11]. The estimator $\widehat{\tau}$ is unbiased, and its variance approaches zero as n tends to infinity, so the estimator $\widehat{\tau}$ is consistent due to Chebyshev's inequality. The consistency of $\widehat{\theta}$ comes from the Continuous mapping theorem.

The asymptotic normality of $\widehat{\theta}$ follows from the asymptotic normality of $\widehat{\tau}$, which is again shown in Gibbons and Chakraborti [2011, Chapter 11]. The statistic $\widehat{\tau}$ is U-statistics, so the normality of $\widehat{\tau}$ follows from the general theory of U-statistics. For the definition and normality of U-statistics, see Van der Vaart [2000]. The asymptotic normality of $\widehat{\theta}$ then follows from the Delta method, see Van der Vaart [2000].

□

2.5.3 Non-parametric estimation

Finally, the non-parametric approach is described. This method can be used if we want to avoid making restrictive assumptions about the parametric shape of the copula. This estimator is based on empirical distribution functions. Let us denote $\widehat{F}$ the bivariate empirical distribution function

$$\widehat{F}(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_{1,i} \le t_1,\, X_{2,i} \le t_2\}, \quad (t_1, t_2)^\top \in \mathbb{R}^2, \tag{2.11}$$

and $\widehat{F}_1^{-1}(u_1)$, $\widehat{F}_2^{-1}(u_2)$ be the empirical quantile functions of F_1, respectively F_2.

Functions $\widehat{F}_1^{-1}(u_1)$ and $\widehat{F}_2^{-1}(u_2)$ are defined as an inversion of the empirical distribution function

$$\widehat{F}_d(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_{d,i} \leq t\}, \quad t \in \mathbb{R},\, d = 1,\, 2.$$

However, it is possible to define $\widehat{F}_1^{-1}(u_1)$ and $\widehat{F}_2^{-1}(u_2)$ as an inversion of an adjusted empirical distribution function defined in (2.8).

Then the non-parametric empirical copula $\widehat{C}(u_1,\, u_2)$ is derived

$$\widehat{C}(u_1,\, u_2) = \widehat{F}\Big(\widehat{F}_1^{-1}(u_1),\, \widehat{F}_2^{-1}(u_2)\Big), \quad (u_1,\, u_2)^\top \in [0,1]^2.$$

The properties of this estimate, such as consistency and asymptotic normality, will vary according to the situation. Further, details can be found in Choroś et al. [2010].

Definition of empirical copula $\widehat{C}(u_1,\, u_2)$ can be equivalently written into the following form

$$\widehat{C}(u_1,\, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\widehat{U}_{1,i} \leq u_1,\, \widehat{U}_{2,i} \leq u_2\}, \quad (u_1,\, u_2)^\top \in [0,1]^2, \tag{2.12}$$

where pseudo-sample $\widehat{\boldsymbol{U}}_1,\, \ldots,\, \widehat{\boldsymbol{U}}_n$ is created in the same way as in Section 2.5.2 using empirical cumulative distribution function defined in (2.11). Note that this pseudo-sample was created non-parametrically.

Also, it is possible to construct empirical copula using pseudo-sample created by parametric estimation of their distribution. However, this is no longer a non-parametric estimate.

There exist other methods how we can make a non-parametric estimation. For example, estimation based on kernels. More information about kernels and estimation based on it can be found in Chen and Huang [2007].

2.6 Copula selection

In the previous part, the model was estimated. However, we assumed that the copula family was given in Sections 2.5.1 and 2.5.2. This section discusses the methods for selecting the most suitable copula family for the model.

One general method is to estimate several copulas and then use the data to compare which copula fits best. It is easier to use the Akaike information criterion (AIC) or Bayesian information criterion (BIC) for the decision. Also, it is possible to perform a goodness-of-fit (GOF) test. The AIC and BIC are defined in the Definition 3 and in the following Section 3.3, a goodness-of-fit test for time series data is presented.

Applying these criteria to copula models will be illustrated in more detail in the next Section 3.

Another method which can be used for copula selection is restricted only to Archimedean copulas. This is based on the generator function $\phi(x)$ defined in Definition 8. The whole procedure can be found in Trivedi et al. [2007] in the Copula Evaluation and Selection part.

2.7 Example

In the last part of the chapter, we show an example. The main goal is to perform all estimating procedures presented above. Statistical calculations were performed using R software, see R Core Team [2020], specifically for copula work, we chose the copula package, created by Hofert, Kojadinovic, Maechler, and Yan [2022]. We attach the script Copulas.R to this example.

To demonstrate the estimation methods, we generate data with a sample size of $n = 150$ from a bivariate distribution created by Clayton copula with normal and exponential margins. Formally, let us assume we have *i.i.d.* bivariate data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{150}$, $\boldsymbol{X}_i = (X_{1,i}, X_{2,i})^\top$, $i = 1, \ldots, 150$ fulfilling

$$\begin{aligned} &\boldsymbol{X}_i \sim C^{Cl}_{\theta=1}\Big(F_1(x_1), F_2(x_2)\Big), \quad x_1 \in \mathbb{R}, x_2 \in \mathbb{R}^+_0, i = 1, \ldots, 150, \\ &F_1(x_1) \sim \mathcal{N}(0, 4), \quad F_2(x_2) \sim \mathsf{Exp}(\lambda = 2), \end{aligned}$$

where $C^{Cl}_{\theta=1}(u_1, u_2)$ denote the Clayton copula with parameter $\theta = 1$, for the definition of Clayton copula see equation (2.2). The true distribution function of this bivariate distribution is plotted in Figure 2.5(a) and the density in Figure 2.5(b).

Figure 2.6 provides scatter plots of the simulated data. Sub-figure 2.6(a) compares simulated data with the contour of the true copula density with the specified marginals, and sub-figure 2.6(b) compares true pseudo-observed data created from simulated data using true marginals with the contour of the density of the Clayton copula with parameter $\theta = 1$.

Figure 2.7 shows the marginal histograms with their true densities - red curves. In sub-figure 2.7(a), we can check that the first variable has a normal distribution. In Figure 2.7(b), we can check the second marginal, which has an exponential distribution.

The estimation of Kendall's tau is expressed by an equation (2.10)

$$\widehat{\tau} = 0.2546.$$

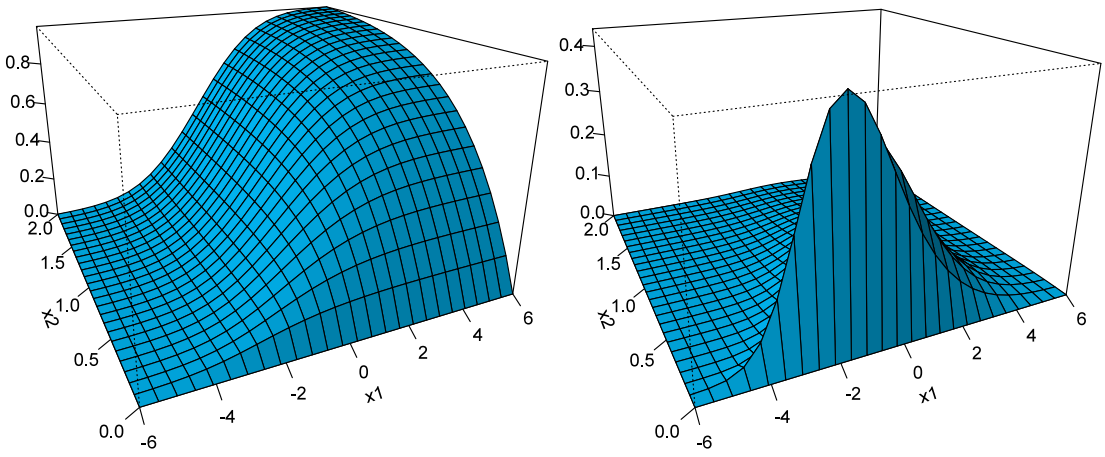
The true value is from (2.3) equal to 1/3.

Parametric estimation

The maximum likelihood (ML) is used for parametric estimation. We perform the estimation described in Section 2.5.1, specifically, we perform a two-stage estimation.

Firstly, it is necessary to find the distribution family of marginals. From the histograms in Figure 2.7, we can conclude that we have a normal and exponential distribution. It is easily derived that maximum likelihood estimation (MLE) for normal distribution is

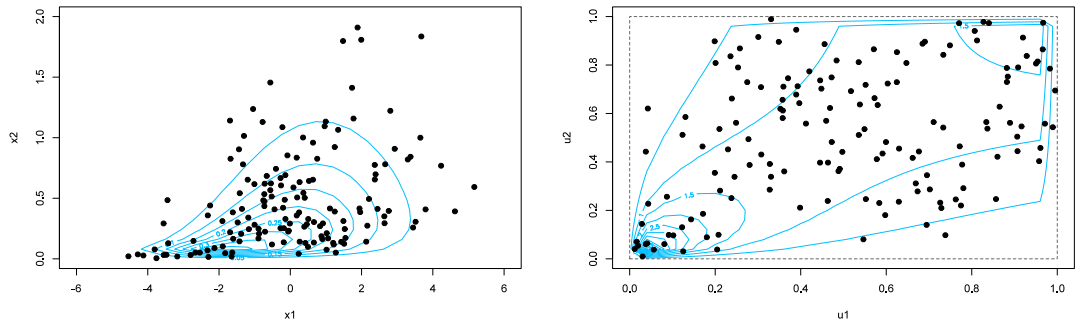
$$\begin{aligned} \widehat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_{i,1} = \overline{X}_{n,1} = 0.0298, \\ \widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\Big(X_{i,1} - \overline{X}_{n,1}\Big)^2 = 3.7724. \end{aligned}$$



(a) Distribution function

(b) Probability density function

Figure 2.5: Bivariate distribution generated by Clayton copula with normal and exponential marginals.



(a) Scatter plot of simulated data

(b) Scatter plot of true pseudo-observed simulated data

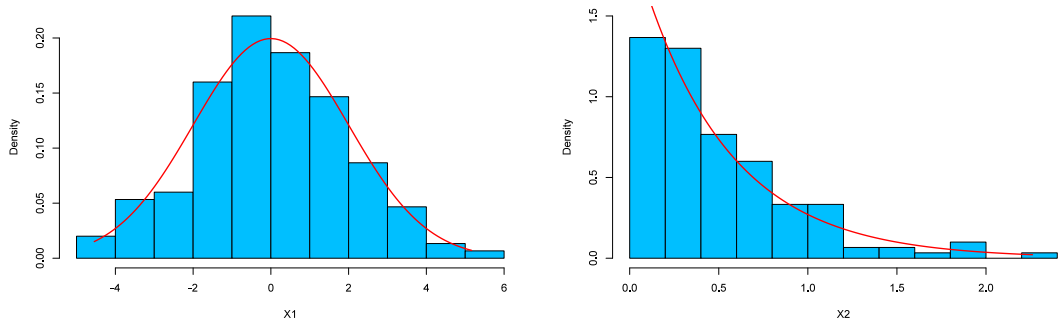
Figure 2.6: Scatter plots of simulated data and true pseudo-observed simulated data with contours of true densities.

MLE for exponential distribution follows

$$\widehat{\lambda} = \frac{1}{\overline{X}_{n,2}} = 2.0505.$$

The second stage is an estimation of a copula parameter. Table 2.2 compares the estimation of Clayton, Gumbel, Gaussian and Student's t-copula parameters and their 95 % confidence intervals based on the assumption of normality.

A note on the estimated degrees of freedom of the Student's t-distribution would be helpful here. From Table 2.2, the degree of freedom of the Student's t-copula is estimated to be $\widehat{\nu} = 10178.3842$. It is well known that the Student's t-distribution converges to a normal distribution as the degree of freedom tends to infinity. From the prescription of the Student's t-copula, see equation (2.4), it is easily seen that the Student's t-copula converges to a Gaussian copula when the degree of freedom tends to infinity. Thus the Gaussian and Student's t fitted copulas are similar.



(a) Histogram of the first variable

(b) Histogram of the second variable

Figure 2.7: Histograms of the marginals compared with the true density - red curve.

Table 2.3 compares the values of AIC and BIC. These values were calculated from the maximum log-likelihood using equations (1.7) for AIC and (1.8) for BIC defined in Definition 3. The log-likelihood used for the calculation corresponds to the log-likelihood of the copula from the second stage of estimation, where we already assume the estimated parameters of the marginal distributions. The log-likelihood is computed

$$\ell(\theta) = \sum_{i=1}^{n} \log c_\theta\Big(\widehat{F}_1(X_{1,i}),\ \widehat{F}_2(X_{2,i})\Big),$$

where $c_\theta(u_1,\, u_2)$ is the density of examined copula (Clayton, Gumbel, Gaussian and Student's t) and $\widehat{F}_1$, resp. $\widehat{F}_2$ are already estimated distribution functions from the first stage. It can be seen that the Clayton copula, which is the true copula, has the lowest AIC and also BIC.

It is possible to plot the estimated copula and compare it to the true copula, but the two plots would be almost identical, so this step is omitted.

Copula	Parameter estimation	Standard deviation	95 % confidence interval 2.5 %	97.5 %
Clayton	$\widehat{\theta} = 0.9139$	0.1587	0.6028	1.2250
Gumbel	$\widehat{\theta} = 1.3187$	0.0846	1.1530	1.4845
Gaussian	$\widehat{\rho} = 0.4795$	0.0596	0.3627	0.5962
Student's t	$\widehat{\rho} = 0.4794$	0.0596	0.3626	0.5962
	$\widehat{\nu} = 10262.2863$	42.8080	10178.3842	10346.1884

Table 2.2: Fully parametric estimation of an unknown parameter of different copulas.

Copula	AIC	BIC
Clayton	-44.0498	-41.0391
Gumbel	-17.7009	-14.6903
Gaussian	-31.9777	-28.9671
Student's t	-29.9768	-23.9555

Table 2.3: Comparison of AIC and BIC values for fully parametric estimation.

Semi-parametric estimation

The estimation procedure based on Section 2.5.2 differs from the fully parametric estimator in the non-parametric estimation of marginals. Marginals are estimated using an empirical cumulative distribution function. The estimation of the copula parameter is based on ML or Kendall's tau. Table 2.4 summarises both estimation procedure results. The first part of the table summarises the results based on the ML method, and the bottom part summarises the results based on Kendall's tau.

In Table 2.4 of the ML section, the degrees of freedom of the Student's t-copula are estimated as $\widehat{\nu} = 2814.9094$. Similar to the parametric estimator above, it can be concluded that the Student's t-copula and the Gaussian copula estimated by ML are similar.

Note that the Student's t-copula estimation based on Kendall's tau assumes a fixed degrees of freedom default set to 4.

The table of estimates, see 2.4, shows that the MLE for Clayton copula is less biased than the estimation based on Kendall's tau. However, it has a higher standard deviation.

Table 2.5 contains AIC and BIC for the MLE. Their calculation is again based on the equation (1.7) for AIC and (1.8) for BIC defined in Definition 3. The log-likelihood calculation is based on the equation (2.9). The lowest AIC and also BIC are for Clayton copula, which is the true copula.

