Vine Copulas For Multivariate Time Series Modelling

With An Application In Energy Finance

Janus S. Valberg-Madsen

Master's Thesis, Mathematics-Economics

AALBORG UNIVERSITY
DENMARK

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

Title
Vine Copulas For Multivariate Time Series Modelling, With An Application In Energy Finance

Project Type
Master's Thesis, Mathematics-Economics

Project Period
Summer 2019

Author
Janus S. Valberg-Madsen

Supervisor
Esben Høg

Page Numbers
93

Date of Completion
2019-09-16

Abstract

We present the theory of *copulas*, functions used to link marginal variables into a multivariate distribution, and give the definitions and properties necessary to prove the main result, Sklar's Theorem, and show how a set of 2-dimensional copulas can be used to construct a high-dimensional distribution called a *vine copula*.
We use this model to describe the joint behaviour of German day-ahead electricity prices, power consumption, and solar and wind power generation.
Based on this model, we quantify the effect that extreme weather scenarios has on the price, concluding that wind power generation especially has a strong effect on prices.
Using simulations of the model, we investigate the payoff distributions of portfolios of *power purchase agreements* based on the variables considered, and we find that such distributions have long left tails, resulting in large values for risk measures such as VaR and expected shortfall.

Preface

This master's thesis is written in the summer of 2019 by Janus S. Valberg-Madsen, student in Mathematics-Economics at the Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark.

The document is typeset with LaTeX, data used in Part II is downloaded from the data platform for the European Network of Transmission System Operators for Electricity [ENTSO-E, 2019], and computations, modelling, and most figures are performed using the R language [R Core Team, 2019] with the packages

- `readr`[1] and `feather`[2] for data import and export
- `dplyr`[3], `tibble`[4], and `purrr`[5] for data processing
- `rugarch`[6], `copula`[7], and `VineCopula`[8] for modelling
- `ggplot2`[9] and `xtable`[10] for making plots and tables

In addition, some custom functionality has been written in a companion package to this project, `rjsvmmt`. This package, and the code in general, is not given or described in the text of the document, but is available upon request.

The author would like to acknowledge the project advisor, Esben P. Høg, associate professor, and thank him for his inputs—and his patience—throughout the project period.

[1] Wickham et al. [2018]
[2] Wickham [2019]
[3] Wickham et al. [2019]
[4] Müller and Wickham [2019]
[5] Henry and Wickham [2019]
[6] Ghalanos [2019]
[7] Hofert et al. [2018]
[8] Nagler et al. [2019]
[9] Wickham [2016]
[10] Dahl et al. [2019]

Contents

List of Figures

List of Tables

1 Introduction

In recent decades, the European electricity market has become liberalised, transitioning from an industry in which regulators fixed the price based on costs to a competitive market in which prices are determined from supply and demand.

The supply curve can be considered as a step function, where each step corresponds to a type of generation. These upwards jumps reflect the nature of the providers' preferences: if the spot market price is higher than the marginal costs of production, the provider turns their generators on and produce as much as possible, and if the prices are lower than the costs, they keep the generators off.[1] At the same time, the demand for electricity is fairly inelastic. The end consumer does in general not care about spot prices, as they will have price agreements with their distributor, and the same is the case for most industrial consumers. This makes the price very sensitive to the quantity produced by the generators with the lowest production costs, since the price is set by the market to ensure enough production to match the demand.

Renewables such as solar and wind power generally are always running, and have very small marginal costs, but the quantity fluctuates with the weather. In the other end of the curve, fossil fuel based generators are associated with large costs, but their production quantity is easily controlled. Therefore, when the renewables are not producing a lot, the power from low-cost generator types are not enough to cover the consumption, so the price goes up to meet the cost for more expensive types of production. When renewables produce a lot of electricity, it suffices to run low-cost plants to meet demand, and the price will therefore be lower.

A sketch of the situation is shown on Figure 1.1. Here, the supply curve is drawn for two different scenarios: one in which the wind power production is low (with dashed lines) and one in which it is high (the solid line). The arrows indicate the right shift of the supply curve, when the size of the renewable production grows, which moves the intersection with the demand curve down a step—from the high price level P_1 to the lower P_2.

The European Union is right now in the process of transforming its economy, aiming to minimise its emission of greenhouse gases, and as the system transitions towards a low-carbon economy, the role of renewable energy sources is increasing. The share of total energy used comprised by electricity is also expected to grow, with e.g. the focus on switching to electricity based transportation.

[1] This is of course a simplification of the real choices the providers face. Due to large costs associated with turning on or ramping up production or physical limitations on uptime on certain types of generators, such providers will have to take into account more factors than just the difference between spot and cost.

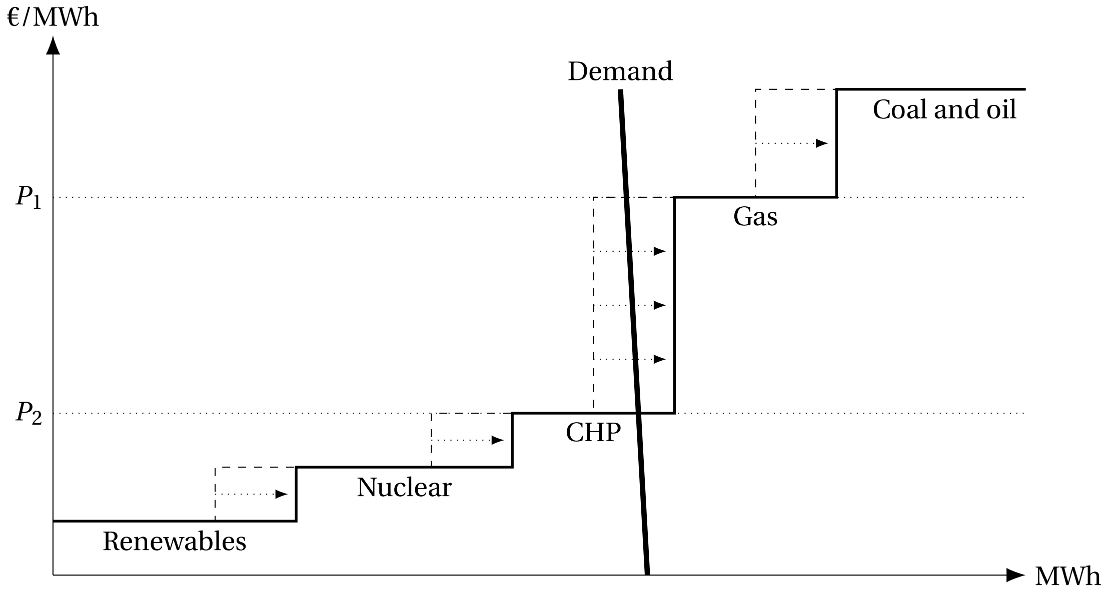


Figure 1.1: *Sketch of the supply and demand curves on the electricity market for a low wind scenario (the dashed line) and a high wind scenario (the solid line). Renewables comprise solar, wind, and hydro power; CHP (combined heat and power) are heat engines that generate electricity and useful heat simultaneously; gas, coal, and oil represent combustion based turbines based on the respective fuels.*

As a larger and larger part of the available power production capacity is comprising variable generators such as photovoltaic panels and wind turbines, the dependence between spot prices and generation of renewable energy is also increasing, and agents trading on the market are increasingly exposed to variations in the weather.

In this project, we seek to describe the relationship between a small subset of the energy market variables—namely the day-ahead spot prices, the consumption of electricity (also called *load*, and representing demand), the quantity of electricity produced with solar power generators, and the quantity produced by wind turbines—for Germany, as it is a country for which a sizeable part of its energy generation comes from renewables.

To this end, we employ the *copula* of [Sklar, 1959], who showed that any joint distribution can be expressed in terms of its marginal distributions and a copula, effectively allowing us to model marginal and joint behaviour separately. The copula is a flexible tool for modelling multivariate distributions, and it has seen much use in financial mathematics in recent years. We present the parts of copula theory necessary to prove the main theorem, and use it to develop a model for the joint distribution of the four variables. We then use this model to simulate payoff distributions for portfolios consisting of instruments with the four variables as the underlying.

1.1 Thesis Structure

The thesis is split into two parts, with the first part focusing on developing the theoretical foundations and the second part focusing on applying the results to real data. The theoretical part is further subdivided into two chapters:

- Chapter 2 starts from the basics of probability theory and introduces concepts and results about copulas and their properties, building up to the statement and proof of Sklar's Theorem. Then, a handful of specific types of copulas and dependence measures, which will be used in the second part, are given.
- Chapter 3 introduces the concept of *vines*, making use of the copulas defined in Chapter 2, and describes how parameters of such constructions are estimated and how to simulate realisations of the distributions they represent.

The second part takes the theoretical results and puts them to practical use through the joint analysis of data from the German energy market. This analysis is described in three separate chapters:

- Chapter 4 presents the data and goes into details about how it is preprocessed and aggregated prior to modelling.
- Chapter 5 describes the modelling steps. First, the marginal models are described and fitted to each variable separately, with considerations specific to the variable given along the way. Then, the variables are tied together with a vine copula model, and the structure and parameters are interpreted.
- Chapter 6 imagines a trading scenario in which the joint model could be useful, presenting a portfolio setup for which payoff distributions are simulated. Risks associated with these distributions are assessed, and a portfolio with minimal risk is found.

Finally, the results and considerations are summarised in Chapter 7, where also some ideas for further research that arose during the project are presented. An appendix with overview tables for Chapter 5 is included in the backmatter, along with a list of references.

Part I
Theoretical Foundation

2 Copula Theory

The main object of interest in this project is the *copula* (pl. copulas or copulae), a multivariate distribution function with uniform marginals. The name comes from the Latin for "link", and the reason for this name comes from an important result by [Sklar, 1959], who showed that any multivariate distribution can be represented in terms of its marginals and a copula.

This result is known as *Sklar's Theorem*, included in this project as Theorem 2.20, and in this chapter, a selection of basic results and definitions needed to prove it are given. These results are largely based on [Durante and Sempi, 2016], a textbook in copula theory that collects results from various sources. Where not specified, the proofs given are adapted from that book.

2.1 Notation

The following is a list of conventions and informal definitions that are used as a basis for other results and definitions throughout the project.

- In addition to the usual, well-known sets of numbers, $\mathbb{N} = \{1, 2, \dots\}$, $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$, $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$, $\mathbb{R} = (-\infty, +\infty)$, and $\mathbb{R}_+ = [0, +\infty)$, we also make use of shorthand notation for the *unit interval*, $\mathbb{I} := [0, 1]$.

- $\mathbb{R}^d$ is the cartesian product of $d \in \mathbb{N}$ copies of $\mathbb{R}$,

$$\mathbb{R}^d := \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \mathbb{R}}_{d \text{ times}}, \tag{2.1}$$

  and similarly for other sets. A vector belonging to such a product is said to be *d-dimensional*, and the scalar d will always be used to denote a general number of dimensions.

- The vectors $\mathbf{0}$ and $\mathbf{1}$ are d-dimensional vectors of all zeros or ones, respectively. Note that d is not specified in the notation for these, as it is usually clear enough from the context in which they are used.

- Let $\boldsymbol{x}$ be some d-dimensional vector. For shorthand notation, we define $\boldsymbol{x}_j(t)$, where $j \in \{1, \dots, d\}$, to be the vector obtained by swapping x_j in $\boldsymbol{x}$ with t,

$$\boldsymbol{x}_j(t) := (x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_d). \tag{2.2}$$

- For two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ where $x_j \leq y_j$ for every $j = 1, \dots, d$, then we write $\boldsymbol{x} \leq \boldsymbol{y}$ (and similarly for $\boldsymbol{x} \geq \boldsymbol{y}$, $\boldsymbol{x} > \boldsymbol{y}$, and $\boldsymbol{x} < \boldsymbol{y}$).