The advantage of a method based on ML is providing AIC and BIC values.

Non-parametric estimation

Lastly, non-parametric estimation is shown. Figure 2.8 compares a true distribution with an empirical one. In Figure 2.9, the empirical copula, see (2.12), is plotted with its true version.

Conclusion of the example

The main purpose of the example was to compare different estimation methods. The fully parametric estimator came closest to the actual values. However, to say this is the best method, we would have to make simulations and compare whether this is the case in the other examples. Of course, it also depends on how the estimation is done, a fully parametric estimator is not always possible. Table 2.6 compares all the estimated parameters for the true Clayton copula. A disadvantage of a fully parametric model may be, for example, a poorly chosen family of distributions.

Copula	Parameter estimation	Standard deviation	95 % confidence interval 2.5 %	97.5 %
Estimation based on ML				
Clayton	$\widehat{\theta} = 0.8358$	0.2213	0.4022	1.2695
Gumbel	$\widehat{\theta} = 1.3003$	0.0816	1.1404	1.4603
Gaussian	$\widehat{\rho} = 0.4517$	0.0735	0.3075	0.5958
Student's t	$\widehat{\rho} = 0.4512$ $\widehat{\nu} =$ 2814.9094	0.0737	0.3067	0.5956
Estimation based on Kendall				
Clayton	$\widehat{\theta} = 0.6831$	0.2024	0.2865	1.0797
Gumbel	$\widehat{\theta} = 1.3415$	0.1012	1.1432	1.5398
Gaussian	$\widehat{\rho} = 0.3893$	0.0813	0.2299	0.5488
Student's t	$\widehat{\rho} = 0.3893$ $\widehat{\nu} = 4$ (fixed)	0.0813	0.2299	0.5488

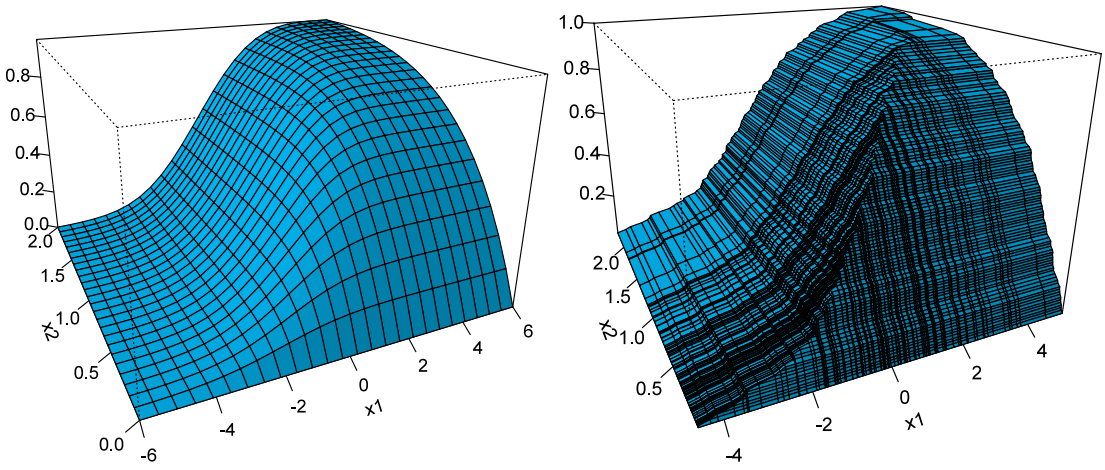
Table 2.4: Semi-parametric estimation of an unknown parameter of different copulas. The first part of the table refers to the estimation based on ML, and the second part is based on Kendall's tau.

Copula	AIC	BIC
Clayton	-41.4908	-38.4802
Gumbel	-16.8266	-13.8160
Gaussian	-29.3603	-26.3497
Student's t	-27.3593	-21.3381

Table 2.5: Comparison of AIC and BIC of semi-parametric estimation based on ML.

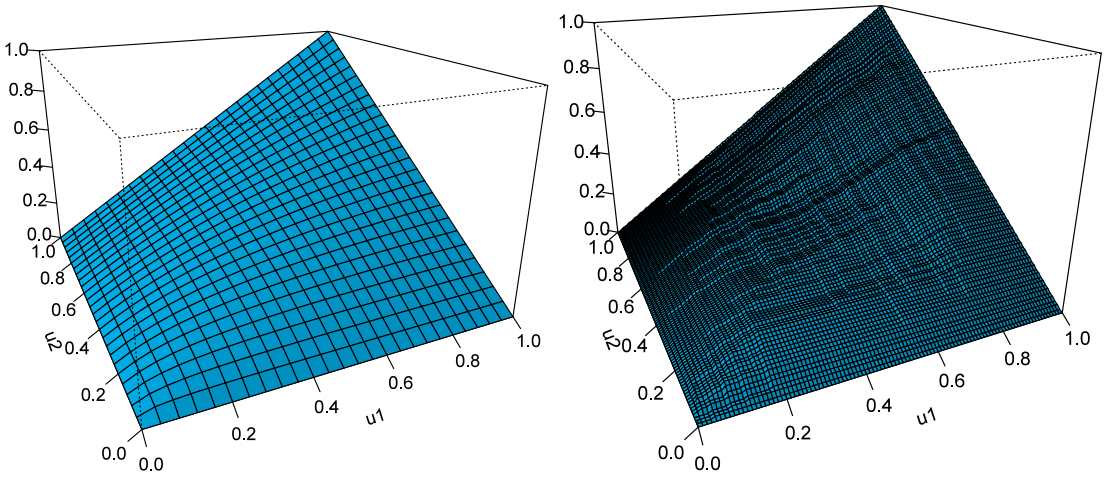
Copula	Parameter	Fully par. MLE	Semi-par. MLE	Semi-par. Kendall
Clayton	$\widehat{\theta}$	0.9139	0.8358	0.6831

Table 2.6: Comparison of all estimated parameters for Clayton copula.



(a) True distribution (b) Empirical distribution

Figure 2.8: Comparison of the true and empirical bivariate distribution function.



(a) True copula (b) Empirical copula

Figure 2.9: Comparison of the true and empirical copula distribution functions.

3. Models for multivariate time series based on copulas

This section connects concepts presented in the previous two chapters to perform models for multivariate time series based on copulas. The model, its estimation and prediction are presented. Three different algorithms for the prediction intervals are proposed. This chapter is mainly inspired by Patton [2012] and Chen and Fan [2006].

3.1 Model specification

A copula multivariate time series model is defined similarly to Patton [2012]. Let us assume we have a multivariate time series $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$ fulfilling Definition 4 and assume that each univariate time series $\{Y_{d,t},\, t \in \mathbb{Z}\}$, $d = 1, \ldots, D$ can be written in the form of (1.1) as follows

$$Y_{d,t} = \mu_d(\mathcal{F}_{t-1}) + \sigma_d(\mathcal{F}_{t-1})\, \varepsilon_{d,t}, \quad t \in \mathbb{Z}, \tag{3.1}$$

where $\mathcal{F}_{t-1}$ is the smallest σ-algebra generated by all past values of multivariate time series $\boldsymbol{Y}_{t-1}$, $\boldsymbol{Y}_{t-2}$, ..., mathematically $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \ldots\}$.

Furthermore, $\varepsilon_{d,t}$ are innovations satisfying the following conditions

$$\begin{aligned} \mathsf{E}\,(\varepsilon_{d,t} \mid \mathcal{F}_{t-1}) &= 0, \quad d = 1, \ldots, D, \\ \operatorname{var}(\varepsilon_{d,t} \mid \mathcal{F}_{t-1}) &= 1, \quad d = 1, \ldots, D. \end{aligned}$$

By the law of total expectation and total variance, the conditional mean and variance of $\varepsilon_{d,t}$ written above also hold unconditionally. For $d = 1, \ldots, D$, it holds that

$$\begin{aligned} \mathsf{E}\,(\varepsilon_{d,t}) &= \mathsf{E}\left(\mathsf{E}\,(\varepsilon_{d,t} \mid \mathcal{F}_{t-1})\right) = \mathsf{E}\,(0) = 0, \\ \operatorname{var}(\varepsilon_{d,t}) &= \mathsf{E}\left(\operatorname{var}(\varepsilon_{d,t} \mid \mathcal{F}_{t-1})\right) + \operatorname{var}\left(\mathsf{E}\,(\varepsilon_{d,t} \mid \mathcal{F}_{t-1})\right) = \mathsf{E}\,(1) + \operatorname{var}(0) = 1. \end{aligned}$$

The vector of innovations is denoted as $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{D,t})^\top$ and it is assumed to be continuously distributed. The cumulative distribution functions of innovations are denoted as follows

$$\begin{aligned} \varepsilon_{d,t} \mid \mathcal{F}_{t-1} &\sim F_{d,t}, \quad t \in \mathbb{Z},\, d = 1, \ldots, D, \\ \boldsymbol{\varepsilon}_t \mid \mathcal{F}_{t-1} &\sim F_t, \quad t \in \mathbb{Z}, \end{aligned}$$

where $F_{d,t}$, $d = 1, \ldots, D$ are univariate cumulative distribution functions with zero mean and unit variance and F_t is the corresponding D-dimensional cumulative distribution function.

For the sake of completeness of the thesis, we state here Sklar's theorem for conditional random variables. For example, this theorem can be found in Patton [2001, Theorem 3].

Theorem 8. *(Sklar's theorem for conditional distribution) Let F be a conditional joint distribution function defined on $\mathbb{R}^D$ with conditional marginal distribution functions $F_1, \ldots, F_D$ and let us denote $\mathcal{F}$ conditioning set. Then there exists a conditional copula $C : [0, 1]^D \to [0, 1]$ such that for all $(x_1, \ldots, x_D)^\top \in \mathbb{R}^D$ it holds*

$$F(x_1, \ldots, x_D \mid \mathcal{F}) = C\Big(F_1(x_1 \mid \mathcal{F}), \ldots, F_D(x_D \mid \mathcal{F})\Big|\mathcal{F}\Big). \tag{3.2}$$

Furthermore, if $F_1, \ldots, F_D$ are additionally absolutely continuous distribution functions, the conditional copula C is unique. Otherwise, C is uniquely defined only on $Ran(F_1) \times \ldots \times Ran(F_D)$ where $Ran(F_d)$, $d = 1, \ldots, D$ denotes the range of the function F_d.

Conversely, having a conditional copula $C : [0, 1]^D \to [0, 1]$ and marginal conditional distribution functions $F_1, \ldots, F_D$, the function F defined by equation (3.2) is a conditional joint D-dimensional distribution function with marginals $F_1, \ldots, F_D$.

Proof. For the proof of conditional Sklar's theorem, see an appendix in Patton [2001].

□

For simplicity we assume that for each $d = 1, \ldots, D$ innovations $\varepsilon_{d,t}$ are *i.i.d.* and independent of $\mathcal{F}_{t-1}$ for each $t \in \mathbb{Z}$. Therefore also $\boldsymbol{\varepsilon}_t$ are *i.i.d.* random vectors. Further, it is denoted

$$\begin{aligned} \varepsilon_{d,t} &\sim F_d, \quad t \in \mathbb{Z}, d = 1, \ldots, D, \\ \boldsymbol{\varepsilon}_t &\sim F, \quad t \in \mathbb{Z}. \end{aligned}$$

From Sklar's theorem, see Theorem 1, there exists a unique copula C such that for all $(x_1, \ldots, x_D)^\top$ it holds

$$F(x_1, \ldots, x_D) = C\Big(F_1(x_1), \ldots, F_D(x_D)\Big). \tag{3.3}$$

In summary, univariate time series are linked to multivariate series through innovations.

A stationary time series is assumed for the estimation of the model. The estimation can be divided into three parts which need to be estimated. These are the estimation of the univariate time series, for example, ARMA-GARCH, the marginal distributions of innovations and the copula. It is possible to estimate all parts parametrically and non-parametrically, so it is possible to create an estimate using any combination.

It is possible to build a fully parametric model where all three parts of the model are parameterised and estimated using the maximum likelihood method. The disadvantage of this approach lies in the large number of parameters to be estimated if we estimate all the parameters simultaneously. However, we also need to identify the model class correctly, i.e. correctly identify the type of each univariate time series, correctly identify the family of the distribution of the innovations and correctly select the copula family. We will not deal with this method, it can be found in the article Patton [2006] or in Patton [2013, Section 3.1] where a multi-stage procedure is shown.

Another way to estimate the time series is the semi-parametric model. It differs from the fully parametric estimator only in the non-parametric estimation of the innovations. The marginal distribution function of innovations is usually estimated using the empirical distribution function of residuals. The copula parameter can be estimated using estimators based on both maximum likelihood and Kendall's tau. This estimation is most often done in a multi-stage procedure.

The last option is the fully non-parametric approach. The non-parametric estimation was discussed by Scaillet and Fermanian [2002], and we will not discuss this approach further.

In the next section of the thesis, we describe the SCOMDY model, which represents a semi-parametric option.

3.2 SCOMDY

The SCOMDY model is a semi-parametric approach to estimating pre-specified multivariate time series $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$. This model was created by Chen and Fan [2006]. SCOMDY is an abbreviation for Semi-parametric Copula-Based Multivariate Dynamic Models.

Each univariate time series is estimated using a parametric model such as the ARMA-GARCH model. The cumulative distribution functions of innovations are then estimated non-parametrically from univariate residuals. Finally, a parametric estimation of the copula is performed. The estimation process is described in the next three sub-chapters. For simplicity, we again assume a two-dimensional time series ($D = 2$), but the extension to a higher dimension is straightforward.

Before starting the estimations, we parameterise the time series defined in equation (3.1). By $\boldsymbol{\psi}_d$, we denote the vector of unknown parameters for d-th time series. The time series is then written same as (1.9)

$$Y_{d,t} = \mu_d(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d) + \sigma_d(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d)\, \varepsilon_{d,t}, \quad t \in \mathbb{Z},\, d = 1,\, 2. \tag{3.4}$$

It is assumed that $C \in \mathcal{C} = \{C_\theta : \theta \in \Theta\}$, where Θ is a set of all possible values for copula parameter. The copula C in equation (3.3) is then denoted as C_θ. It is further assumed that there is a density c_θ of the copula C_θ.

For the estimation procedure and illustration, it is assumed to have a bivariate time series $\{\boldsymbol{Y}_1,\, \ldots,\, \boldsymbol{Y}_n\}$, where $\boldsymbol{Y}_t = (Y_{1,t},\, Y_{2,t})^\top$ for $t = 1,\, \ldots,\, n$.

To show that the model can be estimated in three steps, we introduce the log-likelihood of the SCOMDY model.