- If $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ and $\boldsymbol{x} \leq \boldsymbol{y}$, then the (left-open) d*-box* $(\boldsymbol{x}, \boldsymbol{y}]$ is defined by

$$(\boldsymbol{x}, \boldsymbol{y}] := (x_1, y_1] \times (x_2, y_2] \times \cdots \times (x_d, y_d], \tag{2.3}$$

  and similarly for closed, open, and right-open d-boxes.

- $\mathbb{1}_A(\boldsymbol{x})$ denotes the indicator function of a set $A \subseteq \mathbb{R}^d$,

$$\mathbb{1}_A(\boldsymbol{x}) = \mathbb{1}(\boldsymbol{x} \in A) := \begin{cases} 1, & \text{if } \boldsymbol{x} \in A \\ 0, & \text{otherwise} \end{cases} \tag{2.4}$$

- The *left* and *right limits* of a univariate function $\phi : \mathbb{R} \to \mathbb{R}$ at t are defined as

$$\ell^- \phi(t) := \lim_{s \uparrow t} \phi(s) \quad \text{and} \quad \ell^+ \phi(t) := \lim_{s \downarrow t} \phi(s), \tag{2.5}$$

  respectively, where such limits exist.

- When referring to the monotone properties of a univariate function, f, we use the following terms to avoid any ambiguity:
  - f is *non-decreasing* if $\forall x, y \in \operatorname{dom} f$ such that $x \leq y$, one has $f(x) \leq f(y)$.
  - f is *non-increasing* if $\forall x, y \in \operatorname{dom} f$ such that $x \leq y$, one has $f(x) \geq f(y)$.
  - f is *strictly increasing* if $\forall x, y \in \operatorname{dom} f$ such that $x < y$, one has $f(x) < f(y)$.
  - f is *strictly decreasing* if $\forall x, y \in \operatorname{dom} f$ such that $x < y$, one has $f(x) > f(y)$.

- A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where
  - Ω is an arbitrary set of elements called *outcomes*.
  - $\mathcal{F}$ is a σ-algebra over Ω, i.e. $\mathcal{F} \subseteq 2^\Omega$ such that $\Omega \in \mathcal{F}$, and $\mathcal{F}$ is closed under complements and countable unions. Elements of $\mathcal{F}$ are referred to as *events*.
  - $\mathbb{P} : \mathcal{F} \to \mathbb{I}$ is a countably additive measure with $\mathbb{P}(\Omega) = 1$ called the *probability measure*, i.e. it assigns a non-negative probability to each event in $\mathcal{F}$.

- When considering events on a probability space in the context of some random variable $X : \Omega \to \mathbb{R}$, we denote such a set as

$$\{\text{some condition about } X\} := \{\omega \in \Omega : \text{some condition about } X(\omega)\}.$$

  In general, we omit ω everywhere, as it is never considered by itself, and we will use the shorthand notation $X \in \mathbb{R}$ to mean $\operatorname{ran} X \subseteq \mathbb{R}$.

- When evaluating a multivariate function H of d arguments, this can be denoted as e.g. $H(x_1, \ldots, x_d)$ or $H(\boldsymbol{x})$, where $\boldsymbol{x}$ is a d-dimensional vector. In this project, we use both forms — the former usually in low dimensions (e.g. $d = 2$), the latter usually for general dimensions.

  A mix of notation may also be used if the arguments logically belong in different groups, e.g. $H(\boldsymbol{y}; \boldsymbol{\theta})$, where d arguments, $y_1, \ldots, y_d$ belong to some "data" domain, and k arguments, $\theta_1, \ldots, \theta_k$ belong to some "parameter" domain.

- For a vector $\boldsymbol{x} \in \mathbb{R}^d$, given a subset of indices $\mathcal{J} \subseteq \{1, \ldots, d\}$, we use the notation $\boldsymbol{x}_{\mathcal{J}}$ for the subvector $(x_j : j \in \mathcal{J})$ and $\boldsymbol{x}_{-\mathcal{J}}$ for $(x_j : j \notin \mathcal{J})$.

2.2 Basic Definitions and Results

We begin this section by recalling some fundamental definitions and results of probability theory. Note that some of the proofs draws from the wider area of measure theory, which we will not be covering in details here.

Definition 2.1. Let X be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then its corresponding *distribution function*, $F : \mathbb{R} \to \mathbb{I}$ is defined by

$$F(x) := \mathbb{P}(X \leq x). \tag{2.6}$$

A random variable is fully characterised by its distribution function, and we use the notation $X \sim F$ to mean "X has the distribution function F". Such a function has the following analytical properties:

Theorem 2.2. *A function $F : \mathbb{R} \to \mathbb{I}$ is a distribution function for a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if and only if*

(a) F is right-continuous on $\mathbb{R}$.

(b) F is non-decreasing.

(c) F satisfies the limits

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1. \tag{2.7}$$

Proof. Theorem 2.2 is not proven in [Durante and Sempi, 2016], but it can proven by using properties of measures. Let $A_x = \{X \leq x\}$, and note that F can equivalently be written as

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(A_x). \tag{2.8}$$

(a): A function F is right-continuous at x if $\ell^+ F(x) = F(x)$. Let $x \in \mathbb{R}$ and let $\{x_n\}_{n=1}^{\infty}$ be an arbitrary, non-increasing sequence such that $\lim_{n \to \infty} x_n = x$. Then, by definition, $A_x \subseteq A_{x_n}$ for all $n \in \mathbb{N}$, and

$$\bigcap_{n=1}^{\infty} A_{x_n} = A_x, \tag{2.9}$$

and, since $\mathbb{P}$ is a measure, we have

$$\ell^+ F(x) = \lim_{n\to\infty} F\left(x_n\right) = \lim_{n\to\infty} \mathbb{P}\left(A_{x_n}\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_{x_n}\right) = \mathbb{P}\left(A_x\right) = F(x). \tag{2.10}$$

(b): For any $x, y \in \mathbb{R}$ such that $x < y$, we have that $A_x \subseteq A_y$, and therefore

$$\mathbb{P}\left(A_x\right) \le \mathbb{P}\left(A_y\right) \implies F(x) \le F\left(y\right). \tag{2.11}$$

(c): Let $\left\{x_n\right\}_{n=1}^{\infty}$ be an arbitrary, non-increasing sequence such that $\lim_{n\to\infty} x_n = -\infty$. Then, for every $n \in \mathbb{N}$, $x_n \ge x_{n+1}$ and $A_{x_n} \supseteq A_{x_{n+1}}$, and thus

$$\bigcap_{n=1}^{\infty} A_{x_n} = \emptyset, \tag{2.12}$$

which implies

$$\lim_{x\to-\infty} F(x) = \lim_{n\to\infty} F\left(x_n\right) = \lim_{n\to\infty} \mathbb{P}\left(A_{x_n}\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_{x_n}\right) = \mathbb{P}(\emptyset) = 0. \tag{2.13}$$

A similar argument can be made for $\lim_{x\to+\infty} F(x)$.

To show the converse assertion, we need to use $F : \mathbb{R} \to \mathbb{I}$ with the above properties to construct a probability space $\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ on which a mapping $X : \Omega \to \mathbb{R}$ is defined. In order to do this, define the following events:

- $A_x := \left\{X \le x\right\}$ for all $x \in \mathbb{R}$
- $B_{x,y} := A_x \setminus A_y$ for all $x, y \in \mathbb{R}$ such that $y < x$
- $C_{\boldsymbol{x},\boldsymbol{y}} := \bigcup_{\substack{x\in\boldsymbol{x}\\ y\in\boldsymbol{y}}} B_{x,y}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ such that $\boldsymbol{y} < \boldsymbol{x}$ and $B_{x,y} \cap B_{x',y'} = \emptyset$ for all $x \neq x'$ and $y \neq y'$ for all $d \in \mathbb{N}$

Let $\mathcal{F}$ be set of all such $A_x, B_{x,y}$, and $C_{\boldsymbol{x},\boldsymbol{y}}$. Clearly, this set is a σ-algebra. Define $\mathbb{P} : \mathcal{F} \to \mathbb{I}$ by assigning to the events the following values:

$$\begin{aligned}
\mathbb{P}\left(A_x\right) &:= F(x) \\
\mathbb{P}\left(B_{x,y}\right) &:= F(x) - F\left(y\right) \\
\mathbb{P}\left(C_{\boldsymbol{x},\boldsymbol{y}}\right) &:= \sum_{x\in\boldsymbol{x}} F(x) - \sum_{y\in\boldsymbol{y}} F\left(y\right).
\end{aligned}$$

By construction, $\mathbb{P}$ this function satisfies the countable additivity property; i.e. for all countable collections $\left\{E_n\right\}_{n=1}^{\infty}$ of pairwise disjoint events in $\mathcal{F}$, $\mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(E_n\right)$. Using property (c) of F, we see that

$$\begin{aligned}
1 &= \lim_{x\to+\infty} F(x) = \lim_{x\to+\infty} \mathbb{P}\left(A_x\right) = \mathbb{P}\left(\lim_{x\to+\infty} A_x\right) = \mathbb{P}\left(\Omega\right) \\
0 &= \lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \mathbb{P}\left(A_x\right) = \mathbb{P}\left(\lim_{x\to-\infty} A_x\right) = \mathbb{P}(\emptyset),
\end{aligned}$$

and using property (b) of F, it's easy to see that $\mathbb{P}(E) \geq 0$ for all $E \in \mathcal{F}$. Finally, property (a) of F implies that $\mathbb{P}$ is continuous under countable intersections, since for any non-increasing, real sequence $\{x_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} x_n = x$, we have

$$\begin{aligned} \ell^+ F(x) &= \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}(A_{x_n}), \\ F(x) &= \mathbb{P}(A_x) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_{x_n}\right), \end{aligned}$$

and since F is right-continuous, $\ell^+ F(x) = F(x)$, and thus $\bigcap_{n=1}^{\infty} A_{x_n} = \lim_{n\to\infty} \mathbb{P}(A_{x_n})$. All this together implies that $\mathbb{P}$ is a probability measure, and we have thus specified a probability space on which X lives in terms of F. ■

The concept of a distribution function naturally extends to the d-dimensional case,

$$H(\boldsymbol{x}) := \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}), \tag{2.14}$$

where $\boldsymbol{X}$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, $\boldsymbol{x} \in \mathbb{R}^d$, and $H : \mathbb{R}^d \to \mathbb{I}$. Likewise, Theorem 2.2 can be generalised to d dimensions, as shown in e.g. [Billingsley, 1995]. A d-dimensional distribution function has the nice property that it's variation is bounded by the variation of its univariate margins, as stated in the below lemma:

Lemma 2.3. *Let $H : \mathbb{R}^d \to \mathbb{I}$ be a d-dimensional distribution function with marginals $F_1, \ldots, F_d$. Then, for every pair of points $vx, \boldsymbol{y} \in \mathbb{R}^d$,*

$$\left| H(\boldsymbol{x}) - H(\boldsymbol{y}) \right| \leq \sum_{j=1}^{d} \left| F_j(x_j) - F_j(y_j) \right|. \tag{2.15}$$

Proof. Since H is a distribution function for some random variable $\boldsymbol{X}$, we have that for every $j \in \{1, \ldots, d\}$, every $t, t' \in \mathbb{R}$ such that $t < t'$, and every $\boldsymbol{x} \in \mathbb{R}^d$,

$$\begin{aligned} H(\boldsymbol{x}_j(t')) - H(\boldsymbol{x}_j(t)) &= \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}_j(t')) - \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}_j(t)) \\ &= \mathbb{P}(X_1 \leq x_1, \ldots, t < X_j \leq t', \ldots, X_d \leq x_d) \leq F_j(t') - F_j(t). \end{aligned}$$ ■

Another nice property of distribution functions is that under certain conditions, it is possible to transform any random variable into one that's uniform on $\mathbb{I}$. This result is stated in Theorem 2.6 below and requires the following definition:

Definition 2.4 (Quasi-inverse)**.** Let $F : \mathbb{R} \to \mathbb{I}$ be a distribution function. The *quasi-inverse*, $F^{(-1)} : \mathbb{I} \to \mathbb{R}$, of F is defined as

$$F^{(-1)}(t) := \begin{cases} \inf\{x \in \mathbb{R} : F(x) \geq t\}, & t \in (0, 1] \\ \inf\{x \in \mathbb{R} : F(x) > 0\}, & t = 0 \end{cases} \tag{2.16}$$

The quasi-inverse of a distribution function, F, is also sometimes called a *quantile* or *percentile* function, and it allows us to assign a meaningful notion of an "inverse" function to a distribution function which is not bijective. Note that when F is continuous and strictly increasing, $F^{(-1)}$ coincides with F^{-1}, the standard inverse.

Theorem 2.5. *Let $F : \mathbb{R} \to \mathbb{I}$ be a distribution function with quasi-inverse $F^{(-1)}$. Then*

(a) *$F^{(-1)}$ is non-decreasing, and if F is continuous on $\mathbb{R}$, strictly increasing.*

(b) *$F^{(-1)}$ is left-continuous on $\mathbb{I}$.*

(c) *If $t \in \operatorname{ran} F$, then $F\big(F^{(-1)}(t)\big) = t$, and if F is continuous, then it holds for all $t \in \mathbb{I}$.*

(d) *$F^{(-1)}\big(F(x)\big) \le x$ for all $x \in R$, with equality if F is strictly increasing.*

(e) *For every $t \in \mathbb{I}$ and $x \in \mathbb{R}$, $F(x) \ge t$ if and only if $x \ge F^{(-1)}(t)$.*

Proof. (a): Let $A_t = \{x \in \mathbb{R} : F(x) \ge t\}$ and $t_1, t_2 \in \mathbb{I}$ such that $t_1 < t_2$. Then $A_{t_2} \subseteq A_{t_1}$, making $\inf A_{t_1} \le \inf A_{t_2}$ which by Definition 2.4 makes $F^{(-1)}(t_1) \le F^{(-1)}(t_2)$, i.e. non-decreasing. Now assume furthermore that F is continuous in some $x \in \mathbb{R}$ and that $F^{(-1)}(t_1) = F^{(-1)}(t_2) = x$. Then for every $\epsilon > 0$, we have that

$$F(x-\epsilon) < t_1 < t_2 \le F(x+\epsilon), \tag{2.17}$$

but since this means that $\ell^- F(x) < \ell^+ F(x)$, which is a contradiction due to the continuity of F, $F^{(-1)}$ must hence be strictly increasing.

(b): Let $t_0 \in (0,1]$ and $(t_n)_{n\in\mathbb{N}}$ an arbitrary, monotonically increasing sequence with $t_n \to t_0$ as $n \to \infty$. Let $y_0 = F^{(-1)}(t_0)$ and $y_n = F^{(-1)}(t_n)$ for all $n \in \mathbb{N}$. We have from (a) that y_n is a non-decreasing sequence bounded above by y_0. Suppose that its limit $y = \lim_{n\to\infty} y_n < y_0$. Then, for every $\epsilon > 0$, Definition 2.4 gives us that

$$F(y_n - \epsilon) < t_n \le F(y_n + \epsilon) \tag{2.18}$$

But if we choose $\epsilon < (y_0 - y_n)/2$, then $y_n + \epsilon < y_0 - \epsilon$, which implies that

$$t_0 = \lim_{n\to\infty} t_n \le \lim_{n\to\infty} F(y_n + \epsilon) = F(y+\epsilon) \le F(y_0 - \epsilon) < t_0, \tag{2.19}$$

which is a contradiction, and hence $y = y_0$ and $F^{(-1)}$ is left-continuous.

(c): Since we have that $t \in \operatorname{ran} F$, $\exists y$ such that $F(y) = t$. This y is not necessarily unique, so let $\tilde{y} = \inf\Big\{y \in \mathbb{R} : F(y) = t\Big\}$. From Theorem 2.2 we have that F is right-continuous, so $F(\tilde{y}) = t$, and we have

$$F\big(F^{(-1)}(t)\big) = F\Big(\inf\{y \in \mathbb{R} : F(y) \ge t\}\Big) = F(\tilde{y}) = t. \tag{2.20}$$

(d): For every $x \in \mathbb{R}$,

$$F^{(-1)}\big(F(x)\big) = \inf\{y \in \mathbb{R} : F(y) \ge F(x)\} \le x, \tag{2.21}$$

and if one further assumes that F is strictly increasing, then there is no $\tilde{x} < x$ such that $F(\tilde{x}) \ge F(x)$, and thus $F^{(-1)}\big(F(x)\big) = x$.

(e): If $F(x) \ge t$, then $x \ge F^{(-1)}(t)$ by Definition 2.4. Conversely, if $x \ge F^{(-1)}(t)$, then it follows from the right-continuity of F that $F\big(F^{(-1)}(t)\big) \ge t$. ■

With those properties, the following important result can be shown:

Theorem 2.6 (Probability integral transformation)**.** *Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with the distribution function F. Then,*

(a) If F is continuous, then $F \circ X \sim \mathrm{Unif}(0,1)$.

(b) If $U \sim \mathrm{Unif}(0,1)$ then $F^{(-1)} \circ U \sim F$.

Proof. Theorem 2.6 is proven differently in [Durante and Sempi, 2016], but the continuity of F allows a very simply proof of (a): by direct calculation, for all $u \in \operatorname{ran} F = \mathbb{I}$, we have

$$\begin{aligned} F_{F\circ X}(u) &= \mathbb{P}(F \circ X \le u) \\ &= \mathbb{P}\Big(F\big(X\big) \le u\Big) \\ &= \mathbb{P}\Big(X \le F^{(-1)}(u)\Big), \end{aligned}$$

due to F being non-decreasing and continuous, and

$$\begin{aligned} &= F\Big(F^{(-1)}(u)\Big) \\ &= u, \end{aligned}$$

which exactly characterises a uniformly distributed variable on $\mathbb{I}$.

(b): For every $t \in \mathbb{I}$, part (e) of Theorem 2.2 gives us that

$$\mathbb{P}\big(F^{(-1)} \circ U \le t\big) = \mathbb{P}\big(U \le F(t)\big) = F(t). \tag{2.22}$$

■

In other words, transforming a (continuous) random variable by its distribution function will always result in a standard uniform variable. This theorem is an important result, not just in the theory of copulas, which deals with modelling of uniform variables, but also in simulation in general, as it describes a way to generate arbitrarily distributed variables, given uniformly distributed ones, as long as the distribution function is known.

Definition 2.7 (H-volume)**.** Let A be a d-box in $\mathbb{R}^d$. For a function $H : A \to \mathbb{R}$, the *H-volume*, V_H, of the d-box $(\boldsymbol{a}, \boldsymbol{b}] \subseteq A$ is defined by

$$V_H\Big(\big(\boldsymbol{a},\boldsymbol{b}\big]\Big) := \sum_{\boldsymbol{v} \in \mathrm{ver}(\boldsymbol{a},\boldsymbol{b}]} \mathrm{sign}(\boldsymbol{v})\, H(\boldsymbol{v}), \tag{2.23}$$

where

$$\mathrm{sign}(\boldsymbol{v}) := \begin{cases} 1, & \text{if } v_j = a_j \text{ for en even number of indices,} \\ -1, & \text{otherwise,} \end{cases} \tag{2.24}$$

and $\mathrm{ver}\Big(\big(\boldsymbol{a},\boldsymbol{b}\big]\Big) = \{a_1, b_1\} \times \{a_2, b_2\} \times \cdots \times \{a_d, b_d\}$ is the set of vertices of $(\boldsymbol{a}, \boldsymbol{b}]$.

Another way to consider this quantity is (under certain circumstances, see the remark after Lemma 2.9) as

$$V_H\big(A\big) = \int_A \mathrm{d}H(\boldsymbol{v}), \quad \boldsymbol{v} \in A, \tag{2.25}$$

which is just simplified in Definition 2.7 due to A being specifically a d-box. In particular, if H is the distribution function of a random vector $\boldsymbol{X}$, then $V_H(A) = \mathbb{P}(\boldsymbol{X} \in A)$.

An example of this concept is shown for $d = 2$ on Figure 2.1, where the 2-box $(a_1, b_1] \times (a_2, b_2]$ is drawn with bold lines. Here, the H-volume of the box is calculated by adding the values of H evaluated in the points marked with a + sign and subtract the values of H evaluated in the points marked with a − sign. In other words,

$$V_H\Big((\boldsymbol{a},\boldsymbol{b}]\Big) = H(a_1, b_1) - H(a_1, b_2) - H(a_2, b_1) + H(a_2, b_2). \tag{2.26}$$

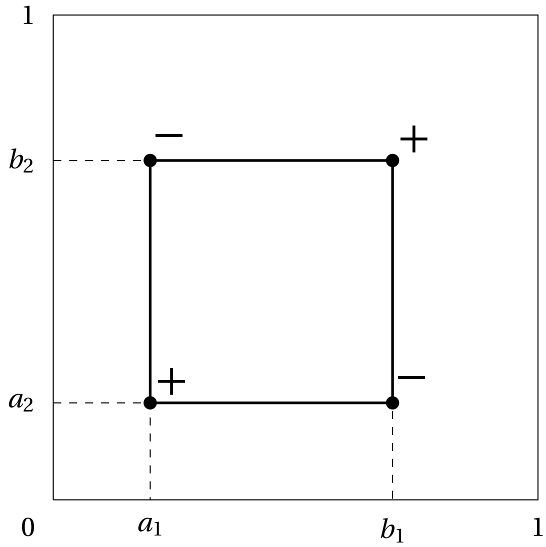


Figure 2.1: *Visualisation of H-volume in two dimensions*

Definition 2.8 (d-increasing)**.** Let A be a d-box in $\mathbb{R}^d$. A function $H : A \to \mathbb{R}$ is said to be *d-increasing* if

$$V_H\Big((\boldsymbol{a},\boldsymbol{b}]\Big) \geq 0, \quad \forall\, (\boldsymbol{a},\boldsymbol{b}] \subseteq A. \tag{2.27}$$

This property can be thought of as a generalisation of the non-decreasing property for univariate functions to multivariate functions, and it is also sometimes known as the Δ-*monotone property* when defined in terms of a finite difference operator (see e.g. [Durante and Sempi, 2016, Remark 1.2.12]).