Theorem 9. *The log-likelihood function of a SCOMDY model of a bivariate time series $\{\boldsymbol{Y}_1,\, \ldots,\, \boldsymbol{Y}_n\}$ connected with a copula C_θ with density c_θ is*

$$\ell(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2,\, \theta) = \sum_{t=1}^{n} \log f_{\boldsymbol{Y}_t}(y_1,\, y_2) = \\ \sum_{t=1}^{n} \Big\{ \log c_\theta\Big(F_1(\varepsilon_{1,t}),\, F_2(\varepsilon_{2,t})\Big) + \sum_{d=1}^{2} \log\Big(f_d(\varepsilon_{d,t})\Big) - \sum_{d=1}^{2} \log\Big(\sigma_d(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d)\Big) \Big\}, \tag{3.5}$$

where $f_{\boldsymbol{Y}_t}(y_1,\, y_2)$ is density of $\boldsymbol{Y}_t \mid \mathcal{F}_{t-1}$.

Proof. For deriving the log-likelihood function, we use the theorem of transformation of random vectors, see Fessler [1998]. It is assumed

$$\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t},\, \varepsilon_{2,t})^\top \sim C_\theta\Big(F_1(x_1),\, F_2(x_2)\Big).$$

Thus the random vector of innovations $(\varepsilon_{1,t},\, \varepsilon_{2,t})^\top$ has the density expressed in equation (2.6) in Theorem 6

$$f_{\boldsymbol{\varepsilon}_t}(x_1,\, x_2) = c_\theta\Big(F_1(x_1),\, F_2(x_2)\Big) \prod_{d=1}^{2} f_d(x_d).$$

The aim is to derive the density of $\boldsymbol{Y}_t \mid \mathcal{F}_{t-1}$, since the log-likelihood function is defined as

$$\ell(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2,\, \theta) = \sum_{t=1}^{n} \log f_{\boldsymbol{Y}_t}(y_1,\, y_2).$$

The bivariate time series can be written in a matrix form. For $t = 1, \ldots, n$ it holds

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1) \\ \mu_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2) \end{pmatrix} + \begin{pmatrix} \sigma_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1) & 0 \\ 0 & \sigma_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2) \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}.$$

Let us denote transformation function $g : \mathbb{R}^2 \to \mathbb{R}^2$

$$g(x_1,\, x_2) = \begin{pmatrix} \mu_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1) \\ \mu_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2) \end{pmatrix} + \begin{pmatrix} \sigma_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1) & 0 \\ 0 & \sigma_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Transformation g fulfils all the assumptions related to the theorem of transformation. The inverse function of g follows

$$\begin{aligned} g^{-1}(y_1,\, y_2) &= \begin{pmatrix} \sigma_1^{-1}(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1) & 0 \\ 0 & \sigma_2^{-1}(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2) \end{pmatrix} \begin{pmatrix} y_1 - \mu_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1) \\ y_2 - \mu_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2) \end{pmatrix} \\ &= \left(\frac{y_1 - \mu_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1)}{\sigma_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1)},\, \frac{y_2 - \mu_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2)}{\sigma_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2)} \right). \end{aligned}$$

The last building block we need is the Jacobian matrix of the transformation g^{-1}

$$J = \begin{pmatrix} \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_2} \\ \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_2} \end{pmatrix} = \begin{pmatrix} \sigma_1^{-1}(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1) & 0 \\ 0 & \sigma_2^{-1}(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2) \end{pmatrix}.$$

Finally, the density of $\boldsymbol{Y}_t \mid \mathcal{F}_{t-1}$ is due to the theorem of transformation

$$\begin{aligned} f_{\boldsymbol{Y}_t}(y_1,\, y_2) &= f_{\boldsymbol{\varepsilon}_t}\Big(g^{-1}(y_1,\, y_2)\Big) \cdot \Big|\det(J)\Big| \\ &= c_\theta\left\{ F_1\left(\frac{y_1 - \mu_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1)}{\sigma_1(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_1)} \right),\, F_2\left(\frac{y_2 - \mu_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2)}{\sigma_2(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_2)} \right) \right\} \\ &\quad \cdot \prod_{d=1}^{2} f_d\left(\frac{y_1 - \mu_d(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d)}{\sigma_d^{-1}(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d)} \right) \cdot \left| \prod_{d=1}^{2} \sigma_d^{-1}(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d) \right| \\ &= c_\theta\Big\{ F_1\Big(\varepsilon_{1,t}\Big),\, F_2\Big(\varepsilon_{2,t}\Big) \Big\} \cdot \prod_{d=1}^{2} f_d(\varepsilon_{d,t}) \cdot \prod_{d=1}^{2} \sigma_d^{-1}(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d). \end{aligned}$$

The log-likelihood for the SCOMDY model is then

$$\ell(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \theta) = \sum_{t=1}^{n} \log f_{\boldsymbol{Y}_t}(y_1, y_2) = \\ \sum_{t=1}^{n} \Big\{ \log c_\theta\Big(F_1(\varepsilon_{1,t}), F_2(\varepsilon_{2,t})\Big) + \sum_{d=1}^{2} \log\Big(f_d(\varepsilon_{d,t})\Big) - \sum_{d=1}^{2} \log\Big(\sigma_d(\mathcal{F}_{t-1}, \boldsymbol{\psi}_d)\Big) \Big\}.$$

□

The unknown parameters $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_2$ and θ are found as an argument maxima of a log-likelihood function from Theorem 9, in equation (3.5).

$$\Big(\widehat{\boldsymbol{\psi}}_1, \widehat{\boldsymbol{\psi}}_2, \widehat{\theta}\Big)^\top = \underset{\boldsymbol{\psi}_1 \in \boldsymbol{\Psi}_1, \boldsymbol{\psi}_2 \in \boldsymbol{\Psi}_2, \theta \in \Theta}{\arg\max} \ell(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \theta),$$

where $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ are spaces of all possible values of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, respectively. The subspace of copula parameter θ is Θ. The initial values for computation of $\ell(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \theta)$ can be set to zero. The initial values can also be selected as more sophisticated.

It can be seen that the log-likelihood (3.5) can be divided into three sums that can be maximised separately. The parameters $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ are hidden in $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$, respectively and in the conditional variance $\sigma_1(\mathcal{F}_{t-1}, \boldsymbol{\psi}_1)$ and $\sigma_2(\mathcal{F}_{t-1}, \boldsymbol{\psi}_2)$, respectively. The copula parameter θ is only in the first term of (3.5). Therefore, we can split the estimation into three steps instead of optimising in one step, which is computationally very demanding.

From the last term of (3.5) we firstly estimate unknown parameters of time series $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$. From the middle term of (3.5) estimate a distribution of innovations given already estimated parameters $\widehat{\boldsymbol{\psi}}_1$ and $\widehat{\boldsymbol{\psi}}_2$ from the first step. Finally, estimate θ from the first term of (3.5) using the previously estimated parts.

Remark. The second and last steps of the estimation procedure refer to Section 2.5.2, where the cumulative distribution functions of innovations are estimated non-parametrically, and the copula parameter is estimated parametrically.

3.2.1 Parametric estimation of time series

The first part of the estimation focuses on individual time series from which is multivariate series composed. For this part, we have introduced ARMA-GARCH models in Chapter 1, which are suitable for modelling financial time series. We assume each of the univariate series to have a model in the following form

$$Y_{d,t} = \mu_d(\mathcal{F}_{t-1}, \boldsymbol{\psi}_d) + \sigma_d(\mathcal{F}_{t-1}, \boldsymbol{\psi}_d)\, \varepsilon_{d,t}, \quad t = 1, \ldots, n,\ d = 1, 2.$$

The estimator of $\boldsymbol{\psi}_d$, $d = 1, 2$ is denoted as $\widehat{\boldsymbol{\psi}}_d$ and is derived in Section 1.3. Recall that under regularity conditions is the estimator consistent, and if $\mathsf{E}\,(\varepsilon_t^4)$ is finite is the estimator asymptotically normal, see Francq and Zakoian [2019, part 7.2].

In general, limiting yourself only to ARMA-GARCH models is not necessary. Chen and Fan [2003] discussed VAR, Markov switching models, etc.

3.2.2 Non-parametric estimation of the distribution function of innovations

The second part deals with the estimation of the distribution of innovations. It is estimated non-parametrically from residuals defined as

$$\widehat{\varepsilon}_{d,t} = \frac{Y_{d,t} - \mu_d(\mathcal{F}_{t-1},\, \widehat{\boldsymbol{\psi}}_d)}{\sigma_d(\mathcal{F}_{t-1},\, \widehat{\boldsymbol{\psi}}_d)}, \quad t = 1,\, \ldots,\, n,\, d = 1,\, 2. \tag{3.6}$$

The empirical distribution function of these residuals is used as a non-parametric estimation of the distribution of innovations

$$\widehat{F}_d(s) = \frac{1}{n+1} \sum_{t=1}^{n} \mathbb{1}\{\widehat{\varepsilon}_{d,t} \leq s\}, \quad s \in \mathbb{R},\, d = 1,\, 2.$$

3.2.3 Parametric estimation of copula

The estimation procedure finishes with the estimation of the copula parameter θ. We already estimated time series parameters and marginal distributions of innovations. So, it is possible to create pseudo-sample $\widehat{\boldsymbol{U}}_1,\, \ldots,\, \widehat{\boldsymbol{U}}_n$, where

$$\widehat{\boldsymbol{U}}_t = \Big(\widehat{F}_1(\widehat{\varepsilon}_{1,t}),\, \widehat{F}_2(\widehat{\varepsilon}_{2,t})\Big)^\top, \quad t = 1,\, \ldots,\, n. \tag{3.7}$$

The copula parameter, denoted as θ, is estimated by maximising the likelihood

$$\mathcal{L}(\theta) = \prod_{t=1}^{n} c_\theta(\widehat{\boldsymbol{U}}_t),$$

where c_θ is the density of C_θ from (2.5).

Same as in (2.9) we can derive log-likelihood

$$\ell(\theta) = \log \mathcal{L}(\theta) = \sum_{t=1}^{n} \log\Big(c_\theta(\widehat{\boldsymbol{U}}_t)\Big). \tag{3.8}$$

Finally, the estimator is found as an argument of the maxima

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \ell(\theta),$$

where Θ is the set of all possible values of the copula parameter θ.

The unknown copula parameter θ can also be estimated using Kendall's tau in the same way as in Section 2.5.2 estimation based on Kendall's tau.

The resulting estimator $\widehat{\theta}$ based on maximum likelihood is consistent and asymptotically normal under suitable regularity conditions, for more information, see Choroś et al. [2010]. The alternative estimator based on Kendall's tau is also consistent and asymptotically normal, see theorem 7.

3.3 Goodness-of-fit test

The goodness-of-fit test helps to determine whether the data can be modelled by a particular family of copulas $\mathcal{C}$, or whether another family should be chosen. The

goodness-of-fit test is described in detail in Rémillard [2017]. This test examines whether the estimated copula C was selected from a suitable copula family $\mathcal{C}$. The null hypothesis and alternative follow

$$H_0 : C \in \mathcal{C}, \qquad H_1 : C \notin \mathcal{C}.$$

Remember $\mathcal{C} = \{C_\theta : \theta \in \Theta\}$ is some parametric family of copula C and Θ is a set of all possible values for the copula parameter θ.

To make a test statistics, denote $C_{\widehat{\theta}}$ estimated parametric copula and $\widehat{C}$ denotes the empirical copula defined in (2.12). To test the hypothesis H_0, it is defined so-called Cramér-von Mises test statistics

$$S_n = \int_{[0,1]^2} n\Big(\widehat{C}(\boldsymbol{u}) - C_{\widehat{\theta}}(\boldsymbol{u})\Big)^2 d\widehat{C}(\boldsymbol{u}) = \sum_{t=1}^{n}\Big(\widehat{C}\Big(\widehat{\boldsymbol{U}}_t\Big) - C_{\widehat{\theta}}\Big(\widehat{\boldsymbol{U}}_t\Big)\Big)^2, \tag{3.9}$$

where $\boldsymbol{u} = (u_1,\, u_2)^\top \in [0,1]^2$ and $\widehat{\boldsymbol{U}}_t$ are the pseudo-observations defined in (3.7). It is easy to see that if the parametric family is correctly determined, then the test statistic S_n would be low, so large values of S_n indicate a violation of the null hypothesis. To derive the p-value of the test, we need to know the distribution of S_n, assuming that H_0 holds. However, this is difficult to derive, so the p-value is obtained from the bootstrap. The following algorithm is proposed in Rémillard [2017] and follows

1. Estimate the time series $\{Y_{d,1},\, \ldots,\, Y_{d,n}\}$, $d = 1,\, 2$, see Section 3.2.1. Compute residuals from equation (3.6) and find the empirical cumulative distribution function of these residuals, see Section 3.2.2. Compute pseudo-sample $\widehat{\boldsymbol{U}}_1,\, \ldots,\, \widehat{\boldsymbol{U}}_n$ from equation (3.7).

2. From computed pseudo-sample $\widehat{\boldsymbol{U}}_1,\, \ldots,\, \widehat{\boldsymbol{U}}_n$ compute the empirical copula $\widehat{C}$ defined in (2.12). Estimate the unknown copula parameter θ using this pseudo-sample above to find $C_{\widehat{\theta}}$. It can be estimated using the log-likelihood defined in (3.8).

3. Compute the test statistic S_n defined above in equation (3.9).

4. Choose an integer B sufficiently large. Repeat the following steps for each $b \in \{1,\, \ldots,\, B\}$.

   (a) Generate a random sample $\boldsymbol{U}_1^{(b)},\, \ldots,\, \boldsymbol{U}_n^{(b)}$ where $\boldsymbol{U}_t^{(b)} = \Big(U_{1,t}^{(b)},\, U_{2,t}^{(b)}\Big)^\top$, $t = 1,\, \ldots,\, n$ from estimated copula $C_{\widehat{\theta}}$.

   (b) Compute pseudo-sample $\widehat{\boldsymbol{U}}_1^{(b)},\, \ldots,\, \widehat{\boldsymbol{U}}_n^{(b)}$ defined as

   $$\widehat{\boldsymbol{U}}_t^{(b)} = \Big(\widehat{F}_1\Big(U_{1,t}^{(b)}\Big),\, \widehat{F}_2\Big(U_{2,t}^{(b)}\Big)\Big)^\top, \quad t = 1,\, \ldots,\, n,$$

   where $\widehat{F}_d$ is an empirical cumulative distribution function created from random sample $U_{d,1}^{(k)},\, \ldots,\, U_{d,n}^{(k)}$, $d = 1,\, 2$ and is scaled by $n + 1$.

   (c) Compute empirical copula from the pseudo-sample $\widehat{\boldsymbol{U}}_1^{(b)},\, \ldots,\, \widehat{\boldsymbol{U}}_n^{(b)}$

   $$\widehat{C}^{(b)}(\boldsymbol{u}) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{\widehat{\boldsymbol{U}}_t^{(k)} \le \boldsymbol{u}\}, \quad \boldsymbol{u} \in [0,1]^2,$$

   and estimate copula parameter θ using this generated pseudo-sample. The estimation of copula is denoted $C_{\widehat{\theta}}^{(b)}$.