Later, we will need to show some properties involving H-volumes, where the lemma below can be used to greatly simplify calculations:

Lemma 2.9. *Let $F, G : \mathbb{I}^d \to \mathbb{I}$ be two functions, and let $A = (\boldsymbol{a},\boldsymbol{b}]$ be a d-box in $\mathbb{I}^d$. Then we have*

(a) $V_{F+G}(A) = V_F(A) + V_G(A)$

(b) $V_{\alpha F}(A) = \alpha V_F(A), \forall \alpha > 0$

(c) *If* $A = \bigcup_{j\in\mathcal{J}} A_j$, *where* $\mathcal{J}$ *is a finite index set, and all* A_j *are disjoint left-open* d*-boxes in* $\mathbb{I}^d$, *then*

$$V_F\big(A\big) = \sum_{j\in\mathcal{J}} V_F\Big(A_j\Big). \tag{2.28}$$

Proof. The proof for Lemma 2.9 is not given in [Durante and Sempi, 2016], but it is easy to see that the first two statements (a) and (b) readily follow from Definition 2.7:

$$\begin{aligned} V_{F+G}\big(A\big) &= \sum_{\boldsymbol{v}\in\operatorname{ver} A} \operatorname{sign}(\boldsymbol{v})\big(F(\boldsymbol{v}) + G(\boldsymbol{v})\big) = \sum_{\boldsymbol{v}\in\operatorname{ver} A} \operatorname{sign}(\boldsymbol{v})\, F(\boldsymbol{v}) + \sum_{\boldsymbol{v}\in\operatorname{ver} A} \operatorname{sign}(\boldsymbol{v})\, G(\boldsymbol{v}) \\ &= V_F\big(A\big) + V_G\big(A\big) \end{aligned} \tag{2.29}$$

$$V_{\alpha F}\big(A\big) = \sum_{\boldsymbol{v}\in\operatorname{ver} A} \operatorname{sign}(\boldsymbol{v})\big(\alpha F(\boldsymbol{v})\big) = \alpha \sum_{\boldsymbol{v}\in\operatorname{ver} A} \operatorname{sign}(\boldsymbol{v})\, F(\boldsymbol{v}) = \alpha V_F\big(A\big) \tag{2.30}$$

For statement (c), with inspiration from [Billingsley, 1995], suppose that each side $\big(a_i, b_i\big]$ of A is partitioned into n_i subintervals $I_{i,j} = \Big(t_{i,j-1}, t_{i,j}\Big]$, $j = 1, \ldots, n_i$ such that $a_i = t_{i,0} < t_{i,1} < \cdots < t_{i,n_i} = b_i$. Then, A is partitioned by the $n_1 n_2 \cdots n_d$ d-boxes

$$B_{j_1,\ldots,j_d} = I_{1,j_1} \times \cdots \times I_{d,j_d}, \quad 1 \le j_1 \le n_1, \ldots, 1 \le j_k \le n_k. \tag{2.31}$$

Let P be the set of these d-boxes and let V be the set of all points $\boldsymbol{v}$ that is a vertex of one or more of the boxes in P. Consider the sum of their F-volumes:

$$\sum_{B\in P} \sum_{\boldsymbol{v}\in\operatorname{ver} B} \operatorname{sign}_B(\boldsymbol{v})\, F(\boldsymbol{v}) = \sum_{\boldsymbol{v}\in V} F(\boldsymbol{v}) \sum_{\substack{B\in P:\\ \boldsymbol{v}\in\operatorname{ver} B}} \operatorname{sign}_B(\boldsymbol{v}), \tag{2.32}$$

where sign_B denotes the sign function in the context of the d-box B. Suppose that $\boldsymbol{v}$ is a vertex of one or more $B \in P$, but is not a vertex of A itself. Then there must be an index $i \in \{1, \cdots, d\}$ such that v_i is neither a_i nor b_i. Without loss of generality, assume that $i = 1$. Then $v_1 = t_{1,j}$ with $0 < j < n_1$. The boxes of which $\boldsymbol{v}$ is a vertex come in pairs B', B'', such that $B' = B_{j,j_2,\ldots,j_d}$ and $B'' = B_{j+1,j_2,\ldots,j_d}$ and $\operatorname{sign}_{B'}(\boldsymbol{v}) = -\operatorname{sign}_{B''}(\boldsymbol{v})$, and therefore, the inner sum in (2.32) is 0 whenever $\boldsymbol{v} \notin \operatorname{ver} A$.

Suppose now that $\boldsymbol{v} \in \operatorname{ver} A$ and $\boldsymbol{v} \in \operatorname{ver} B$ for at least one $B \in P$. Then for each index $i \in \{1, \ldots, d\}$, either $v_i = a_i$ or $v_i = b_i$, and $\boldsymbol{v}$ is a vertex of only the $B \in P$ for which $j_i = 1$ or $j_i = n_i$, accordingly. Then we have that $\operatorname{sign}_B(\boldsymbol{v}) = \operatorname{sign}_A(\boldsymbol{v})$, and we can conclude

$$\sum_{B\in P} \sum_{\boldsymbol{v}\in\operatorname{ver} B} \operatorname{sign}_B(\boldsymbol{v})\, F(\boldsymbol{v}) = \sum_{\boldsymbol{v}\in\operatorname{ver} A} \operatorname{sign}_A(\boldsymbol{v})\, F(\boldsymbol{v}) = V_F\big(A\big). \tag{2.33}$$

Note that this only shows that (c) holds for a so-called regular (i.e. grid-like) partition, but it is easy to see that it also holds for an irregular partition, as each B in such a partition can itself be partitioned into a set of d-boxes $\tilde{B}$ such that the set of all $\tilde{B}$'s is a regular partition of A. ■

It turns out that if F is d-increasing and continuous, then V_F corresponds to a unique measure [Billingsley, 1995, Theorem 12.5], which for distribution functions is the probability measure, as noted underneath (2.25).

2.3 The Copula

We now arrive at the definition of a *copula*, the main object of interest in this project. In the literature, there are several different definitions of it, but here we follow [Sklar, 1996] and define it in terms of a related concept:

Definition 2.10 (Subcopula)**.** Let $A_1,\dots,A_d$ be subsets of $\mathbb{I}$ containing both 0 and 1. Then a d-dimensional *subcopula* (or simply d-subcopula) is a function $C' : A_1 \times \cdots \times A_d \to \mathbb{I}$ such that

(a) $C'(\boldsymbol{u}) = C'\left(u_1,\dots,u_d\right) = 0$ if $u_j = 0$ for at least one $j \in \left\{1,\dots,d\right\}$

(b) $C'\left(\mathbf{1}_j\left(u_j\right)\right) = C'\left(1,\dots,1,u_j,1,\dots,1\right) = u_j$ for all $j \in \left\{1,\dots,d\right\}$

(c) $V_{C'}\left(\left(\boldsymbol{a},\boldsymbol{b}\right]\right) \geq 0$ for every d-box $\left(\boldsymbol{a},\boldsymbol{b}\right] \subseteq A_1 \times \cdots \times A_d$

Definition 2.11 (Copula)**.** A *copula*, C, is a subcopula whose domain is the entire unit d-box, $\mathbb{I}^d$. The space of d-copulas is denoted $\mathscr{C}_d$.

By this definition, a d-copula is a d-dimensional distribution function on $\mathbb{I}^d$ with uniform marginals on $\mathbb{I}$; for a copula, C, property (a) and (b) of Definition 2.10 are called the *boundary conditions*, where by (a) it is said to be *grounded* and by (b) the functions $C_j : \mathbb{I} \to \mathbb{I}$, obtained by setting all arguments except for the j'th to 1, correspond to distribution functions for onedimensional, uniform variables. Property (c) corresponds to a copula being d-increasing, and if the function $c(\boldsymbol{u}) = \partial^d C(\boldsymbol{u}) \,/\, \left(\partial u_1 \dots \partial u_d\right)$ exists, then

$$C(\boldsymbol{u}) = \int_{[0,\boldsymbol{u}]} c(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}, \tag{2.34}$$

and property (c) is equivalent to

$$\int_{(\boldsymbol{a},\boldsymbol{b}]} c(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \geq 0. \tag{2.35}$$

The function c is called the *copula density*.

Some fundamental examples of copulas are given below.

Example 2.12 (The comonotonicity copula)**.** Let $U \sim \mathrm{Unif}\left(0,1\right)$ and consider the random vector consisting of d copies of U,

$$\boldsymbol{U} := \Big(\underbrace{U,\dots,U}_{d\text{ times}}\Big). \tag{2.36}$$

Then, for every $\boldsymbol{u} \in \mathbb{I}^d$,

$$\mathbb{P}\left(\boldsymbol{U} \leq \boldsymbol{u}\right) = \mathbb{P}\left(U \leq \min\left\{u_1,\dots,u_d\right\}\right) = \min\left\{u_1,\dots,u_d\right\}, \tag{2.37}$$

and the distribution function $M_d : \mathbb{I}^d \to \mathbb{I}$ defined by

$$M_d := \min\left\{u_1,\dots,u_d\right\}, \tag{2.38}$$

is a copula in $\mathscr{C}_d$ called the *comonotonicity copula*.

Example 2.13 (The independence copula)**.** Let $U_1,\dots,U_d$ be i.i.d Unif$(0,1)$ and consider the random vector $\boldsymbol{U} = (U_1,\dots,U_d)$. Then, for every $\boldsymbol{u} \in \mathbb{I}^d$,

$$\mathbb{P}(\boldsymbol{U} \leq \boldsymbol{u}) = \mathbb{P}(U_1 \leq u_1)\cdots\mathbb{P}(U_d \leq u_d) = \prod_{j=1}^{d} u_j, \tag{2.39}$$

and the distribution function $\Pi_d : \mathbb{I}^d \to \mathbb{I}$ defined by

$$\Pi_d(\boldsymbol{u}) := \prod_{j=1}^{d} u_j \tag{2.40}$$

is a copula in $\mathscr{C}_d$ called the *independence copula.*

Example 2.14 (The countermonotonicity copula)**.** Let $U \sim$ Unif$(0,1)$ and consider the random vector $\boldsymbol{U} = (U, 1-U)$. Then for every $\boldsymbol{u} \in \mathbb{I} \times \mathbb{I}$,

$$\mathbb{P}(\boldsymbol{U} \leq \boldsymbol{u}) = \mathbb{P}(U_1 \leq u_1, 1-U_1 \leq u_2) = \max\{0, u_1 + u_2 - 1\}, \tag{2.41}$$

and the distribution function $W_2 : \mathbb{I} \times \mathbb{I} \to \mathbb{I}$ defined by

$$W_2(\boldsymbol{u}) := \max\{0, u_1 + u_2 - 1\} \tag{2.42}$$

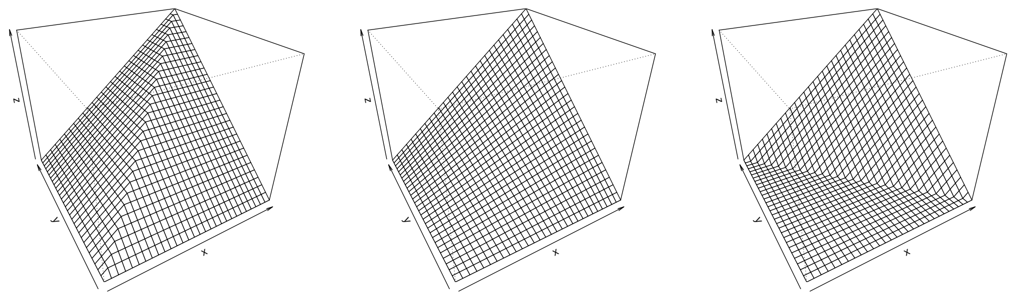
is a copula in $\mathscr{C}_2$ called the *countermonotonicity copula.*

Notice that W_2 is explicitly only defined as a 2-copula, not a d-copula. There is an analoguous function, $W_d : \mathbb{I}^d \to \mathbb{I}$ defined by

$$W_d(\boldsymbol{u}) := \max\left\{0, \sum_{j=1}^{d} u_j - (d-1)\right\}, \tag{2.43}$$

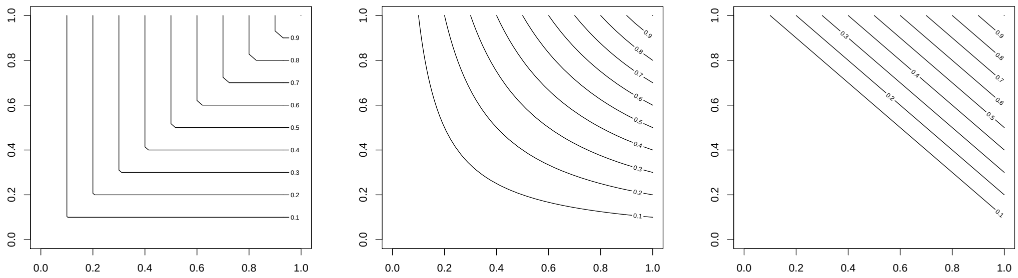
but it is not generally a copula for $d > 2$, as it is not necessarily d-increasing.

Examples for the above copulas in $d = 2$ are visualised as 3D plots on Figure 2.2 and as contour plots on Figure 2.3.



(a) M_2 ***(b)*** Π_2 ***(c)*** W_2

Figure 2.2: *3D plots of the 2-copulas M_2, Π_2, and W_2*



(a) M_2 ***(b)*** Π_2 ***(c)*** W_2

Figure 2.3: *Contour plots of the 2-copulas* M_2, Π_2, *and* W_2

2.3.1 Properties of Copulas

Here follows some properties of copulas that either we need in order to prove the main result in the next section, or that are interesting in and of themselves.

Theorem 2.15 (Fréchet-Hoeffding bounds)**.** *Let* W_d *and* M_d *be defined as in* (2.43) *and* (2.38)*, respectively. For every* d*-copula* C *and every point* $\boldsymbol{u} \in \mathbb{I}^d$*,*

$$W_d(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M_d(\boldsymbol{u})\,. \tag{2.44}$$

Proof. Since C is a distribution function with univariate margins on $\mathbb{I}$, we have for every $\boldsymbol{u} \in \mathbb{I}^d$ that

$$\bigcap_{k=1}^{d} \{U_k \leq u_k\} \subseteq \Big\{U_j \leq u_j\Big\}, \quad j = 1, \ldots, d, \tag{2.45}$$

implying that

$$C(\boldsymbol{u}) = \mathbb{P}\left(\bigcap_{k=1}^{d} \{U_k \leq u_k\}\right) \leq \min_{j \in \{1,\ldots,d\}} \mathbb{P}\Big(U_j \leq u_j\Big) = M_d(\boldsymbol{u})\,, \tag{2.46}$$

and conversely, we have

$$\begin{aligned} C(\boldsymbol{u}) &= \mathbb{P}\left(\bigcap_{j=1}^{d} \Big\{U_j \leq u_j\Big\}\right) = 1 - \mathbb{P}\left(\bigcup_{j=1}^{d} \Big\{U_j > u_j\Big\}\right) \\ &\geq 1 - \sum_{j=1}^{d} \mathbb{P}\Big(U_j > u_j\Big) = 1 - \sum_{j=1}^{d} \Big(1 - u_j\Big) = \sum_{j=1}^{d} u_j - (d-1). \end{aligned}$$

■

The functions W_d and M_d are referred to as the *lower* and *upper Fréchet-Hoeffding bound*, respectively. Visually, one can think of Theorem 2.15 as stating that the surface of any copula lies between those shown (in 2 dimensions) on Figure 2.2c and Figure 2.2a.

Theorem 2.16. *The set of copulas,* $\mathscr{C}_d$ *is a convex set, i.e.* $\forall \alpha \in \mathbb{I}$ *and* $C_0, C_1 \in \mathscr{C}_d$*,*

$$C = \alpha C_0 + (1-\alpha) C_1 \in \mathscr{C}_d \tag{2.47}$$

Proof. This proof is largely skipped in [Durante and Sempi, 2016], but it follows by direct application of Lemma 2.9. Since $C_0, C_1 \in \mathscr{C}_d$, their domain (and therefore C's domain) is $\mathbb{I}^d$, so to prove that C is a d-copula, we show that it satisfies each condition of Definition 2.10:

(a): Let $\boldsymbol{u} \in \mathbb{I}^d$ have $u_j = 0$ for at least one index $j \in \{1, \ldots, d\}$. Then we have

$$\begin{aligned} C(\boldsymbol{u}) &= \alpha C_0(\boldsymbol{u}) + (1-\alpha) C_1(\boldsymbol{u}) \\ &= \alpha \cdot 0 + (1-\alpha) \cdot 0 = 0, \end{aligned}$$

since $C_0, C_1 \in \mathscr{C}_d$ and thus satisfy the condition themselves.

(b): Let $\mathbf{1}_j$ be a vector of ones, except with u_j in j'th entry. Then we have

$$C\big(\mathbf{1}_j(u_j)\big) = \alpha C_0\big(\mathbf{1}_j(u_j)\big) + (1-\alpha) C_1\big(\mathbf{1}_j(u_j)\big)$$

and since $C_0, C_1 \in \mathscr{C}_d$, they have uniform marginals, and thus

$$\begin{aligned} &= \alpha u_j + (1-\alpha) u_j \\ &= u_j. \end{aligned}$$

(c): By Lemma 2.9, we have for every d-box $A = (\boldsymbol{a}, \boldsymbol{b}] \subseteq \mathbb{I}^d$ that

$$V_C(A) = V_{\alpha C_0 + (1-\alpha) C_1}(A),$$

then property (a) and (b) of the lemma gives us that

$$= \alpha V_{C_0}(A) + (1-\alpha) V_{C_1}(A),$$

and since $C_0, C_1 \in \mathscr{C}_d$ and are thus d-increasing, we have

$$\geq 0,$$

which makes C d-increasing as well, and thus a copula. ∎

Theorem 2.16 ensures convex combinations of existing copulas also are copulas, which may be useful for constructing new forms of copulas. An example of such copulas are *Fréchet copulas*, defined for $d = 2$ as

$$C^{\mathbf{Fre}}_{\alpha,\beta}(u_1, u_2) := \alpha M_2(u_1, u_2) + (1 - \alpha - \beta)\Pi_2(u_1, u_2) + \beta W_2(u_1, u_2). \tag{2.48}$$

Theorem 2.17. *Let C be a d-copula. Then for all $\mathbf{u}, \mathbf{v} \in \mathbb{I}^d$ the following inequality holds:*

$$\big|C(\boldsymbol{u}) - C(\boldsymbol{v})\big| \leq \sum_{j=1}^{d} \big|u_j - v_j\big|. \tag{2.49}$$

Proof. The proof follows from Lemma 2.3 since C is a distribution function and its marginals are univariate on $\mathbb{I}$. ∎

The inequality (2.49) is equivalent with saying that C is Lipschitz continuous with constant 1, since it can be written as

$$\left|C(\boldsymbol{u}) - C(\boldsymbol{v})\right| \leq \left\|\boldsymbol{u} - \boldsymbol{v}\right\|_1, \tag{2.50}$$

where $\|\cdot\|_1$ is the ℓ^1-norm on $\mathbb{R}^d$.

We will also need certain metric properties of copulas for the proof in section 2.4. For this, let $\left(\mathbf{C}\left(\mathbb{I}^d\right), \delta_\infty\right)$ be the space of continuous, real-valued functions on $\mathbb{I}^d$ with the distance measure $\delta_\infty : \mathbf{C}\left(\mathbb{I}^d\right) \times \mathbf{C}\left(\mathbb{I}^d\right) \to \mathbb{R}$ defined by

$$\delta_\infty\left(f_1, f_2\right) = \left\|f_1 - f_2\right\|_\infty = \sup_{\boldsymbol{u}\in\mathbb{I}^d}\left|f_1(\boldsymbol{u}) - f_2(\boldsymbol{u})\right| = \max_{\boldsymbol{u}\in\mathbb{I}^d}\left|f_1(\boldsymbol{u}) - f_2(\boldsymbol{u})\right|. \tag{2.51}$$

Theorem 2.18. *If a sequence $\left\{C_n\right\}_{n=1}^\infty \subseteq \mathscr{C}_d$ converges pointwise to C, i.e.* $\lim_{n\to\infty} C_n(\boldsymbol{u}) = C(\boldsymbol{u}), \forall \boldsymbol{u} \in \mathbb{I}^d$*, then C is a copula.*

Proof. It's easy to see that C has uniform marginals on $\mathbb{I}$. To see that it is also d-increasing, note that for all d-boxes $A \subseteq \mathbb{I}^d$, $V_C\left(A\right)$ can be expressed as the pointwise limit of $\left\{V_{C_n}\left(A\right)\right\}_{n=1}^\infty$, and hence

$$V_C\left(A\right) = \lim_{n\to\infty} V_{C_n}\left(A\right) \geq 0, \tag{2.52}$$

which makes C a copula. ■

Theorem 2.19. *The space of d-copulas, $\mathscr{C}_d$, is a compact subspace in $\left(\mathbf{C}\left(\mathbb{I}^d\right), d_\infty\right)$.*

Proof. Because $\left(\mathbf{C}\left(\mathbb{I}^d\right), \delta_\infty\right)$ is complete and, by Theorem 2.18, $\mathscr{C}_d$ is closed in $\mathbf{C}\left(\mathbb{I}^d\right)$, $\mathscr{C}_d$ is also complete. Since

$$\sup_{\substack{\boldsymbol{u}\in\mathbb{I}^d,\\ C\in\mathscr{C}_d}} |C(\boldsymbol{u})| \leq 1, \tag{2.53}$$

$\mathscr{C}_d$ is uniformly bounded, and since every copula is 1-Lipschitz by Theorem 2.17, it is equicontinuous. As a result of this, the Arzelà-Ascoli Theorem [Arzelà, 1895] makes $\mathscr{C}_d$ totally bounded w.r.t δ_∞, which together with its completeness makes it compact. ■

We now have all the definitions and properties we need in order to state and prove the main result, which is done in detail in the following section.

2.4 Sklar's Theorem

The main result of copula theory is *Sklar's Theorem* from [Sklar, 1959], which states that *any* multivariate distribution function can be represented as a composition of its univariate margins and a copula. This fact makes copulas a very flexible tool for statistical analysis, as one need not specify a complete joint model for all variables at once, but rather, one can model each (univariate) variable individually and then model their dependence with a copula after the fact, which we will make heavy use of in the applied part of this project.

The contents of the theorem is given in the following.

Theorem 2.20 (Sklar's Theorem)**.** *Let $\boldsymbol{X} \in \mathbb{R}^d$ be a d-dimensional random vector on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the joint distribution function $H(\boldsymbol{x}) = \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d)$, and let $F_j\left(x_j\right) = \mathbb{P}\left(X_j \le x_j\right)$ for $j = 1, \ldots, d$ be its marginals. Then there exists a d-copula C such that $\forall \boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$,*

$$H(\boldsymbol{x}) = C\Big(F_1(x_1), \ldots, F_d(x_d)\Big). \tag{2.54}$$

Furthermore, if the marginals $F_1, \ldots, F_d$ are continuous, the copula C is uniquely defined.

Corollary 2.21. *Under the same assumptions as in Theorem 2.20, assume furthermore that the joint distribution of $\boldsymbol{X}$ has the density function h and the marginal distribution of X_j has the density f_j for $j = 1, \ldots, d$. Then the full joint density h can be expressed as*

$$h(\boldsymbol{x}) = \frac{\partial^d H(\boldsymbol{x})}{\partial x_1 \ldots \partial x_d} = \frac{\partial^d C(F_1(x_1), \ldots, F_d(x_d))}{\partial x_1 \ldots \partial x_d} = c\Big(F_1(x_1), \ldots, F_d(x_d)\Big) f_1(x_1) \cdots f_d(x_d), \tag{2.55}$$

and the conditional density $h_{\mathcal{J}|-\mathcal{J}}$ of $(x_j : j \in \mathcal{J})$ conditioned on $(x_k : k \notin \mathcal{J})$ as

$$h_{\mathcal{J}|-\mathcal{J}}(\boldsymbol{x}_{\mathcal{J}} | \boldsymbol{x}_{-\mathcal{J}}) = c\Big(F_1(x_1), \ldots, F_d(x_d)\Big) \prod_{j \in \mathcal{J}} f_j(x_j), \tag{2.56}$$

where c is the copula density, $\mathcal{J} \subseteq \{1, \ldots, d\}$ is a subset of indices.

In the above corollary, (2.55) follows from the chain rule of probability, and (2.56) follows from (2.55) by simply dividing by the densities of the marginals conditioned on.

In the case of continuous margins, the proof of Theorem 2.20 follows directly from Theorem 2.6, and since this result has practically useful consequences, it is stated as a lemma below.