(d) Compute $S_n^{(b)}$

$$S_n^{(b)} = \sum_{t=1}^{n} \left(\widehat{C}^{(b)}\left(\widehat{\boldsymbol{U}}_t^{(b)}\right) - C_{\widehat{\theta}}^{(b)}\left(\widehat{\boldsymbol{U}}_t^{(b)}\right) \right)^2.$$

Approximated p-value is then computed from the $S_n^{(1)}, \ldots, S_n^{(B)}$

$$\widehat{p} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{S_n^{(b)} > S_n\}.$$

The package "copula", see Hofert et al. [2022] in statistical software R uses the adjusted approximation for p-value

$$\widehat{p} = \frac{1}{B+1} \left(\sum_{b=1}^{B} \mathbb{1}\{S_n^{(b)} \geq S_n\} + 0.5 \right).$$

More detailed information about the algorithm and its technical assumptions can be found in the article written by Rémillard [2017]. There, one can find information that it is not necessary to replicate the whole time series. It is enough to replicate the $\boldsymbol{U}_1^{(b)}, \ldots, \boldsymbol{U}_n^{(b)}$ from the copula $C_{\widehat{\theta}}$.

3.4 Predictions

We most often estimate the time series to determine their future evolution. This section shows three different algorithms to predict multivariate time series based on copulas. All three algorithms are based on the bootstrap technique, allowing us to find $1 - \alpha$ confidence intervals, $\alpha \in (0, 1)$, usually $\alpha = 0.05$.

It is assumed a multivariate time series $\{\boldsymbol{Y}_t, t \in \mathbb{Z}\}$. For simplicity, we assume a bivariate time series and denote it as $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$, where $\boldsymbol{Y}_t = (Y_{1,t}, Y_{2,t})^\top$ for $t = 1, \ldots, n$. The methods can be easily extended to higher dimensions. We also consider the same notation for all parameters as in the section on SCOMDY models, see Section 3.2. We look for the K-step ahead prediction for $\boldsymbol{Y}_{n+K}$, $K \in \mathbb{N}$ conditioned on the time series up to time n. The predicted value is denoted as $\widehat{\boldsymbol{Y}}_{n+K}(n) = \left(\widehat{Y}_{1,n+K}(n), \widehat{Y}_{2,n+K}(n)\right)^\top$, $K \in \mathbb{N}$.

3.4.1 First algorithm

The first described algorithm is simple and intuitive. However, its disadvantage is that it does not consider the parameter estimation's variability, i.e., the estimation is random. Thus, this approximate method works if a sample size of time series n is sufficiently large, i.e. the variability of the estimate will already be small. The algorithm follows.

1. Estimate the time series $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$ to obtain $\widehat{\boldsymbol{\psi}}_d$, $d = 1, 2$. It can be estimated using a parametric approach, see Section 3.2.1.

2. Compute residuals $\widehat{\varepsilon}_{d,1}, \ldots, \widehat{\varepsilon}_{d,n}$, $d = 1, 2$ using an equation (3.6).

3. Estimate copula parameter from residuals $\widehat{\varepsilon}_{d,1}, \ldots, \widehat{\varepsilon}_{d,n}$, $d = 1, 2$, see Section 3.2.3 and denote $C_{\widehat{\theta}}$ estimated parametric copula.

4. Due to the assumption of zero expectation and unit variance of innovations, standardise the residuals

$$\widetilde{\varepsilon}_{d,t} = \frac{\widehat{\varepsilon}_{d,t} - \frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_{d,t}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\Big(\widehat{\varepsilon}_{d,t} - \frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_{d,t}\Big)^2}}, \quad t = 1, \ldots, n,\ d = 1, 2.$$

5. Choose sufficiently large integer B of bootstraps. Repeat the following steps for each $b \in \{1, \ldots, B\}$.

   (a) For each prediction time $k \in \{1, \ldots, K\}$ repeat the following steps.

      i. Generate pseudo-observation $\boldsymbol{U}^{(b,k)} = \Big(U_1^{(b,k)}, U_2^{(b,k)}\Big)^\top$ from estimated copula $C_{\widehat{\theta}}$.

      ii. Compute simulated innovation $\boldsymbol{\varepsilon}^{(b,k)} = \Big(\varepsilon_1^{(b,k)}, \varepsilon_2^{(b,k)}\Big)^\top$, where it holds

$$\varepsilon_d^{(b,k)} = \widetilde{F}_d^{-1}\Big(U_d^{(b,k)}\Big), \quad d = 1, 2,$$

      where $\widetilde{F}_d^{-1}(s)$ is a quantile function of an empirical distribution function $\widetilde{F}_d(s)$ defined as

$$\widetilde{F}_d(s) = \frac{1}{n+1}\sum_{i=1}^{n}\mathbb{1}\{\widetilde{\varepsilon}_{d,t} \le s\}, \quad s \in \mathbb{R},\ d = 1, 2.$$

      iii. Generate the possible future value of the series using the estimated parameters $\widehat{\boldsymbol{\psi}}_d$, $d = 1, 2$ and simulated innovation $\varepsilon_d^{(b,k)}$. In the case of the ARMA-GARCH model, generate the new future value from the prescription (1.2) using the sufficient number of last observed values of the original series or use already generated values. In short, we write it as follows and use this notation in the following algorithms

$$Y_{d,n+k}^{(b,k)}(n) = \mu_d(\mathcal{F}_n, \widehat{\boldsymbol{\psi}}_d) + \sigma_d(\mathcal{F}_n, \widehat{\boldsymbol{\psi}}_d)\,\varepsilon_d^{(b,k)}, \quad d = 1, 2,$$

      where $\mathcal{F}_n$ is the full information of the original time series up to time n.

6. For each time $k \in \{1, \ldots, K\}$ and each series $d = 1, 2$ compute median of bootstrapped data $Y_{d,n+k}^{(b,k)}(n)$, $b = 1, \ldots, B$ to obtain predicted value $\widehat{Y}_{d,n+k}(n)$ and compute $\alpha/2$ and $1-\alpha/2$ quantiles to obtain $1-\alpha$ confidence interval of prediction. The predicted value $\widehat{Y}_{d,n+k}(n)$ can also be obtained analytically using the procedure in Section 1.5.

3.4.2 Second algorithm

We describe the second algorithm based on the article by Pascual et al. [2004]. This algorithm takes into account the variability of parameter estimation of time series. The algorithm is based on generating new time series, which are subsequently estimated. These new estimates are then used to generate predictions from the original series.

1. Steps 1. – 4. are same as in the first algorithm, see Section 3.4.1.

2. Choose sufficiently large integer B of bootstraps to be performed. Repeat the following steps for each $b \in \{1, \ldots, B\}$.

   (a) For each $t = 1, \ldots, n$ generate pseudo sample $\boldsymbol{U}_t^{(b)} = \left(U_{1,t}^{(b)}, U_{2,t}^{(b)}\right)^\top$ from estimated copula $C_{\widehat{\theta}}$.

   (b) Compute simulated innovation $\boldsymbol{\varepsilon}_t^{(b)} = \left(\varepsilon_{1,t}^{(b)}, \varepsilon_{2,t}^{(b)}\right)^\top$, where it holds

   $$\varepsilon_{d,t}^{(b)} = \widetilde{F}_d^{-1}\left(U_{d,t}^{(b)}\right), \quad t = 1, \ldots, n,\ d = 1, 2,$$

   where $\widetilde{F}_d^{-1}(s)$ is a quantile function of an empirical distribution function $\widetilde{F}_i(s)$ defined as

   $$\widetilde{F}_d(s) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}\{\widetilde{\varepsilon}_{d,t} \leq s\}, \quad s \in \mathbb{R},\ d = 1, 2.$$

   (c) Construct a bootstrap series $\{\boldsymbol{Y}_1^{(b)}, \ldots, \boldsymbol{Y}_n^{(b)}\}$, $\boldsymbol{Y}_t^{(b)} = \left(Y_{1,t}^{(b)}, Y_{2,t}^{(b)}\right)^\top$, $t = 1, \ldots, n$ where each univariate time series is constructed as follows

   $$Y_{d,t}^{(b)} = \mu_d(\mathcal{F}_{t-1}, \widehat{\boldsymbol{\psi}}_d) + \sigma_d(\mathcal{F}_{t-1}, \widehat{\boldsymbol{\psi}}_d)\, \varepsilon_{d,t}^{(b)}, \quad t = 1, \ldots, n,\ d = 1, 2.$$

   Use the first few observations to initialise the bootstrapped series from the original time series $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$.

   (d) Estimate bootstrap series $\{\boldsymbol{Y}_1^{(b)}, \ldots, \boldsymbol{Y}_n^{(b)}\}$ using the same model as in the first step to obtaining $\widehat{\boldsymbol{\psi}}_d^{(b)}, d = 1, 2$.

   (e) For each prediction time $k \in \{1, \ldots, K\}$ repeat the following steps.

      i. Generate pseudo-observation $\boldsymbol{U}^{(b,k)} = \left(U_1^{(b,k)}, U_2^{(b,k)}\right)^\top$ from estimated copula $C_{\widehat{\theta}}$.

      ii. Compute simulated innovation $\boldsymbol{\varepsilon}^{(b,k)} = \left(\varepsilon_1^{(b,k)}, \varepsilon_2^{(b,k)}\right)^\top$, where it holds

      $$\varepsilon_d^{(b,k)} = \widetilde{F}_d^{-1}\left(U_d^{(b,k)}\right), \quad d = 1, 2,$$

      where $\widetilde{F}_d^{-1}(s)$ is a quantile function of an empirical distribution function $\widetilde{F}_d(s)$ defined above.

      iii. Generate the possible future value of the series using the newly estimated parameters $\widehat{\boldsymbol{\psi}}_d^{(b)}$, $d = 1, 2$ and simulated innovation $\varepsilon_d^{(b,k)}$

      $$Y_{d,n+k}^{(b,k)}(n) = \mu_d(\mathcal{F}_n, \widehat{\boldsymbol{\psi}}_d^{(b)}) + \sigma_d(\mathcal{F}_n, \widehat{\boldsymbol{\psi}}_d^{(b)})\, \varepsilon_d^{(b,k)}, \quad d = 1, 2,$$

      where $\mathcal{F}_n$ is the full information of the original time series up to time n. For completeness, we use the original time series $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$ for predictions, not the bootstrapped time series $\{\boldsymbol{Y}_1^{(b)}, \ldots, \boldsymbol{Y}_n^{(b)}\}$.

3. For each time $k \in \{1, \ldots, K\}$ compute median of bootstrapped data $Y_{d,n+k}^{(b,k)}(n)$, $b = 1, \ldots, B$ to obtain predicted value $\widehat{Y}_{d,n+k}$ and compute $\alpha/2$ and $1 - \alpha/2$ quantiles to obtain $1 - \alpha$ confidence interval of prediction. The predicted value $\widehat{Y}_{d,n+k}(n)$ can also be obtained analytically using the procedure in Section 1.5.

3.4.3 Third algorithm

The last algorithm we show for predictions is the most theoretically correct but time-consuming. This algorithm is similar to the second one, except that the copula parameter is also estimated from bootstrapped time series. The newly estimated copula parameter is then used to simulate innovations, which we use for predictions similar to the second algorithm. This algorithm takes into account the variability of all estimates.

1. Steps $1. - 4.$ are same as in the first algorithm, see Section 3.4.1.

2. Choose sufficiently large integer B of bootstraps. Repeat the following steps for each $b \in \{1, \ldots, B\}$.

   (a) For each $t = 1, \ldots, n$ generate pseudo sample $\boldsymbol{U}_t^{(b)} = \left(U_{1,t}^{(b)}, U_{2,t}^{(b)}\right)^\top$ from estimated copula $C_{\widehat{\theta}}$.

   (b) Compute simulated innovation $\boldsymbol{\varepsilon}_t^{(b)} = \left(\varepsilon_{1,t}^{(b)}, \varepsilon_{2,t}^{(b)}\right)^\top$, where it holds

   $$\varepsilon_{d,t}^{(b)} = \widetilde{F}_d^{-1}\left(U_{d,t}^{(b)}\right), \quad d = 1, 2,$$

   where $\widetilde{F}_d^{-1}(s)$ is a quantile function of empirical distribution function $\widetilde{F}_d(s)$ defined as

   $$\widetilde{F}_d(s) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}\{\widetilde{\varepsilon}_{d,t} \leq s\}, \quad s \in \mathbb{R},\, d = 1, 2.$$

   (c) Construct a bootstrap series $\{\boldsymbol{Y}_1^{(b)}, \ldots, \boldsymbol{Y}_n^{(b)}\}$, $\boldsymbol{Y}_t^{(b)} = \left(Y_{1,t}^{(b)}, Y_{2,t}^{(b)}\right)^\top$, $t = 1, \ldots, n$ where each univariate time series is generated using the estimated parameters $\widehat{\boldsymbol{\psi}}_d$, $d = 1, 2$ and simulated innovations $\varepsilon_{d,t}^{(b)}$

   $$Y_{d,t}^{(b)} = \mu_d(\mathcal{F}_{t-1}, \widehat{\boldsymbol{\psi}}_d) + \sigma_d(\mathcal{F}_{t-1}, \widehat{\boldsymbol{\psi}}_d)\, \varepsilon_{d,t}^{(b)}, \quad t = 1, \ldots, n,\, d = 1, 2.$$

   Use the first few observations to initialise the bootstrapped series from the original time series $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$.

   (d) Estimate bootstrap series $\{\boldsymbol{Y}_1^{(b)}, \ldots, \boldsymbol{Y}_n^{(b)}\}$ using the same model as in the first step to obtaining $\widehat{\boldsymbol{\psi}}_d^{(b)}$, $d = 1, 2$.

   (e) From this newly estimated time series, compute residuals $\widehat{\varepsilon}_{d,1}^{(b)}, \ldots, \widehat{\varepsilon}_{d,n}^{(b)}$, $d = 1, 2$ using an equation (3.6).

   (f) Estimate a copula parameter using $\widehat{\varepsilon}_{d,1}^{(b)}, \ldots, \widehat{\varepsilon}_{d,n}^{(b)}$, $d = 1, 2$, see Section 3.2.3, and denote $C_{\widehat{\theta}}^{(b)}$ estimated parametric copula.