Lemma 2.22 (Sklar's Theorem for continuous margins)**.** *Under the same assumptions as in Theorem 2.20, assume furthermore that the margins $F_1, \ldots, F_d$ are continuous. Then, there exists a unique copula C associated with $\boldsymbol{X}$ that is the distribution function of the random vector $(F_1 \circ X_1, \ldots, F_d \circ X_d)$, and for every $\boldsymbol{u} \in \mathbb{I}^d$ it is specified by*

$$C(\boldsymbol{u}) = H\Big(F_1^{(-1)}(u_1), \ldots, F_d^{(-1)}(u_1)\Big). \tag{2.57}$$

Proof. By Theorem 2.6, since F_j is continuous, then $F_j \circ X_j \sim \text{Unif}(0,1)$, $j = 1, \ldots, d$. Hence, the random vector $(F_1 \circ X_1, \ldots, F_d \circ X_d)$ has uniform univariate margins and its distribution function is thus a copula, and for every $\boldsymbol{x} \in \mathbb{R}^d$, we have

$$\begin{aligned} H(\boldsymbol{x}) &= \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d) \\ &= \mathbb{P}\Big(F_1(X_1) \le F_1(x_1), \ldots, F_d(X_d) \le F_d(x_d)\Big) \\ &= C\Big(F_1(x_1), \ldots, F_d(x_d)\Big). \end{aligned}$$

■

Lemma 2.22 provides a method for constructing a copula, when you know the joint distribution. This method of constructing copulas is called the *inversion method*, and it will be used later in section 2.6 to define certain types of copulas.

When one or more margins are not continuous, the copula is not uniquely defined on all of $\mathbb{I}^d$ and one needs a suitable constraint for referring to a single copula as "the" copula of $\boldsymbol{X}$. In the following section, a proof of Theorem 2.20 that allows for margins with discontinuities is presented.

2.4.1 Proof of Sklar's Theorem

Because of both its simplicity and its relevance in the field of statistical analysis, Theorem 2.20 has been rediscovered in the literature by many different authors, and as such, several proofs exist that base their argument in different properties of copulas.

The theorem was first stated without proof in [Sklar, 1959] and later proved in detail for $d = 2$ in [Schweizer and Sklar, 1974] and for the general case in [Sklar, 1996]. Both proofs involves an extension argument, in which one shows that a similar result holds for a subcopula and that such a subcopula can be extended to a copula. This strategy was also used in [Carley and Taylor, 2003], who showed that the extension amounts to a multiliniar interpolation of the subcopula.

A proof based on probabilistic arguments was given in [Moore and Spruill, 1975], who used a generalised version of the probability integral transform to extend the result to margins with possible discontinuities in their distribution functions.

[Durante et al., 2012] gave a proof showing the existence of a copula, but not its form, in which the authors approximated a distribution function H by a sequence of such functions, $\{H_n\}_{n=1}^{\infty}$ and used it to construct a copula associated with H via a sequence of copulas $\{C_n\}_{n=1}^{\infty}$, each associated with H_n.

In this project, we follow the method of [Sklar, 1996] and present a proof by extension, divided into the following lemmas:

Lemma 2.23. *For every d-dimensional distribution function H with marginals $F_1, \dots, F_d$ there exists a unique subcopula, $C' : \operatorname{ran} F_1 \times \cdots \times \operatorname{ran} F_d \to \mathbb{I}$, such that for all $\boldsymbol{x} \in \mathbb{R}^d$,*

$$H(\boldsymbol{x}) = C'\Big(F_1(x_1), \dots, F_d(x_d)\Big), \tag{2.58}$$

and it is given by

$$C'(\boldsymbol{u}) = H\Big(F_1^{(-1)}(u_1), \dots, F_d^{(-1)}(u_d)\Big), \tag{2.59}$$

for all $\boldsymbol{u} \in \operatorname{ran} F_1 \times \cdots \times F_d$.

Proof. For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, Lemma 2.3 implies that of $F_j(x_j) = F_j(y_j)$ for all $j = 1, \dots, d$, then $H(\boldsymbol{x}) = H(\boldsymbol{y})$. This means that for all $\boldsymbol{x} \in \mathbb{R}^d$, the value of $H(\boldsymbol{x})$ only depends on the numbers $F_j(x_j)$, $j = 1, \dots, d$, which implies that there exists a unique function $C' : \operatorname{ran} F_1 \times \cdots \times \operatorname{ran} F_d \to \mathbb{I}$ that satisfies (2.54). The properties of H directly imply that C' is a subcopula; (a) and (b) of Definition 2.10 follow by direct calculation, and (c) follows from the fact that H is d-increasing.

Next, for every $u_j \in \operatorname{ran} F_j$, $j = 1, \dots, d$, we have from Theorem 2.5 that $F_j \circ F_j^{(-1)}(u_j) = u_j$, so that

$$H\Big(F_1^{(-1)}(u_1), \dots, F_d^{(-1)}(u_d)\Big) = C'\Big(F_1 \circ F_1^{(-1)}(u_1), \dots, F_d \circ F_d^{(-1)}(u_d)\Big) = C'(\boldsymbol{u}), \tag{2.60}$$

for all $\boldsymbol{u} \in \operatorname{ran} F_1 \times \cdots \times \operatorname{ran} F_d$. ■

Lemma 2.24. *For every supcopula, $C' : A_1 \times \cdots A_d \to \mathbb{I}$, where $A_j \subseteq \mathbb{I}$ for $j = 1, \dots, d$, there exists a copula, C, that extends it, i.e. for all $\boldsymbol{u} \in A_1 \times \cdots \times A_d$,*

$$C(\boldsymbol{u}) = C'(\boldsymbol{u}). \tag{2.61}$$

Proof. Since every subcopula is uniformly continuous on its domain by Lemma 2.3, it is possible to extend C' to a function $C'' : \overline{A}_1 \times \cdots \times \overline{A}_d \to \mathbb{I}$, where $\overline{A}_j$ denotes the closure of A_j. For each of these $\overline{A}_j$, one can find a sequence of finite sets, $\{A_{j,n}\}_{n\in\mathbb{N}}$, such that $A_{j,1} \subseteq A_{j,2} \subseteq \cdots \subseteq \overline{A}_j$, with $0, 1 \in A_{j,n}$ for all $n \in \mathbb{N}$ and

$$\overline{\bigcup_{n\in\mathbb{N}} A_{j,n}} = \overline{A}_j. \tag{2.62}$$

Define $S_n : A_{1,n} \times \cdots \times A_{d,n} \to \mathbb{I}$ by $S_n(\boldsymbol{u}) := C''(\boldsymbol{u})$. For all $n \in \mathbb{N}$, S_n is a subcopula, since it is the restriction of a subcopula, and has $\mathbf{0}$ and $\mathbf{1}$ in its domain, and thus we now have a sequence $\{S_n\}_{n\in\mathbb{N}}$ such that for all $n \in \mathbb{N}$,

- $\operatorname{dom} S_n$ is finite,
- S_n is a restriction of C'',
- $\operatorname{dom} S_n \subseteq \operatorname{dom} S_{k+1}$,
- the countable union of $\operatorname{dom} S_n$, $n \in \mathbb{N}$ is a countable, dense subset of $\operatorname{dom} C''$.

For all $n \in \mathbb{N}$, construct a copula C_n by extending the domain of each S_n to $\mathbb{I}^d$; in [Durante and Sempi, 2016], this extension is only demonstrated for $d = 2$, but [Sklar, 1996] showed how it can be done in general dimensions as follows: If a point $\boldsymbol{x} \notin \operatorname{dom} C''$, then $\boldsymbol{x}$ lies in a unique d-box B such that $\operatorname{ver} B \subseteq \operatorname{dom} C''$, and it contains no smaller such d-box. Then $\boldsymbol{x}$ can be uniquely represented as

$$\boldsymbol{x} = \sum_{\boldsymbol{v}\in\operatorname{ver} B} \beta(\boldsymbol{v})\, \boldsymbol{v}, \tag{2.63}$$

where $\beta(\boldsymbol{v})$ is a non-negative number such that $\sum_{\boldsymbol{v}\in\operatorname{ver} B} \beta(\boldsymbol{v}) = 1$. Now, define each $C_n(\boldsymbol{x})$ by

$$C_n(\boldsymbol{x}) = \sum_{\boldsymbol{v}\in\operatorname{ver} B} \beta(\boldsymbol{v})\, S_n(\boldsymbol{v}), \tag{2.64}$$

and set $C_n(\boldsymbol{x}) = S_n(\boldsymbol{x})$ for all $\boldsymbol{x} \in \operatorname{dom} S_n$.

Since $\mathscr{C}_d$ is compact by Theorem 2.19, there exists a subsequence $\{C_{n(k)}\}_{k\in\mathbb{N}}$ of $\{C_n\}_{n\in\mathbb{N}}$ that converges to a copula C cf. Theorem 2.18. It follows that $C(\boldsymbol{x}) = C''(\boldsymbol{x})$ at every point of

$$\bigcup_{n\in\mathbb{N}} \operatorname{dom} S_n \subseteq \overline{\bigcup_{n\in\mathbb{N}} \operatorname{dom} S_n} = \operatorname{dom} C'', \tag{2.65}$$

which concludes the proof. ■

Given Lemma 2.23 and 2.24, Theorem 2.20 can then be proved as follows:

Proof of Sklar's Theorem. Let $H : \mathbb{R}^d \to \mathbb{I}$ be an arbitrary d-dimensional distribution function. By Lemma 2.23, there exists a unique subcopula C' such that for every $\boldsymbol{x} \in \mathbb{R}^d$,

$$H(\boldsymbol{x}) = C'\Big(F_1(x_1), \ldots, F_d(x_d)\Big), \tag{2.66}$$

and by Lemma 2.24, this subcopula can be extended to a copula C. For every $\boldsymbol{x} \in \mathbb{R}^d$, $F_j(x_j) \in \operatorname{ran} F_j$, and therefore, (2.54) holds. ∎

2.5 Measures of Association

When selecting copulas for modelling real data, where in general, the true copula is seldom known, we of couse wish to select one that closely captures the joint behaviour of the data. In this section, we introduce some *measures of association*, which can be used to describe aspects of such behaviour, and which differ from copula to copula.

Definition 2.25 (Pearson's correlation coefficient)**.** For two random variables $X, Y \in \mathbb{R}$ whose second moments exist, *Pearson's correlation coefficient* (named after Karl Pearson) is defined as

$$r(X,Y) := \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)\Big)}{\sqrt{\mathbb{E}(X^2) - \mathbb{E}(X)^2}\sqrt{\mathbb{E}(Y^2) - \mathbb{E}(Y)^2}}, \tag{2.67}$$

where $\mathbb{E}$ is the expectation operator, $\operatorname{cov}(X,Y)$ denotes the covariance of X and Y, and σ_X and σ_Y are the standard deviations of X and Y, respectively.

For a sample, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, the coefficient can be calculated as

$$\hat{r}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2}\sqrt{\sum_{j=1}^n (y_j - \bar{y})^2}}, \tag{2.68}$$

where $\bar{x} = \frac{1}{n}\sum_{j=1}^n x_j$ is the sample mean, and likewise for $\bar{y}$.

Pearson's correlation coefficient is also commonly known as *linear correlation*, as it is invariant under linear transformations of the variables, but not in general under strictly increasing transformations. While this measure is commonly in e.g. linear normal models, it's not commonly used in copula modelling due to this missing property. The below two other measures are used instead.

Definition 2.26 (Spearman's rank correlation coefficient)**.** For two random variables $X, Y \in \mathbb{R}$ whose copula is given by C, *Spearman's rank correlation coefficient* (or *Spearman's rho*, named after Charles Spearman) is defined as

$$\rho(X,Y) := 12\int_{\mathbb{I}^2} C(u,v)\,\mathrm{d}u\,\mathrm{d}v - 3 \tag{2.69}$$

This quantity is related (2.67), since because $\mathbb{E}(F_X \circ X) = \mathbb{E}(F_Y \circ Y) = \frac{1}{2}$ (where F_X and F_Y are the distribution functions for X and Y, respectively), and $\operatorname{var}(F_X \circ X) = \operatorname{var}(F_Y \circ Y) = \frac{1}{12}$ by the uniformity of $F_X \circ X$ and $F_Y \circ Y$, one can write

$$\rho(X,Y) = \frac{\operatorname{cov}(F_X \circ X, F_Y \circ Y)}{\sqrt{\operatorname{var}(F_X \circ X)\operatorname{var}(F_Y \circ Y)}} = r(F_X \circ X, F_Y \circ Y). \tag{2.70}$$

For a sample, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, the coefficient $\hat{\rho}(\boldsymbol{x}, \boldsymbol{y})$ is the Pearson correlation of the *rank scores* of $\boldsymbol{x}$ and $\boldsymbol{y}$. The rank score of a set of observations, $\operatorname{rg}(\boldsymbol{x})$, is an assignment of the numbers 1 through n to the observations, such that $\operatorname{rg}(\boldsymbol{x})_i > \operatorname{rg}(\boldsymbol{x})_j$ if $x_i > x_j$, and thus

$$\hat{\rho}(\boldsymbol{x}, \boldsymbol{y}) = \hat{r}(\operatorname{rg}(\boldsymbol{x}), \operatorname{rg}(\boldsymbol{y})). \tag{2.71}$$

Definition 2.27 (Kendall's tau)**.** For two random variables $X, Y \in \mathbb{R}$ whose copula is given by C, *Kendall's tau* (named after Maurice G. Kendall) is defined as

$$\tau(X,Y) := 4\int_{\mathbb{I}^2} C(u,v)\,\mathrm{d}C(u,v) - 1. \tag{2.72}$$

The above expression is the value of the difference in probability between *concordance* and *discordance*, and can be written as

$$\rho(X,Y) = \mathbb{P}\big(\underbrace{(X_1 - X_2)(Y_1 - Y_2) > 0}_{\text{concordance}}\big) - \mathbb{P}\big(\underbrace{(X_1 - X_2)(Y_1 - Y_2) < 0}_{\text{discordance}}\big), \tag{2.73}$$

where (X_1, Y_1) and (X_2, Y_2) are two independent pairs drawn from the joint distribution of X and Y. In other words, the pairs are said to be concordant if the sort order of X and Y agree.

For a sample, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, this coefficient can be calculated by simply counting the number of concordant and discordant pairs, i.e.

$$\hat{\tau}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\#(\text{concordant pairs}) - \#(\text{discordant pairs})}{\binom{n}{2}}, \tag{2.74}$$

where

$$\#(\text{concordant pairs}) = \sum_{j=2}^{n}\sum_{i=1}^{j}\Big(\mathbb{1}(x_i > x_j)\,\mathbb{1}(y_i > y_j) + \mathbb{1}(x_i < x_j)\,\mathbb{1}(y_i < y_j)\Big) \tag{2.75}$$

$$\#(\text{discordant pairs}) = \sum_{j=2}^{n}\sum_{i=1}^{j}\Big(\mathbb{1}(x_i > x_j)\,\mathbb{1}(y_i < y_j) + \mathbb{1}(x_i < x_j)\,\mathbb{1}(y_i > y_j)\Big) \tag{2.76}$$

Due to these definitions, Spearman's rho and Kendall's tau are both called *rank correlation measures*, and in contrast to linear correlation, these measures are invariant under strictly increasing transformations. They are both fully specified by the copula C of X and Y, if it is known, and as such will often be denoted as $\rho(C)$ and $\tau(C)$.

Definition 2.28 (Tail dependence coefficients)**.** Let $X, Y \in \mathbb{R}$ be random variables with distribution functions F_X and F_Y, respectively. The *upper tail dependence coefficient* λ_U of X and Y is defined by

$$\lambda_U(X,Y) := \lim_{t\uparrow 1} \mathbb{P}\Big(Y > F_Y^{(-1)}(t) \,|\, X > F_X^{(-1)}(t)\Big), \tag{2.77}$$

and the *lower tail dependence coefficient* λ_L of X and Y is defined by

$$\lambda_L(X,Y) := \lim_{t\downarrow 0} \mathbb{P}\Big(Y \le F_Y^{(-1)}(t) \,|\, X \le F_X^{(-1)}(t)\Big), \tag{2.78}$$

when those limits exist.

The tail dependence coefficient measures the degree of dependence in the tails, e.g. how likely it is for extreme events to occur together. These quantities can also be expressed in terms of a copula: let C be the copula of X and Y; then

$$\lambda_U(X,Y) = \lim_{t\uparrow 1} \frac{1-2t+C(t,t)}{1-t}, \quad \text{and} \quad \lambda_L(X,Y) = \lim_{t\downarrow 0} \frac{C(t,t)}{t}. \tag{2.79}$$

Copulas for which either λ_U or λ_L is nonzero are said to be tail dependent.

2.6 Families of Copulas

Theorem 2.20 gives us the ability to separate the analysis of marginals from the analysis of dependence, but if the joint distribution is unknown—which it generally is—one will need to choose an appropriate copula in some way when doing emperical work.

When considering a *family* of copulas, we are dealing with functions that, given some parameters, are copulas, and the functional form is different for each valid set of parameters. Formally put, we define a family of copulas as follows:

Definition 2.29. Let Θ be some set. A mapping $\theta \in \Theta \mapsto C_\theta \in \mathscr{C}_d$ is called a *family of copulas*.

A family of copulas is, in other words, some subset $\{C_\theta\}_{\theta\in\Theta} \subseteq \mathscr{C}_d$ that's indexed by a suitable set Θ, which is often referred to as the *parameter space*. In some cases, we will further group families of copulas together into *classes* of copula families, which are similarly defined. Some properties that make a family of copulas appealing from a practical viewpoint are as follows:

- *Identifiability*: A family of copulas $\{C_\theta\}_{\theta\in\Theta}$ is said to be *identifiable* if a copula in it cannot be parameterised in two different ways, i.e. if $\theta \mapsto C_\theta$ is injective.

- *Interpretability*: Members of a family of copulas may have natural, probabilistic interpretations, which can suggest what kind of situations they are appropriate for.

- *Flexibility*: Another desirable property is that a family of copula covers the space between the Fréchet-Hoeffding bounds, and even includes the bounds (possibly as limiting cases). A family of copulas that includes both Π_d, W_d, and M_d is said to be *comprehensive*.

- *Ease of use*: Much desirable is that members of a family of copulas can be expressed on closed form, or at least are analytically tractable.

Many parametric families of copulas have been proposed in the literature, each of which imposes a different dependence structure on the data. Below, the ones we will be using in the application are specified.

2.6.1 Elliptical Copulas

A random vector $\boldsymbol{X} \in \mathbb{R}^d$ is said to have an elliptical distribution, if it can be represented as an affine transformation of a random vector $\boldsymbol{Z} \in \mathbb{R}^d$ with a spherical distribution. Such a vector is said to have a d-dimensional spherical distribution, if its characteristic function is on the form

$$\psi_{\boldsymbol{Z}}(\boldsymbol{t}) = \phi\left(\boldsymbol{t}^\top \boldsymbol{t}\right), \tag{2.80}$$

where ϕ is some scalar function. This is often denoted as $\boldsymbol{Z} \sim \mathcal{S}_d(\phi)$. Given such a $\boldsymbol{Z}$, $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z}, \tag{2.81}$$

where $\mu \in \mathbb{R}^d$, $\boldsymbol{A} \in \mathbb{R}^{d\times k}$ with $\boldsymbol{\Sigma} := \boldsymbol{A}\boldsymbol{A}^\top \in \mathbb{R}^{d\times d}$ and $\operatorname{rank}\boldsymbol{\Sigma} = k \le d$. If $\boldsymbol{\Sigma}$ is nonsingular, then the density of $\boldsymbol{X}$ has the form

$$f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{1/2} \phi\left(\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}\right)\right), \tag{2.82}$$

for all $\boldsymbol{x} \in \mathbb{R}^d$.

Definition 2.30. An *elliptical copula* is any copula C that can be obtained from an elliptical distribution with the inversion method of Lemma 2.22.

The most used variants of such families of copulas are given below.

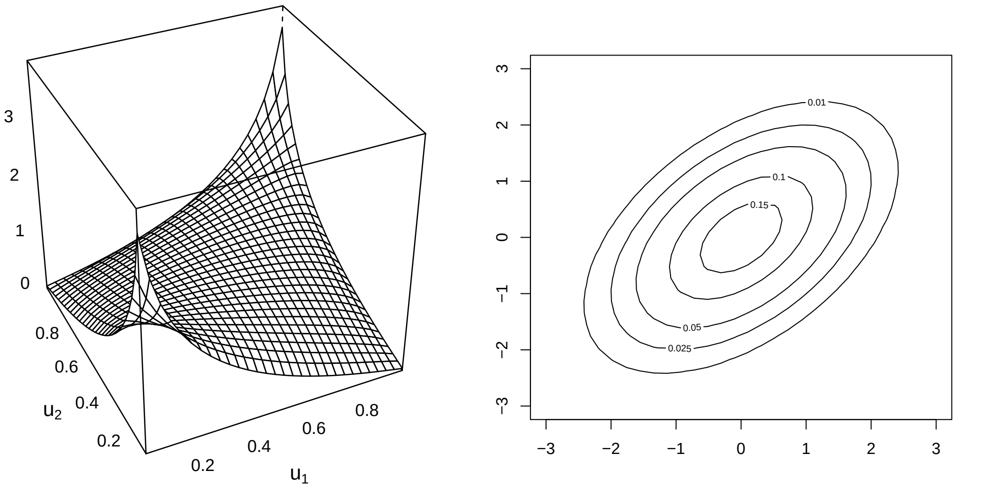
Example 2.31 (Gaussian copula)**.** A *Gaussian copula* is the elliptical d-copula $C_R^{\mathbf{Ga}}$ of a random vector $\boldsymbol{X} \sim \mathrm{N}_d(\mathbf{0}, R)$, i.e. for $\boldsymbol{u} \in \mathbb{I}^d$,

$$C_R^{\mathbf{Ga}}(\boldsymbol{u}) = \Phi_R\left(\Phi^{-1}\left(u_1\right), \ldots, \Phi^{-1}\left(u_d\right)\right), \tag{2.83}$$

where Φ^{-1} is the inverse distribution function for a standard normal variable, $R \in \mathbb{R}^{d\times d}$ is a correlation matrix, and Φ_R is the distribution function of a d-dimensional normal distribution with zero mean and coveriance matrix equal to R.

When $R = I_d$ this becomes the independence copula Π_d. In the bivariate case, the correlation matrix R reduces to a single number $\rho \in (-1, 1)$ and the copula has the form

$$C_\rho^{\mathbf{Ga}}(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \left(2\pi\sqrt{1-\rho^2}\right)^{-1} \exp\left(-\frac{s^2 - 2\rho s t + t^2}{2\left(1-\rho^2\right)}\right) \mathrm{d}t\, \mathrm{d}s, \tag{2.84}$$



***(a)** 3D plot of the copula density*

***(b)** Contour plot of the joint distribution with standard normal margins*

***Figure 2.4:** Example of a Gaussian copula in two dimensions with $\rho = \frac{1}{2}$*

and its association measures are given by

$$\rho\left(C_\rho^{\mathbf{Ga}}\right) = \frac{6}{\pi} \arcsin \frac{\rho}{2} \tag{2.85}$$

$$\tau\left(C_\rho^{\mathbf{Ga}}\right) = \frac{2}{\pi} \arcsin \rho \tag{2.86}$$

$$\lambda_U\left(C_\rho^{\mathbf{Ga}}\right) = \lambda_L\left(C_\rho^{\mathbf{Ga}}\right) = 0. \tag{2.87}$$

An example of a Gaussian copula is plotted on Figure 2.4.