   (g) Due to the assumption of zero expectation and unit variance of innovations, standardise the residuals

   $$\widetilde{\varepsilon}_{d,t}^{(b)} = \frac{\widehat{\varepsilon}_{d,t}^{(b)} - \frac{1}{n}\sum_{t=1}^{n} \widehat{\varepsilon}_{d,t}^{(b)}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\widehat{\varepsilon}_{d,t}^{(b)} - \frac{1}{n}\sum_{t=1}^{n} \widehat{\varepsilon}_{d,t}^{(b)}\right)^2}}, \quad t = 1, \ldots, n,\, d = 1, 2.$$

   (h) For each prediction time $k \in \{1, \ldots, K\}$ repeat the following steps.

i. Generate pseudo-observation $\boldsymbol{U}^{(b,k)} = \left(U_1^{(b,k)}, U_2^{(b,k)}\right)^\top$ from estimated copula $C_{\widehat{\theta}}^{(b)}$.

ii. Compute simulated innovation $\boldsymbol{\varepsilon}^{(b,k)} = \left(\varepsilon_1^{(b,k)}, \varepsilon_2^{(b,k)}\right)^\top$, where it holds

$$\varepsilon_d^{(b,k)} = \left(\widetilde{F}_d^{(b)}\right)^{-1}\left(U_d^{(b,k)}\right), \quad d = 1, 2,$$

where $\left(\widetilde{F}_d^{(b)}\right)^{-1}(s)$ is a quantile function of an empirical distribution function $\widetilde{F}_d^{(b)}(s)$ defined as

$$\widetilde{F}_d^{(b)}(s) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbb{1}\{\widetilde{\varepsilon}_{d,t}^{(b)} \leq s\}, \quad s \in \mathbb{R},\, d = 1, 2.$$

iii. Generate the possible future value of the series using newly estimated parameters $\widehat{\boldsymbol{\psi}}_d^{(b)}$, $d = 1, 2$ and simulated innovation $\varepsilon_d^{(b,k)}$ created with the help of newly estimated copula $C_{\widehat{\theta}}^{(b)}$

$$Y_{d,n+k}^{(b,k)}(n) = \mu_d(\mathcal{F}_n, \widehat{\boldsymbol{\psi}}_d^{(b)}) + \sigma_d(\mathcal{F}_n, \widehat{\boldsymbol{\psi}}_d^{(b)})\, \varepsilon_d^{(b,k)}, \quad d = 1, 2,$$

where $\mathcal{F}_n$ is the full information of the original time series up to time n. For completeness, we use the original time series $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$ for predictions, not the bootstrapped time series $\{\boldsymbol{Y}_1^{(b)}, \ldots, \boldsymbol{Y}_n^{(b)}\}$.

3. For each time $k \in \{1, \ldots, K\}$ compute median of bootstrapped data $Y_{d,n+k}^{(b,k)}(n)$, $b = 1, \ldots, B$ to obtain predicted value $\widehat{Y}_{d,n+k}(n)$ and compute $\alpha/2$ and $1 - \alpha/2$ quantiles to obtain $1 - \alpha$ confidence interval of prediction. The predicted value $\widehat{Y}_{d,n+k}$ can also be obtained analytically using the procedure in Section 1.5.

To clarify, all three algorithms use an estimator based on the SCOMDY model in the first three steps, see Section 3.2.

All three algorithms are compared in a simulation study in Section 4.3.

Another prediction method is based on the backward representation, see Thombs and Schucany [1990]. The disadvantage of this method is the need to express the time series backwards. Since this method created B bootstrapped time series similar to algorithm 2 with initialisation as the last observations. For GARCH models, this method is therefore impossible since the GARCH model cannot be expressed in a backwards-looking way.

3.5 Example

In this example, the estimation procedure of the SCOMDY model is shown on a bivariate time series created by two autoregressions of order one whose innovations come from a bivariate distribution obtained by composing two Student's t-distributions using the Clayton copula.

Statistical calculations were performed using R software, see R Core Team [2020], specifically for copula work, we chose the copula package, created by Hofert et al. [2022]. We attach the script Example_of_SCOMDY.R to this example.

Each univariate time series in an autoregressive model of order 1. It is denoted as AR(1). Generally for $d = 1, 2$, it holds

$$Y_{d,t} = \mu_{0,d} + \phi_d Y_{d,t-1} + \varepsilon_{d,t}, \quad t \in \mathbb{Z}, \tag{3.10}$$

where $\mu_{0,d}$ and ϕ_d are parameters. Note that autoregressive model in (3.10) is of the form of (3.4) if we denote

$$\begin{aligned} \mu_d(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d) &= \mu_{0,d} + \phi_d Y_{d,t-1}, \\ \sigma_d(\mathcal{F}_{t-1},\, \boldsymbol{\psi}_d) &= 1. \end{aligned}$$

Let us generate bivariate data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ from bivariate stationary time series $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$, where $\boldsymbol{Y}_t = (Y_{1,t},\, Y_{2,t})^\top$, $t \in \mathbb{Z}$, where $Y_{1,t}$ and $Y_{2,t}$ are autoregressions of order one, defined in (3.10) with the following parameter values

$$\begin{aligned} \mu_{0,1} &= 0, \quad \phi_1 = 0.6, \\ \mu_{0,2} &= 0, \quad \phi_2 = 0.4. \end{aligned}$$

Thus it holds

$$\begin{aligned} Y_{1,t} &= 0.6\, Y_{1,t-1} + \varepsilon_{1,t}, \quad t \in \mathbb{Z}, \\ Y_{2,t} &= 0.4\, Y_{2,t-1} + \varepsilon_{2,t}, \quad t \in \mathbb{Z}. \end{aligned}$$

Further, we assume $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t},\, \varepsilon_{2,t})^\top$, $t \in \mathbb{Z}$ to be *i.i.d.* coming from the bivariate distribution generated by Clayton copula with parameter $\theta = 1$, which correspond to Kendall's tau equal to $1/3$, with Student's t marginals with five degrees of freedom. Thus it holds

$$\boldsymbol{\varepsilon}_t \sim C^{Cl}_{\theta=1}\Big(F_1(x_1),\, F_2(x_2)\Big), \quad x_1,\, x_2 \in \mathbb{R},\, t \in \mathbb{Z},$$

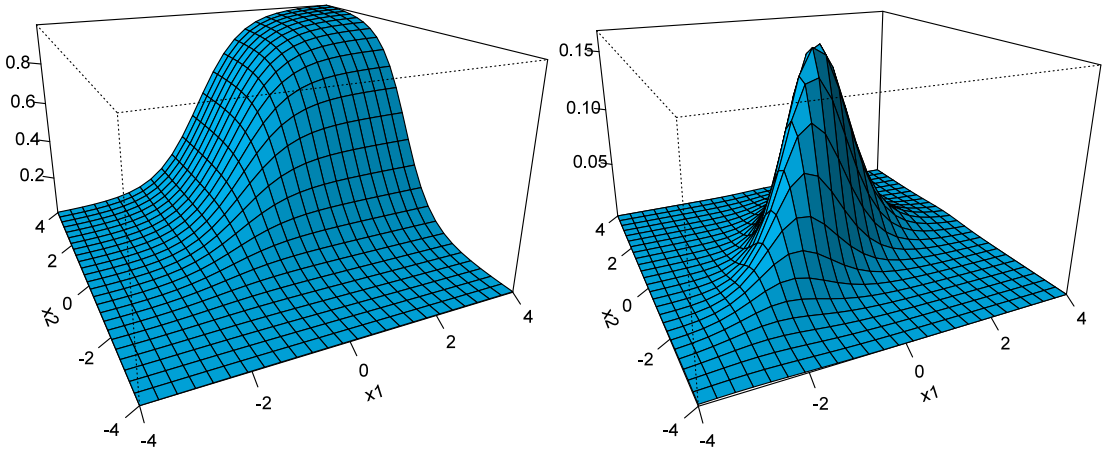
where $F_1(x_1)$ and $F_2(x_2)$ are cumulative distribution functions of Student's t-distribution with five degrees of freedom and $C^{Cl}_{\theta=1}(u_1,\, u_2)$ denote the Clayton copula with parameter $\theta = 1$, see equation (2.2). This distribution is shown in Figure 3.1.

Figure 3.2 plots the generated bivariate time series.

We estimate the simulated time series using the SCOMDY model. Firstly, we estimate the unknown parameters of the univariate time series ϕ_1 and ϕ_2. Secondly, we non-parametrically estimate the marginal distribution functions of innovations F_1 and F_2. In the last third step, we parametrically estimate the unknown parameter of the copula θ. Finally, predictions are made.

Parametric estimation of time series

The first step of the estimation is a parametric estimation of univariate time series. The estimation is based on the theory presented in Section 3.2.1. It is assumed to estimate the autoregressive processes of order one. Table 3.1 provide all estimated parameters. It can be seen that we are close to the true values, which are also included in the confidence intervals.



(a) Distribution function (b) Probability density function

Figure 3.1: Bivariate distribution generated by Clayton copula with Student's t marginals with 5 degrees of freedom.

Parameter	Estimation	Standard deviation	95 % confidence interval 2.5 %	97.5 %
$\widehat{\mu}_{0,1}$	-0.0233	0.1887	-0.3932	0.3466
$\widehat{\phi}_1$	0.5481	0.0588	0.4328	0.6635
$\widehat{\mu}_{0,2}$	0.2139	0.1367	-0.0540	0.4817
$\widehat{\phi}_2$	0.3152	0.0670	0.1838	0.4465

Table 3.1: Estimation of unknown parameters of both autoregressive models of order one.

Non-parametric estimation of the distribution function of innovations

The second step of the estimation computes residuals using equation (3.6). The distribution function of innovations is then estimated non-parametrically using an adjusted empirical cumulative distribution function. Again for more theory, visit Section 3.2.2.

Figure 3.3 provides histograms of residuals of both autoregressions. It is compared with Student's t-distribution with five degrees of freedom, which should be the true distribution of these innovations.

Parametric estimation of copula

The last step of the estimation procedure follows. It estimates the copula parameter. For theory, see Section 3.2.3. Figure 3.4 shows scatter plots with true contours.

Table 3.2 provides results of an estimation of copula parameter for different copulas with its 95% confidence intervals. The first part of the table contains an estimate based on the MLE, which can be compared with the second part, which includes an estimate based on Kendall's tau. To estimate the Student's t-copula

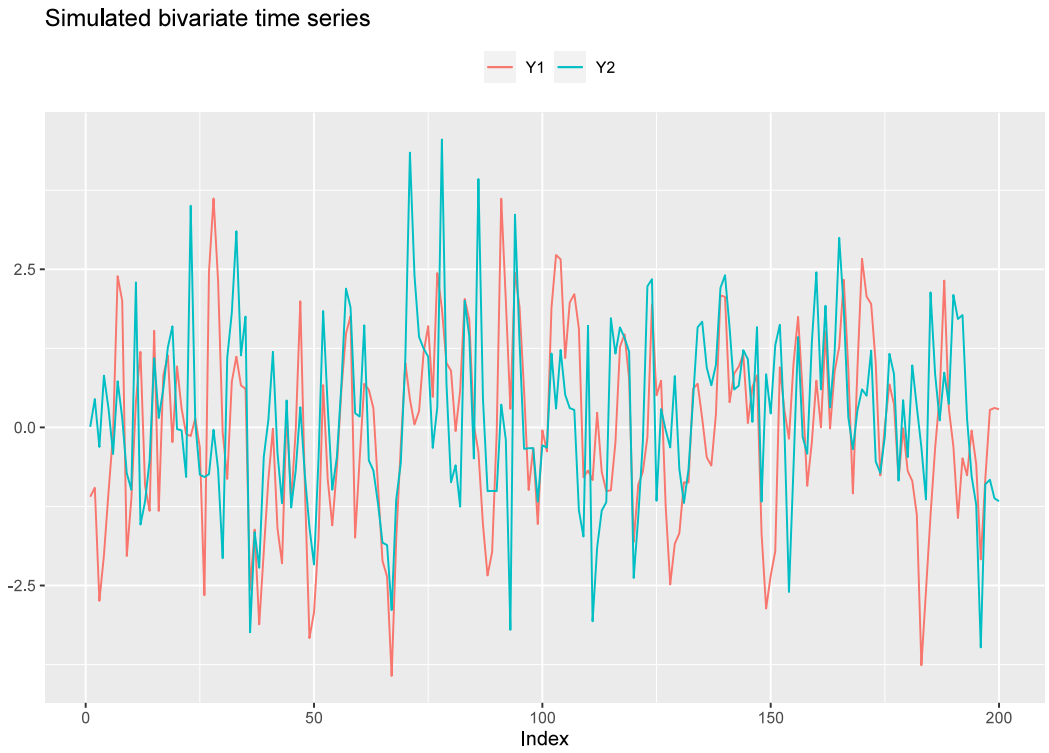


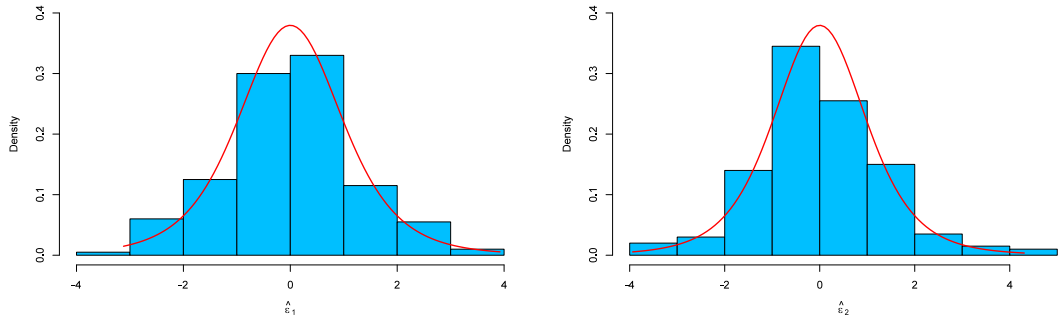
Figure 3.2: Generated bivariate time series. Each univariate time series is an autoregressive model of order 1. Both time series are connected via innovations simulated from Clayton copula with Student's t marginals.

based on Kendall's tau, the degrees of freedom are assumed to be fixed at the 4.

Table 3.3 concludes that the lowest AIC and BIC are for the Gaussian copula, which is incorrect.

Copula	Parameter estimation	Standard deviation	95 % confidence interval 2.5 %	97.5 %
Estimation based on MLE				
Clayton	$\widehat{\theta} = 1.1937$	0.1163	0.9657	1.4217
Gumbel	$\widehat{\theta} = 1.4738$	0.0673	1.3420	1.6057
Gaussian	$\widehat{\rho} = 0.5299$	0.0434	0.4448	0.6150
Student's t	$\widehat{\rho} = 0.5310$ $\widehat{\nu} = 106.1044$	0.0442	0.4443	0.6176
Estimation based on Kendall				
Clayton	$\widehat{\theta} = 1.1937$	0.1947	0.8121	1.5753
Gumbel	$\widehat{\theta} = 1.5969$	0.0974	1.4061	1.7877
Gaussian	$\widehat{\rho} = 0.5540$	0.0499	0.4561	0.6518
Student's t	$\widehat{\rho} = 0.5540$ $\widehat{\nu} = 4$ (fixed)	0.0499	0.4561	0.6518

Table 3.2: Parametric estimation of an unknown parameter of different copulas.



(a) Histogram of the $\widehat{\varepsilon}_{1,t}$

(b) Histogram of the $\widehat{\varepsilon}_{2,t}$

Figure 3.3: Histograms of the residuals of both time series compared with its true distribution, Student's t-distribution with five degrees of freedom - red curve.

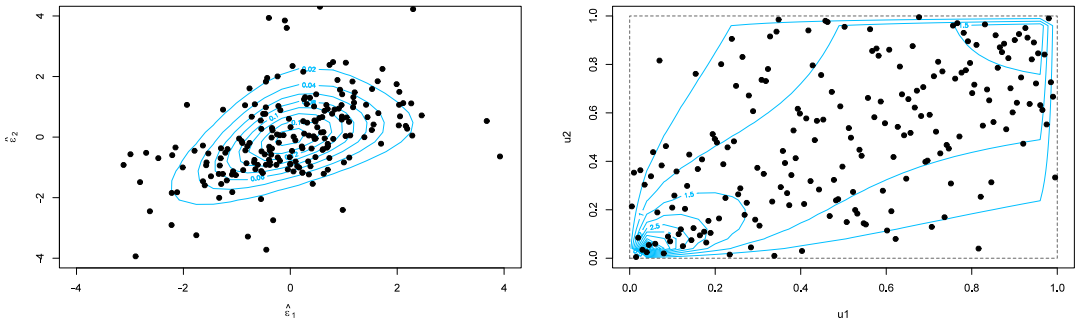
Copula	AIC	BIC
Clayton	-53.3430	-50.0447
Gumbel	-47.0408	-43.7425
Gaussian	-59.9778	-56.6794
Student's t	-57.9926	-51.3960

Table 3.3: Comparison of AIC and BIC values computed from Definition 3 using likelihood of copulas from equation (3.8).

We can also use the goodness-of-fit test defined in Section 3.3 to find the appropriate family of copulas. The p-values and values of the test statistics are given in Table 3.4 for $N = 1\,000$ replications. This table shows that Clayton and Gumbel copulas are rejected at the five per cent significant level, but Clayton copula is true. Conversely, the test does not reject the null hypothesis for Gaussian and Student's t-copulas. The wrong test decision is due to a small sample size n. The appropriate sample size is discussed in Section 4.2 of the simulations chapter.

Copula	S_n	p-value
Clayton	0.0382	0.0215
Gumbel	0.0705	0.0005
Gaussian	0.0293	0.1044
Student's t	0.0290	0.0994

Table 3.4: Goodness-of-fit test statistics and p-value for different copula families for $N = 1\,000$ replications.



(a) Scatter plot of residuals

(b) Scatter plot of pseudo-observations

Figure 3.4: Scatter plots of residuals and pseudo-observations created in equation (3.7). Both figures are compared with true contours. The first figure contains contours of Clayton copula with Student's t marginals with five degrees of freedom, and the second figure contains contours of Clayton copula with uniform marginals. In both figures are Clayton copula's contours plotted with parameter $\theta = 1$.

Predictions

In Section 3.4, predictions are discussed. Table 3.5 gives results of estimating both time series three steps ahead with its 95% confidence intervals. The predicted values can be compared with the true values that were also generated but not used in the estimation part.

The number of bootstraps for the estimation is $B = 10\,000$. Table 3.5 shows that all three algorithms provide similar prediction results, as seen from the MSE. Conclusions regarding the suitability of the prediction algorithms are discussed in the simulation chapter, see Chapter 4. In addition to the prediction interval, we also present the median as an estimate of the mean value of the series. The mean value can also be derived analytically according to Section 1.5.

Prediction performance is measured by MSE, which is defined as

$$\mathrm{MSE}_d = \frac{1}{k} \sum_{t=n+1}^{n+k} \left(Y_{d,t} - \widehat{Y}_{d,t} \right)^2, \quad d = 1,\, 2,$$

where $Y_{d,t}$ is for $t > n$ a test value for d-th time series. MSE is provided in Table 3.5 for $k = 1,\, 2,\, 3$.

Conclusion of example

This example aimed to demonstrate the presented theory of SCOMDY modelling. The model for this example was a bivariate time series consisting of two autoregressions of order one. Innovations of univariate series were linked via a Clayton copula with the marginals of the Student's t-distribution with five degrees of freedom.

Parameters for all mentioned copulas were estimated here with their standard deviations and 95% confidence intervals.

Algorithm	Quantile	$k = 1$	$k = 2$	$k = 3$
		The first time series		
True observations		1.3264	1.2135	2.7571
1st Alg.	97.5%	2.3931	2.6964	2.8122
	50%	0.1623	0.0672	0.0288
	2.5%	-2.5664	-2.7182	-2.8540
	MSE	1.3551	1.3346	3.3709
2nd Alg.	97.5%	2.4123	2.6932	2.8319
	50%	0.1234	0.0076	-0.0410
	2.5%	-2.5611	-2.8200	-2.9813
	MSE	1.4472	1.4507	3.5769
3rd Alg.	97.5%	2.3820	2.6898	2.8047
	50%	0.1333	0.0552	-0.0401
	2.5%	-2.5424	-2.8341	-3.0425
	MSE	1.4235	1.3826	3.5298
		The second time series		
True observations		1.5211	0.3473	0.7850
1st Alg.	97.5%	3.4508	3.3940	3.3989
	50%	-0.2140	0.0558	0.1576
	2.5%	-2.9000	-2.6970	-2.5586
	MSE	3.0106	1.5478	1.1631
2nd Alg.	97.5%	3.4578	3.6457	3.6344
	50%	-0.1539	0.2106	0.3157
	2.5%	-2.9464	-2.5095	-2.4041
	MSE	2.8056	1.4122	1.0149
3rd Alg.	97.5%	2.5767	3.4791	3.6585
	50%	-0.1433	0.2196	0.3174
	2.5%	-2.7790	-2.5392	-2.5609
	MSE	2.7702	1.3933	1.0017

Table 3.5: Prediction of bivariate time series three steps ahead with the 95% confidence interval.

The tricky part is the choice of the correct copula. The data is generated from the Clayton copula, which is rejected by the goodness-of-fit test. However, in this example, the test does not reject the null hypothesis for Gaussian and Student's t-copulas on the five per cent significance level. A poor test result may be due to a small number of observations. Please remember that the series was generated with a length of 200.

Finally, a prediction of both time series is made. Suppose we have two assets whose logarithmic return follows this model with a given dependence. Thus, we can construct a portfolio comprising these two assets and model its return,

including the 95% confidence interval. The power in modelling the dependence structure is the most apparent when modelling the portfolio's return. For more information on the portfolio, see below.

4. Simulations

This part of the thesis examines the functionality of the previously mentioned theory. Using simulations, we can find the weaknesses or strengths of different approaches. All the simulations performed are illustrated on a bivariate time series consisting of two autoregressions connected by a Clayton copula with marginals of the Student's t-distribution. We consider the same model with the same parameters as in Section 3.5, where an example was presented. Recall each univariate time series is an autoregressive model of order 1, fulfilling

$$\begin{aligned} Y_{1,t} &= 0.6\, Y_{1,t-1} + \varepsilon_{1,t}, \quad t \in \mathbb{Z}, \\ Y_{2,t} &= 0.4\, Y_{2,t-1} + \varepsilon_{2,t}, \quad t \in \mathbb{Z}. \end{aligned}$$

For $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t},\, \varepsilon_{2,t})^\top$, $t \in \mathbb{Z}$, it holds

$$\boldsymbol{\varepsilon}_t \sim C^{Cl}_{\theta=1}\Big(F_1(x_1),\, F_2(x_2)\Big), \quad x_1,\, x_2 \in \mathbb{R},\, t \in \mathbb{Z},$$

where $F_1(x_1)$ and $F_2(x_2)$ are cumulative distribution functions of Student's t-distribution with five degrees of freedom and $C^{Cl}_{\theta=1}(u_1,\, u_2)$ denote the Clayton copula with parameter $\theta = 1$. Let us generate bivariate data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ from this model.

We compare the copula estimates based on MLE and Kendall's tau of this bivariate time series with different time horizons. Which estimation method is better? Along with comparison, it is illustrated the consistency.

The behaviour of the goodness-of-fit test introduced in Section 3.3 is also investigated. Specifically, we address the question of the appropriate sample size for the correct test evaluation.

The three prediction algorithms are also compared here. Algorithm one is the least time-consuming to compute, but is it also applicable? Which algorithm is the most relevant?

Finally, we study prediction intervals for linear combinations of time series.

All the simulations were performed using R software, see R Core Team [2020]. Specifically for copula work, the copula package, created by Hofert et al. [2022], was used. We attach the script Simulations.R to this chapter.

4.1 MLE vs Kendall tau's estimation

This simulation compares maximum likelihood-based estimates and estimates based on Kendall's tau proposed in Section 2.5.2. The comparison is performed for series with different sample sizes, namely $n = 200,\, 500,\, 1\,000,\, 2\,000$. We always run 10 000 simulations for each sample size. For each simulation, the parameters of the copula are estimated. Estimation of bias, sample standard deviation and estimation of mean squared error (MSE) are calculated. Recall MSE of an estimator $\widehat{\theta}$ of parameter θ is defined as

$$\mathrm{MSE}(\widehat{\theta}) = \mathsf{E}_\theta\Big[(\widehat{\theta} - \theta)^2\Big] = \mathrm{var}_\theta(\widehat{\theta}) + bias(\widehat{\theta},\, \theta)^2,$$

where $bias(\widehat{\theta}, \theta)$ is defined

$$bias(\widehat{\theta}, \theta) = \mathsf{E}(\widehat{\theta}) - \theta.$$

Table 4.1 provides the results. Recall that for the Clayton copula, the true value of the parameter is $\theta = 1$. We see that the estimate based on Kendall's tau is less biased than the MLE, which on the other hand, has a smaller standard deviation. Estimation based on ML has a smaller MSE.

Sample	MLE			Kendall		
size	bias	SD	MSE	bias	SD	MSE
$n = 200$	0.0322	0.1830	0.0345	0.0067	0.2010	0.0405
$n = 500$	0.0163	0.1149	0.0134	0.0042	0.1271	0.0162
$n = 1\,000$	0.0072	0.0797	0.0064	0.0006	0.0876	0.0077
$n = 2\,000$	0.0033	0.0573	0.0033	-0.0003	0.0623	0.0039

Table 4.1: Comparison of the estimation of Clayton copula parameter based on maximum likelihood and Kendall's tau. The true value of the parameter is 1.

As a by-product, we can observe the consistency of both estimations. Table 4.1 shows that as the sample size n increases, the estimate becomes closer to the true value.

We also tried the same bivariate model but with different types of copulas. For Gumbel copula, Gaussian copula and Student's t-copula, see appendix Tables A.1, A.2 and A.3, respectively.

For the Gumbel and Gaussian copula, the estimation based on Kendall's tau is again less biased than the MLE. However, it is more biased for the Student's t-copula. It's the other way around with the standard deviation. For the Gumbel and Gaussian copula, the standard deviation is lower for estimation based on ML, and for Student's t-copula, is standard deviation lower for estimation based on Kendall's tau. The MSE is lower for MLE in all cases.

4.2 Performance of the GOF test

This section deals with the goodness-of-fit test introduced in Section 3.3. Again, we simulate data from a known bivariate time series with different sample sizes n. We count how many times the null hypothesis of the GOF test was rejected on a significant level of 0.05. We perform the test for Clayton copula, Gumbel copula, Gaussian copula and Student's t-copula. Table 4.2 contains a percentage of rejection of the null hypothesis. Recall that the true copula is Clayton copula. So the number of rejections for Clayton copula should be around 0.05. The number of simulations is set to 500, and for each series, the GOF test is computed from 350 bootstrap selections. These numbers are selected because of the runtime of the algorithm.

Table 4.2 shows that in all simulated sample sizes n, the true copula is rejected in about 5% of the cases. Furthermore, we can see that the Gaussian and Student t-copulas are less distinct from the Clayton copula for smaller n than is the case

for the Gumbel copula. Finally, we can see that the power of the test increases with increasing n and for n large enough (in our case $n \geq 1000$) the power of the test is equal to one.

Copula	$n = 200$	$n = 500$	$n = 1\,000$	$n = 2\,000$
Clayton	0.0680	0.0720	0.0260	0.0720
Gumbel	0.9880	1.0000	1.0000	1.0000
Gaussian	0.6500	0.9860	1.0000	1.0000
Student's t	0.6040	0.9840	1.0000	1.0000

Table 4.2: Percentage rejection of the null hypothesis of the goodness-of-fit test.

4.3 Prediction algorithms

In the next simulation, the prediction algorithms from Section 3.4 are compared for one-step ahead prediction. The prediction is evaluated on the above-defined two-dimensional time series for different sample sizes $n = 200, 500, 1\,000$. The number of simulations is set to 1 000. For each simulation, we generate bivariate time series of sample size $n + 1$. The estimate and prediction are constructed from the first n values, and the last one is used for evaluation, since we make one step ahead. For each simulation, 500 bootstraps were performed for each algorithm. For each simulation and each algorithm, the 95% confidence interval was constructed for both the time series and their sum. We calculated the interval length and found out if the actual value, which we predicted together with the series, is within it.

Table 4.3 presents the average length of simulated 95% confidence intervals and the coverage reliability, which is calculated as the number of actual values inside the interval divided by the number of simulations 1 000.

Sample Size	Algos	Y_1		Y_2		$Y_1 + Y_2$	
		Avg int len	Coverage	Avg int len	Coverage	Avg int len	Coverage
$n = 200$	1st algo	5.2354	0.9520	5.2411	0.9360	9.0277	0.9660
	2nd algo	5.1690	0.9320	5.1740	0.9420	8.9283	0.9500
	3rd algo	5.1638	0.9400	5.1821	0.9460	8.8910	0.9500
$n = 500$	1st algo	5.2292	0.9380	5.1395	0.9440	8.9565	0.9400
	2nd algo	5.1744	0.9480	5.2459	0.9540	9.0215	0.9360
	3rd algo	5.1803	0.9520	5.2048	0.9460	8.9366	0.9480
$n = 1\,000$	1st algo	5.2395	0.9420	5.2123	0.9420	9.0325	0.9540
	2nd algo	5.2169	0.9280	5.1664	0.9400	8.9846	0.9360
	3rd algo	5.1343	0.9460	5.1531	0.9400	8.8860	0.9520

Table 4.3: The average interval length and empirical coverage of one-step ahead prediction intervals based on three different algorithms are compared. Kendall's tau is equal to 1/3.

Table 4.3 shows that the predictive confidence intervals are similar for sample sizes n. The average interval length and coverage are close to each other. However,

for larger sample sizes $n = 1\,000$, the average length of predictive intervals is smaller for the 3rd algorithm. But coverage is still around 95%.

In Table 4.4, we present the same comparison for the above-defined model with a change in dependence. We have increased Kendall's tau to 2/3. The length of the prediction intervals of each univariate series is almost the same. However, the length of the prediction interval for the sum of the two series is larger compared to the length of the prediction interval for the sum of the two series in Table 4.3, where Kendall's tau equals 1/3. The strength of the copulas can be seen here.

Sample Size	Algos	Y_1		Y_2		$Y_1 + Y_2$	
		Avg int len	Coverage	Avg int len	Coverage	Avg int len	Coverage
$n = 200$	1st algo	5.2354	0.9520	5.2439	0.9540	9.7421	0.9540
	2nd algo	5.1690	0.9320	5.1697	0.9440	9.6791	0.9500
	3rd algo	5.1638	0.9400	5.1739	0.9540	9.6205	0.9480
$n = 500$	1st algo	5.2292	0.9380	5.1937	0.9460	9.7222	0.9420
	2nd algo	5.1744	0.9480	5.1696	0.9360	9.7130	0.9500
	3rd algo	5.1803	0.9520	5.1715	0.9460	9.6564	0.9540
$n = 1\,000$	1st algo	5.2095	0.9410	5.2036	0.9410	9.7194	0.9380
	2nd algo	5.2005	0.9360	5.1722	0.9400	9.6631	0.9380
	3rd algo	5.2241	0.9450	5.2037	0.9490	9.7394	0.9390

Table 4.4: The average interval length and reliability of the coverage of one-step ahead prediction intervals based on three different algorithms are compared. Kendall's tau is equal to 2/3.

Simulations indicate that the verification of the proposed 3rd algorithm works as expected. It seems that 1st algorithm, which is the least time-consuming, seems to be sufficient.

5. Practical application

In the last part of the thesis, we apply the acquired knowledge to the practical problem we introduced in the first chapter. We develop a model for a three-dimensional time series based on copulas. The daily closing prices of Apple Inc. (ticker AAPL), Alphabet Inc. (GOOG) and Microsoft Corporation (MSFT) are considered. Given the assumption of stationarity, we model logarithmic returns. Therefore, before the application, we present a brief theory on returns.

5.1 Returns

This section provides a brief background on the returns. Specifically, we look at the simple and logarithmic returns, see Cipra [2020].

Definition 9. *Let P_t denote the price of an asset at time t, $t \in \mathbb{Z}$. The simple return R_t of the asset is defined as*

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

The log return r_t of the asset is defined as

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right).$$

The basic properties and relationships of the returns and prices of the assets are given in the following theorem.

Theorem 10. *Let P_t denote the price of an asset at time t and R_t, r_t denote simple return and log return at time t, $t \in \mathbb{Z}$, respectively. The initial price of an asset is P_0. Then it holds*

1. $r_t = \log(1 + R_t)$,
2. $P_{t-1}\,(1 + R_t) = P_{t-1}\,\exp(r_t) = P_t$,
3. $P_t = \exp\left(\sum_{i=1}^{t} r_i\right)$,
4. $P_t = \prod_{i=1}^{t}(1 + R_i)$.

Proof. The individual points can be easily derived from Definition 9. □

5.2 Analyzing practical series

Now we move on to the analysis. As mentioned above, we model a three-dimensional time series composed of the daily closing prices of Apple Inc. (ticker

AAPL), Alphabet Inc. (GOOG) and Microsoft Corporation (MSFT). We originally wanted to model the series from January 2007 to December 2022. Unfortunately, the data corresponding to this series does not fit any copula. This could be due to the financial crisis in 2008, the COVID-19 pandemic in 2019, or a large amount of inconsistent data. In general, it is difficult to apply models to overly long series. We, therefore, decided to model a shorter series. We use data from January 2012 to December 2016, a total of $1,258$ of observations. All the data has been collected from finance.yahoo.com. The multivariate time series is presented in Figure 5.1.

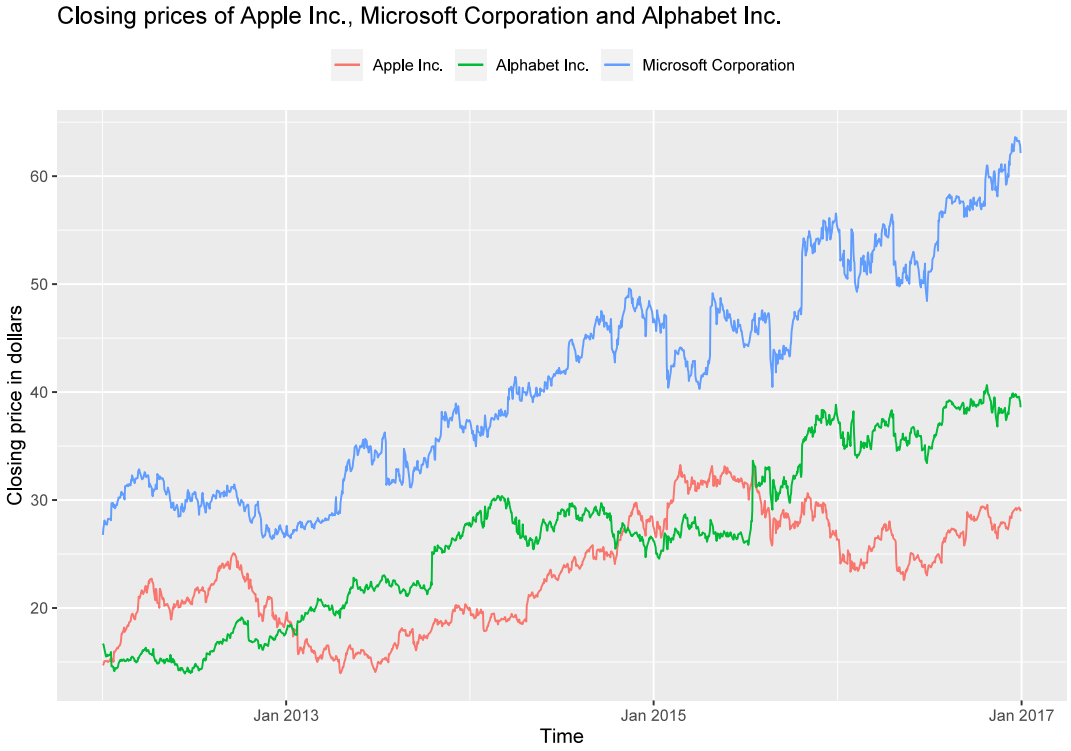


Figure 5.1: Example of multivariate time series with reduced sample size. The figure shows the closing prices of Apple Inc., Microsoft Corporation and Alphabet Inc. in dollars.

Table 5.1 presents the basic properties of examined time series. For simplicity, we refer to each company in the analysis by its ticker.

The main aim of the practical task is to predict all series for five trading days ahead, including 95% prediction intervals. Also, we want to predict the value of a portfolio consisting of these three assets.

All the statistical calculations were performed using R software, see R Core Team [2020], specifically for copula work, we use the copula package, created by Hofert et al. [2022]. For ARMA-GARCH modelling, we used the package called rugarch, see Ghalanos [2022]. We attach the script Practical_example.R to the practical analysis example.

As mentioned above, we model the log returns defined in Definition 9. Figure 5.2 shows a time series of log returns.

Characteristics	AAPL	MSFT	GOOG
minimum	13.9475	26.3700	13.9241
1st quantile	18.9698	31.8250	20.2061
median	23.6527	41.4800	26.8735
mean	23.3404	41.3663	26.6767
3rd quantile	27.5944	48.6350	32.6086
maximum	33.2500	63.6200	40.6555
std deviation	5.0745	9.9780	7.5875

Table 5.1: Basic characteristics of examined time series. All values in dollars.

5.2.1 Building the model

For the time series modelling, we use the SCOMDY model. Firstly, the logarithmic return r_t of each univariate time series is modelled using the ARMA-GARCH model. The prescription of ARMA-GARCH can be found in (1.2). Orders of the model are chosen according to AIC defined in Definition 3. Once we find the order of the model, we look to see if all its parameters are significant. Non-significant parameters are excluded. We use maximum orders for the ARMA series for the AR and MA parts of $p \leq 5$ and $q \leq 5$, respectively. For the GARCH part, we use maximum orders $m \leq 2$, $s \leq 2$. The following models were fitted

$$\begin{aligned} \text{AAPL} &\sim \text{ARMA}(2,\,3)\text{-GARCH}(1,\,1), \\ \text{MSFT} &\sim \text{ARMA}(2,\,2)\text{-GARCH}(1,\,1), \\ \text{GOOG} &\sim \text{ARMA}(5,\,5)\text{-GARCH}(1,\,1). \end{aligned}$$

The estimated parameters with their standard deviations are presented in Table 5.2.

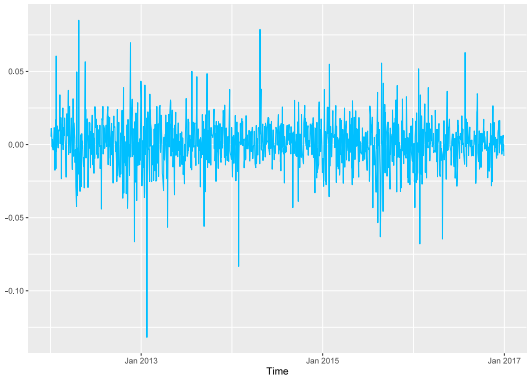
The model needs to be verified. The estimated parameters fulfil the conditions mentioned in (1.3). Correlograms of the standardised residuals and of the squares of the standardised residuals were examined, and no patterns were found. Basic diagnostic tools do not indicate any violation of assumptions, and we can work with estimated models.

For interest, see standardised residuals in Figure 5.3. In Figure 5.3(a) are the standardised residuals of the AAPL series, in 5.3(b) are the standardised residuals of the MSFT series and finally, in 5.3(c) are the standardised residuals of the GOOG series.

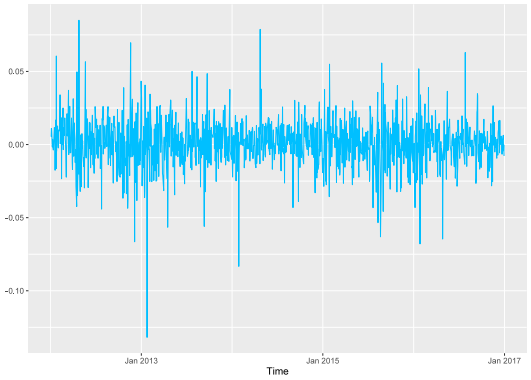
In the second step, we find the empirical distribution function of residuals.

Finally, we can estimate the copula parameter. Before estimating, we need to find the most suitable copula. We choose from the Clayton, Gumbel, Gaussian, and the Student's t-copula. The selection is based on the GOF test, see Section 3.3. From Table 5.3, we cannot reject Student's t-copula on 5% significance level. Also, the AIC and BIC values are the lowest for Student's t-copula. The Student's t-copula seems to be applicable.

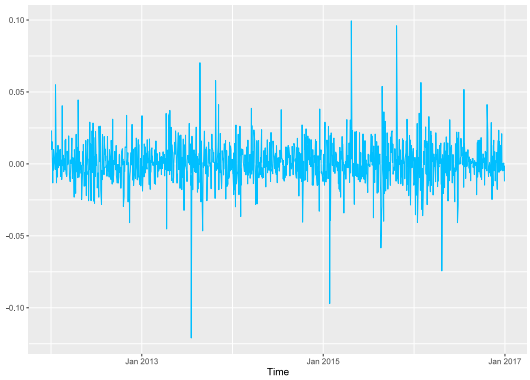
The estimated parameters of Student's t-copula follow. The covariance matrix



(a) Log return of Apple Inc.



(b) Log return of Microsoft Corporation

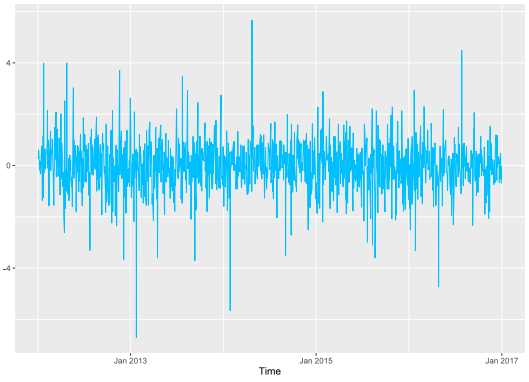


(c) Log return of Alphabet Inc.

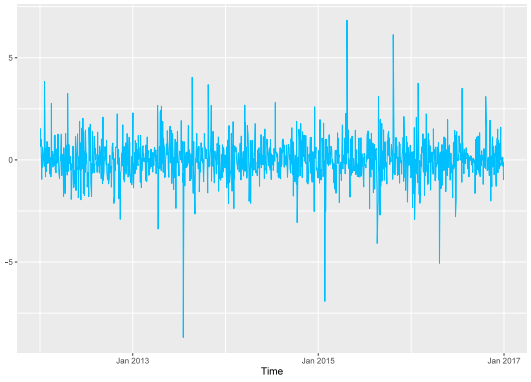
Figure 5.2: Log returns of all the three considered companies.

Param	AAPL		MSFT		GOOG	
	Estim	SD	Estim	SD	Estim	SD
$\widehat{\mu}$	0.0010	0.0005	0.0007	0.0004	0.0006	0.0000
$\widehat{\phi}_1$	-1.3408	0.0020	1.8595	0.0076	0.6834	0.0000
$\widehat{\phi}_2$	-0.9826	0.0053	-0.9597	0.0077	1.8192	0.0002
$\widehat{\phi}_3$					-1.2921	0.0001
$\widehat{\phi}_4$					-0.9762	0.0001
$\widehat{\phi}_5$					0.7089	0.0001
$\widehat{\theta}_1$	1.3655	0.0010	-1.8784	0.0003	-0.7188	0.0001
$\widehat{\theta}_2$	1.0203	0.0001	0.9687	0.0007	-1.8348	0.0001
$\widehat{\theta}_3$	0.0171	0.0003			1.3535	0.0001
$\widehat{\theta}_4$					0.9978	0.0001
$\widehat{\theta}_5$					-0.7378	0.0001
$\widehat{\alpha}_0$	0.00002	0.0000	0.00000	0.0000	0.00001	0.0000
$\widehat{\alpha}_1$	0.0749	0.0184	0.0087	0.0020	0.4298	0.0700
$\widehat{\beta}_1$	0.8637	0.0208	0.9761	0.0046	0.2527	0.0827

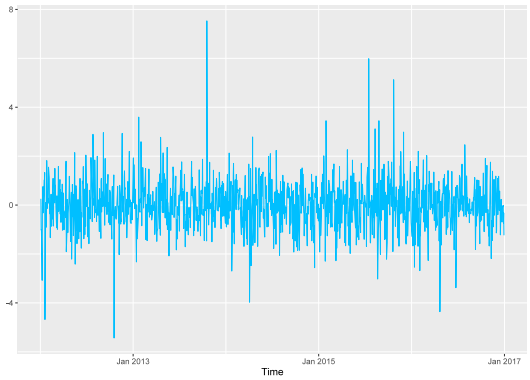
Table 5.2: Estimated parameters for each time series.



(a) Standardised residuals of AAPL



(b) Standardised residual of MSFT



(c) Standardised residuals of GOOG

Figure 5.3: Standardised residuals.

Copula	AIC	BIC	Statistics S_n	p-value
Clayton	-484.0706	-478.9341	0.3679	0.0005
Gumbel	-507.7244	-502.5879	0.2514	0.0005
Gaussian	-533.4013	-517.9918	0.0522	0.0105
Student's t	-640.8629	-620.3170	0.0345	0.1304

Table 5.3: AIC, BIC and p-value of GOF test for copula selection.

is estimated as

$$\widehat{R} = \begin{pmatrix} 1 & 0.3723 & 0.3799 \\ 0.3723 & 1 & 0.4794 \\ 0.3799 & 0.4794 & 1 \end{pmatrix},$$

and estimation of degrees of freedom is $\widehat{\nu} = 5.4643$. We finally have a model built.

5.2.2 Prediction

In this section, we use the model we estimated. Five steps ahead prediction is performed, $K = 5$. We construct the estimation of expectations and their 95% confidence intervals. Section 3.4 presents the theory of confidence interval

creation, and Section 1.5 provides an analytical estimation of the expectation. We create confidence intervals by creating a 10 000 bootstraps interaction. The predicted logarithmic returns were converted to closing prices using Theorem 10.

Table 5.4 provides closing prices for the AAPL time series. Results of MSFT can be found in Table 5.5. Last Table 5.6 provides results regarding GOOG. In these tables, the relevant quantiles (2.5%, 50%, 97.5%) are recorded with an analytical prediction from Section 1.5, which is denoted as "Pred". "True" remains for the observed value not used for the estimation.

Algos	Date	03/01/17 $k=1$	04/01/17 $k=2$	05/01/17 $k=3$	06/01/17 $k=4$	09/01/17 $k=5$
	True	29.0375	29.0050	29.1525	29.4775	29.7475
	Pred	29.2593	29.0778	29.0530	29.2966	29.0265
	97.5%	29.6490	30.0087	30.2764	30.4682	30.7151
1st algo	50%	28.9315	28.9603	28.9779	28.9719	28.9981
	2.5%	28.2643	27.9434	27.6854	27.4613	27.3127
	97.5%	30.5781	31.0064	31.5138	31.9133	32.2699
2nd algo	50%	28.9729	28.9516	28.9945	28.9809	29.0064
	2.5%	27.4285	26.5436	25.6726	25.0489	24.6652
	97.5%	30.6274	30.9732	31.7170	31.9271	32.3033
3rd algo	50%	28.9430	28.9264	28.9598	28.9267	28.9563
	2.5%	27.4526	26.4672	25.8312	25.4051	24.9159

Table 5.4: Point prediction and 95% predictive bootstrapped confidence interval for AAPL with different algorithms. All values are in dollars.

Algos	Date	03/01/17 $k=1$	04/01/17 $k=2$	05/01/17 $k=3$	06/01/17 $k=4$	09/01/17 $k=5$
	True	62.5800	62.3000	62.3000	62.8400	62.6400
	Pred	61.6696	60.5342	59.7048	58.3438	57.3978
	97.5%	63.7100	64.3897	64.4530	64.4281	64.2721
1st algo	50%	62.0532	61.8527	61.4788	61.0248	60.4816
	2.5%	60.4251	59.6244	58.8298	57.9250	57.0299
	97.5%	66.3132	67.7811	72.2152	77.3351	87.1383
2nd algo	50%	62.7088	63.0091	64.2359	64.9341	66.4316
	2.5%	59.3460	57.6057	56.8513	57.1709	58.2618
	97.5%	66.2474	68.0299	71.1211	76.3173	84.6847
3rd algo	50%	62.7026	63.0392	64.3113	64.9994	66.4312
	2.5%	59.7366	57.8730	57.2467	57.2434	58.4250

Table 5.5: Point prediction and 95% predictive bootstrapped confidence interval for MSFT with different algorithms. All values are in dollars.

5.2.3 Portfolio creation

We create a fictitious portfolio consisting of a single stock of Apple Inc, Microsoft Corporation, and Alphabet Inc. We want to know the value of the portfolio. Five days ahead, a prediction is made, including a 95% confidence interval. In Table 5.7, the results can be found.

Algos	Date	03/01/17 $k=1$	04/01/17 $k=2$	05/01/17 $k=3$	06/01/17 $k=4$	09/01/17 $k=5$
	True	39.3070	39.3450	39.7010	40.3075	40.3325
	Pred	38.3597	37.8040	37.5469	37.0913	36.9307
1st algo	97.5%	39.5797	40.0537	40.3857	40.7868	41.2244
	50%	38.6021	38.6303	38.6747	38.8027	38.9127
	2.5%	37.6585	37.2355	37.0108	36.8821	36.6911
2nd algo	97.5%	42.3190	42.9164	46.6534	49.9361	57.1252
	50%	39.0160	39.1355	39.8116	40.0388	40.8724
	2.5%	36.1808	34.4274	34.0166	33.4840	34.1766
3rd algo	97.5%	42.4591	43.3097	46.7600	50.1000	57.3320
	50%	39.0321	39.1450	39.8693	40.0338	40.9816
	2.5%	36.2557	34.7374	34.6064	33.7158	34.8994

Table 5.6: Point prediction and 95% predictive bootstrapped confidence interval for GOOG with different algorithms. All values are in dollars.

Algos	Date	03/01/17 $k=1$	04/01/17 $k=2$	05/01/17 $k=3$	06/01/17 $k=4$	09/01/17 $k=5$
	True	130.9245	130.6500	131.1535	132.6250	132.7200
	Pred	129.2885	127.4160	126.3046	124.7318	123.3550
1st algo	97.5%	131.7606	132.5654	132.9345	133.2049	133.3177
	50%	129.6047	129.4966	129.1459	128.8158	128.4089
	2.5%	127.5800	126.5793	125.7082	124.7794	123.9102
2nd algo	97.5%	135.8885	137.9470	144.4599	152.6146	169.6926
	50%	130.7136	130.9919	133.1124	134.0947	136.8538
	2.5%	126.2024	123.5961	122.8496	122.0269	122.9959
3rd algo	97.5%	136.0510	137.8639	144.8486	151.1231	164.4251
	50%	130.7300	130.9313	133.1592	134.1938	136.9095
	2.5%	126.2311	123.8506	123.5279	122.5839	124.1392

Table 5.7: Point prediction and 95% predictive bootstrapped confidence interval for portfolio created from AAPL, MSFT and GOOG. All values are in dollars.

5.2.4 Conclusion of practical example

In the example, we have shown the functioning of the SCOMDY model on real data. From the prediction part, we can see that predictive intervals are shorter for the first algorithm. From the behaviour of the series, we can conclude that this algorithm will not be suitable for long-term interval predictions as it does not take into account the previously mentioned variability of the estimate. The interval is inaccurate.

If we wanted to model the full, original series from January 2007 to December 2022, it would be necessary to make a complex model. For example, we can use a model that changes the value of the copula parameter over time. Alternatively, we can change the whole copula during the time. However, this topic is beyond the scope of this work.

Conclusion

This thesis discussed an approach to modelling multivariate time series. In particular, we have considered a copula-based approach. To address this topic, it was necessary to introduce ARMA-GARCH models and copulas. Thus, in the first chapter, we discussed ARMA-GARCH models. We discussed how to estimate it and how to make a prediction. We defined multivariate time series and motivated a practical example.

The second chapter summarizes the basic theory of copulas. We introduced Sklar's theorem, which is a key theorem for copula theory, and mentioned the most basic four copulas - Clayton, Gumbel, Gaussian and Student's t-copula. Copula with different values of the dependence parameter and different marginal functions were visualised. Subsequently, we derived different methods for estimating the unknown parameters. Specifically, we discussed parametric, semi-parametric, and non-parametric approaches.

The main chapter of the thesis is chapter three, where the SCOMDY model was introduced. We presented a goodness-of-fit test. Three different bootstrap algorithms have been proposed to build confidence intervals for the prediction of conditional expectations of time series. Furthermore, the prediction algorithms could be extended to the estimation of conditional volatility.

In the fourth chapter, we dealt with simulations. A comparison between the ML-based estimate and Kendal's tau-based estimate is presented. As a result of the study, we found that the MLE has a smaller MSE than the estimator based on Kendal's tau. Furthermore, we tested the functionality of the GOF test. We found that if we increase the sample size n, the power of the test increase. The test rejects the incorrect hypothesis reliably on the sample size of at least $n = 1\,000$. Finally, we tested the proposed prediction algorithms. We found that our proposed third algorithm gives us the expected results. For n large enough, the prediction interval length of the third algorithm is the smallest among the three proposed algorithms. But we also found that the first algorithm, which is time-saving, also gives usable results. Hence, using this algorithm in applications where time plays an important role can be advisable.

Finally, in Chapter 5, we applied the theory to a real problem. We investigated a three-dimensional time series composed of the closing stock prices of Apple Inc., Microsoft Corporation, and Alphabet Inc., Google's parent company. We examined the series over a time horizon of five years, from January 2012 to December 2016. We created prediction and 95% prediction intervals for five steps ahead and then created a fictitious portfolio consisting of one stock of each mentioned company. Thus, we examined the evolution of the portfolio price over time. It is worth noting the lengths of the prediction intervals, which are shorter for the portfolio than if we had created the sum of the individual stocks and considered them to be independent. In fact, the dependency structure modelled by the copula has played a role in this result.

The limitations of the SCOMDY model, are primarily related to the maximum dimension, D. From a practical point of view, a maximum of 10 series is recommended to study. This is because the number of unknown parameters increases, and the model becomes more complex. Oh and Patton [2017] proposes

how to model more dimensions, up to a maximum of 100.

Other models not discussed in this thesis are models with the so-called time-varying copula. This is a model where the conditional volatility of the time series varies over time. This topic is addressed in Patton [2001].

In practice, it is also sometimes necessary to model time series of different lengths. One possibility is to model only part of all series. However, it is possible to use theory concerning multivariate time series of different lengths. This issue is discussed in Patton [2006].

Bibliography

M. Akahira and K. Takeuchi. *Non-regular statistical estimation*, volume 107. Springer Science & Business Media, 2012.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

G. E. Box, G. M. Jenkins, and G. Reinsel. Time series analysis: forecasting and control holden-day san francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day1970*, 1970.

P. J. Brockwell and R. A. Davis. *Introduction to time series and forecasting*. Springer, 2002.

K. P. Burnham and D. R. Anderson. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304, 2004.

H. Carley and M. Taylor. A new proof of sklar's theorem. In *Distributions with given marginals and statistical modelling*, pages 29–34. Springer, 2002.

S. X. Chen and T.-M. Huang. Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, 35(2):265–282, 2007.

X. Chen and Y. Fan. Specification and estimation of a new class of semiparametric copula-based multivariate dynamic models. *Manuscript, New York University and Vanderbilt University*, 2003.

X. Chen and Y. Fan. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of econometrics*, 135(1-2):125–154, 2006.

B. Choroś, R. Ibragimov, and E. Permiakova. Copula estimation. In *Copula theory and its applications*, pages 77–91. Springer, 2010.

T. Cipra. *Time series in economics and finance*. Springer, 2020.

P. Embrechts. Copulas: A personal view. *Journal of Risk and Insurance*, 76(3): 639–650, 2009.

P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223, 2002.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

J. A. Fessler. On transformations of random vectors. *Commun. Signal Process. Lab., Dept. Elect. Eng. Comput. Sci., Univ. Michigan, Ann Arbor, MI, http://www. eecs. umich. edu/˜ fessler*, 1998.

C. Francq and J.-M. Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.

A. Ghalanos. *rugarch: Univariate GARCH models.*, 2022. R package version 1.4-9.

M. B. Ghalibaf. Relationship between kendall's tau correlation and mutual information. *Revista Colombiana de Estadística*, 43(1):3–20, 2020.

J. D. Gibbons and S. Chakraborti. Nonparametric statistical inference. In *International encyclopedia of statistical science*, pages 977–979. Springer, 2011.

M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan. *copula: Multivariate Dependence with Copulas*, 2022. URL `https://CRAN.R-project.org/package=copula`. R package version 1.1-0.

H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of multivariate Analysis*, 94(2):401–419, 2005.

B. G. Kibria and A. H. Joarder. A short review of multivariate t-distribution. *Journal of Statistical research*, 40(1):59–72, 2006.

D. H. Oh and A. J. Patton. Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35(1):139–154, 2017.

L. Pascual, J. Romo, and E. Ruiz. Bootstrap predictive inference for arima processes. *Journal of Time Series Analysis*, 25(4):449–465, 2004.

A. Patton. Copula methods for forecasting multivariate time series. *Handbook of economic forecasting*, 2:899–960, 2013.

A. J. Patton. Modelling time-varying exchange rate dependence using the conditional copula. 2001.

A. J. Patton. Estimation of multivariate models for time series of possibly different lengths. *Journal of applied econometrics*, 21(2):147–173, 2006.

A. J. Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

B. Rémillard. Goodness-of-fit tests for copulas of multivariate time series. *Econometrics*, 5(1):13, 2017.

O. Scaillet and J.-D. Fermanian. Nonparametric estimation of copulas for time series. *FAME Research paper*, (57), 2002.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

L. A. Thombs and W. R. Schucany. Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, 85(410):486–492, 1990.

P. K. Trivedi, D. M. Zimmer, et al. Copula modeling: an introduction for practitioners. *Foundations and Trends® in Econometrics*, 1(1):1–111, 2007.

R. S. Tsay. *Analysis of financial time series*. John wiley & sons, 2005.

A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

A. Appendix

Sample	MLE			Kendall		
size	bias	SD	MSE	bias	SD	MSE
$n = 200$	0.0172	0.0991	0.0101	0.0023	0.1006	0.0101
$n = 500$	0.0080	0.0614	0.0038	0.0014	0.0630	0.0040
$n = 1000$	0.0043	0.0426	0.0018	0.0004	0.0439	0.0019
$n = 2000$	0.0024	0.0305	0.0009	0.0003	0.0313	0.0010

Table A.1: Comparison of the estimation of Gumbell copula parameter based on maximum likelihood and Kendall's tau. The true value of the parameter is 1.5.

Sample	MLE			Kendall		
size	bias	SD	MSE	bias	SD	MSE
$n = 200$	0.0072	0.0541	0.0030	-0.0035	0.0574	0.0033
$n = 500$	0.0029	0.0341	0.0012	-0.0022	0.0363	0.0013
$n = 1000$	0.0017	0.0237	0.0006	-0.0010	0.0252	0.0006
$n = 2000$	0.0009	0.0168	0.0003	-0.0007	0.0178	0.0003

Table A.2: Comparison of the estimation of Gauss copula parameter based on maximum likelihood and Kendall's tau. The true value of the parameter is 0.5.

Sample	MLE			Kendall		
size	bias	SD	MSE	bias	SD	MSE
$n = 200$	-0.0001	0.0643	0.0041	-0.0039	0.0638	0.0041
$n = 500$	0.0005	0.0404	0.0016	-0.0010	0.0403	0.0016
$n = 1000$	-0.0003	0.0283	0.0008	-0.0010	0.0283	0.0008
$n = 2000$	-0.0004	0.0198	0.0004	-0.0007	0.0198	0.0004

Table A.3: Comparison of the estimation of Student's t-copula parameter based on maximum likelihood and Kendall's tau. The true value of the parameter is 0.5.