Example 2.32 (Student's t-copula)**.** A *Student's t-copula* (or simply t-copula) is the elliptical d-copula $C_{R,\nu}^{\mathbf{t}}$ of a random vector $\boldsymbol{X} \sim \mathbf{t}_d(\nu, \mathbf{0}, R)$, i.e. for all $\boldsymbol{u} \in \mathbb{I}^d$,

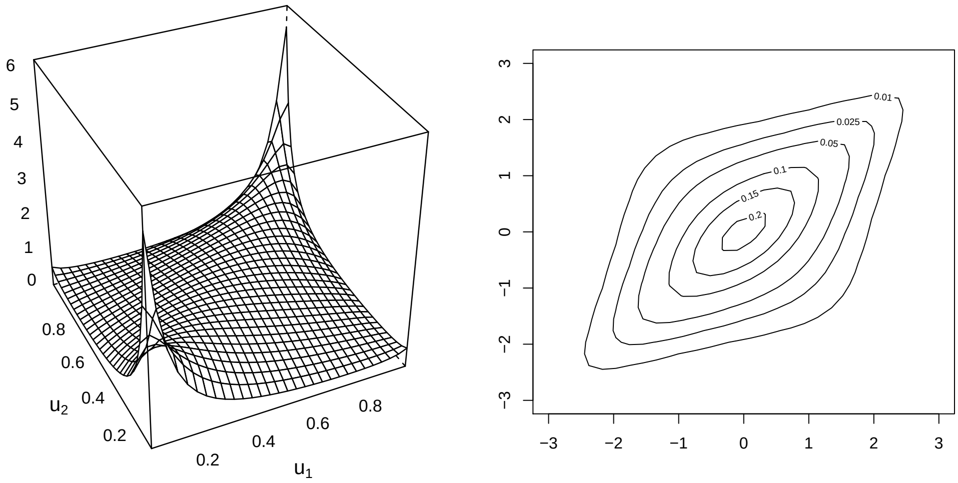
$$C_{R,\nu}^{\mathbf{t}}(\boldsymbol{u}) = \mathbf{t}_{R,\nu}\left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\right), \tag{2.88}$$

where $\mathbf{t}_{R,\nu}$ is the distribution function of a d-dimensional t-distributed variable with ν degrees of freedom and the correlation matrix R, and t_ν is the distribution function for a univariate standard t-distribution with ν degrees of freedom.

The t-copula has the Gaussian copula as a limiting case for $\nu \to \infty$. The rank correlation measures of $C_{\rho,\nu}^{\mathbf{t}}$ do not depend on the degrees of freedom ν, and they are identical to those of the Gaussian copula. Its tail dependence coefficients are identical due to the radial symmetry of the copula, and they are given by

$$\lambda_U\left(C_{\rho,\nu}^{\mathbf{t}}\right) = \lambda_L\left(C_{\rho,\nu}^{\mathbf{t}}\right) = 2\, t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right). \tag{2.89}$$

An example of a bivariate Student's t-copula is plotted on Figure 2.5.



*(**a**) 3D plot of the copula density*

*(**b**) Contour plot of the joint distribution with standard normal margins*

Figure 2.5: *Example of a t-copula in two dimensions with $\rho = \frac{1}{2}$ and $\nu = 3$*

2.6.2 Archimedian Copulas

The *Archimedian* class of copulas comprises families that are parameterised via a univariate generator function of the following type:

Definition 2.33. A function $\phi : \mathbb{R}_+ \to \mathbb{I}$ is called an *additive generator*, if

(a) it is continuous,

(b) it is non-increasing,

(c) $\phi(0) = 1$ and $\lim_{t\to\infty} \phi(t) = 0$

(d) it is strictly decreasing on $[0, t_0]$, where $t_0 := \inf\{t > 0 : \phi(t) = 0\}$,

and its *pseudo-inverse* is defined by

$$\phi^{(-1)}(t) := \begin{cases} \phi^{-1}(t), & t \in (0,1] \\ t_0, & t = 0. \end{cases} \tag{2.90}$$

Using any such function, one can construct a copula belonging to the class of Archimedian copulas, defined below.

Definition 2.34. A d-copula C is said to be *Archimedian* if $\exists \phi : \mathbb{R}_+ \to \mathbb{I}$ such that $\forall \boldsymbol{u} \in \mathbb{I}^d$,

$$C(\boldsymbol{u}) = \phi\Big(\phi^{(-1)}(u_1) + \cdots + \phi^{(-1)}(u_d)\Big). \tag{2.91}$$

Example 2.35 (Clayton copulas)**.** The *Clayton* family of d-copulas are parameterised by $\alpha \geq -1/(d-1), \alpha \neq 0$ and have the form

$$C_{\alpha}^{\mathbf{Clay}}(\boldsymbol{u}) = \max\left\{ \left(\sum_{j=1}^{d} u_i^{-\alpha} - (d-1) \right)^{-1/\alpha}, 0 \right\}, \tag{2.92}$$

and its generator is given by

$$\phi_{\alpha}(t) = \Big(\max\{1 + \alpha t, 0\} \Big)^{-1/\alpha}. \tag{2.93}$$

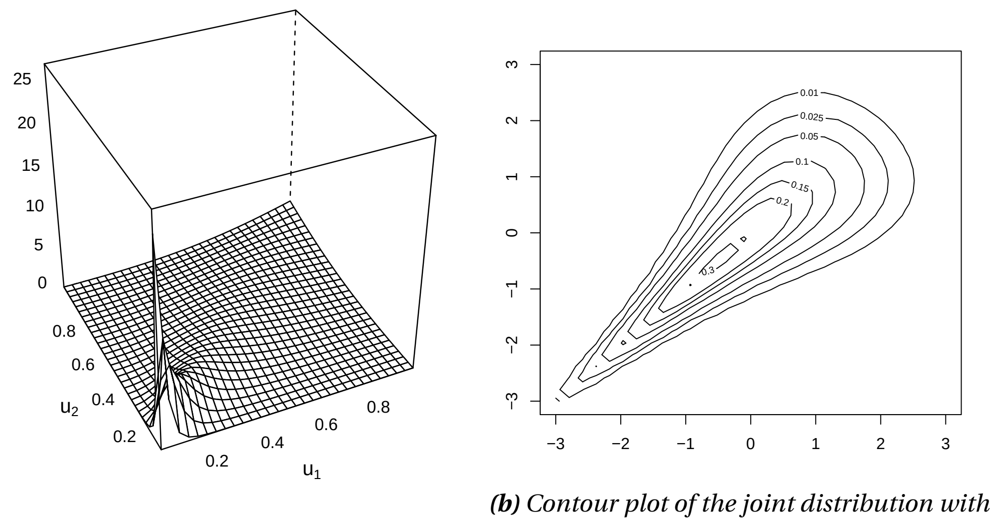
The Kendall's tau of this copula in its bivariate form is given by

$$\tau\Big(C_{\alpha}^{\mathbf{Clay}}\Big) = \frac{\alpha}{\alpha+2}, \tag{2.94}$$

and its tail dependence measures are

$$\lambda_L\Big(C_{\alpha}^{\mathbf{Clay}}\Big) = \begin{cases} 2^{-1/\alpha}, & \alpha > 0, \\ 0, & \alpha \in [-1, 0], \end{cases} \qquad \lambda_U\Big(C_{\alpha}^{\mathbf{Clay}}\Big) = 0, \tag{2.95}$$

meaning that Clayton copulas only capture lower tail dependence. Spearman's rho is not given for this family, since in contrast to the simple form of the Kendall's tau, however, the association between the copula parameter and the Spearman's rho is very complicated. The limiting case $\alpha \to 0$ is the independence copula Π_d. An example of a Clayon copula is plotted on Figure 2.6.



(a) 3D plot of the copula density

(b) Contour plot of the joint distribution with standard normal margins

Figure 2.6: *Example of a Clayton copula in two dimensions with $\alpha = 3$*

Example 2.36 (Gumbel copulas)**.** The *Gumbel* family of copulas are parameterised by $\alpha \geq 1$ and have the functional form

$$C_\alpha^{\mathbf{Gum}}(\boldsymbol{u}) = \exp\left(-\left(\sum_{j=1}^{d} \left(-\log u_j \right)^{\alpha} \right)^{1/\alpha} \right), \tag{2.96}$$

and its generator function is given by

$$\phi_\alpha(t) = \exp\left(-t^{1/\alpha} \right). \tag{2.97}$$

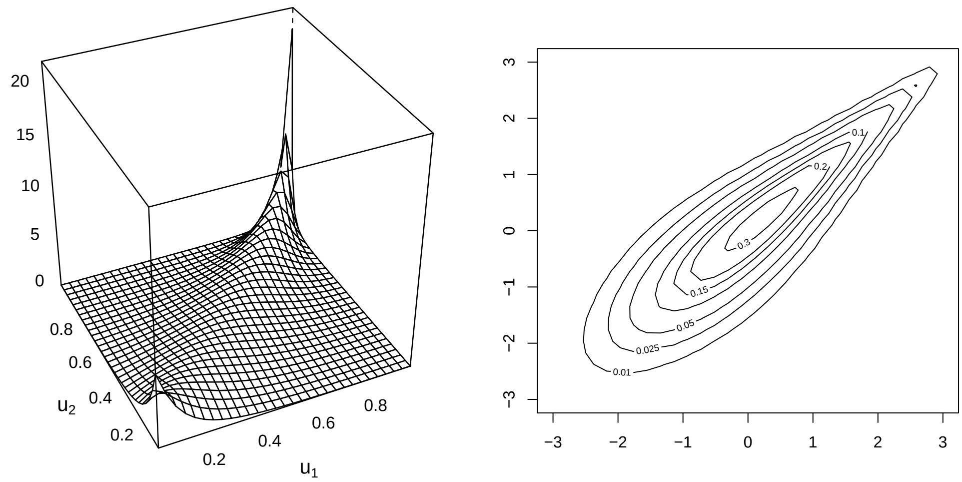
The bivariate Gumbel family of copulas has the Kendall's tau

$$\tau\left(C_\alpha^{\mathbf{Gum}} \right) = 1 - \alpha^{-1}, \tag{2.98}$$

but its Spearman's rho does not have a closed form. Its tail dependence measures are given by

$$\lambda_L\left(C_\alpha^{\mathbf{Gum}} \right) = 0, \quad \lambda_U\left(C_\alpha^{\mathbf{Gum}} \right) = 2 - 2^{1/\alpha}, \tag{2.99}$$

meaning that Gumbel copulas exhibits upper tail dependence, but no lower tail dependence. When $\alpha = 1$, the copula reduces to the independence copula Π_d, and in the limiting case for $\alpha \to \infty$, one obtains the comonotonicity copula M_d. An example of a bivariate Gumbel copula is plotted on Figure 2.7.



(a) *3D plot of the copula density*

(b) *Contour plot of the joint distribution with standard normal margins*

Figure 2.7: *Example of a Gumbel copula in two dimensions with* $\alpha = 3$

Example 2.37 (Frank copulas)**.** The *Frank* family of d-copulas is parameterised by $\alpha > 0$ and has the functional form

$$C_\alpha^{\textbf{Frank}}(\boldsymbol{u}) = -\frac{1}{\alpha}\log\left(1 + \frac{\prod_{j=1}^d \left(\mathrm{e}^{-\alpha u_j} - 1\right)}{\left(\mathrm{e}^{-\alpha} - 1\right)^{d-1}}\right), \tag{2.100}$$

and its generator is given by

$$\phi_\alpha(t) = \frac{1}{\alpha}\log\Big(1 - \left(1 - \mathrm{e}^{-\alpha}\right)\mathrm{e}^{-t}\Big). \tag{2.101}$$

Like the elliptical copulas, Frank copulas are radially symmetric, and it has no tail dependence, i.e. $\lambda_L\Big(C_\alpha^{\textbf{Frank}}\Big) = \lambda_U\Big(C_\alpha^{\textbf{Frank}}\Big) = 0$. In contrast to the Clayton and Gumbel copulas, the Spearman's rho of Frank copulas does have a closed form,

$$\rho\Big(C_\alpha^{\textbf{Frank}}\Big) = 1 - \frac{12}{\alpha}\left(D_1(\alpha) - D_2(\alpha)\right), \tag{2.102}$$

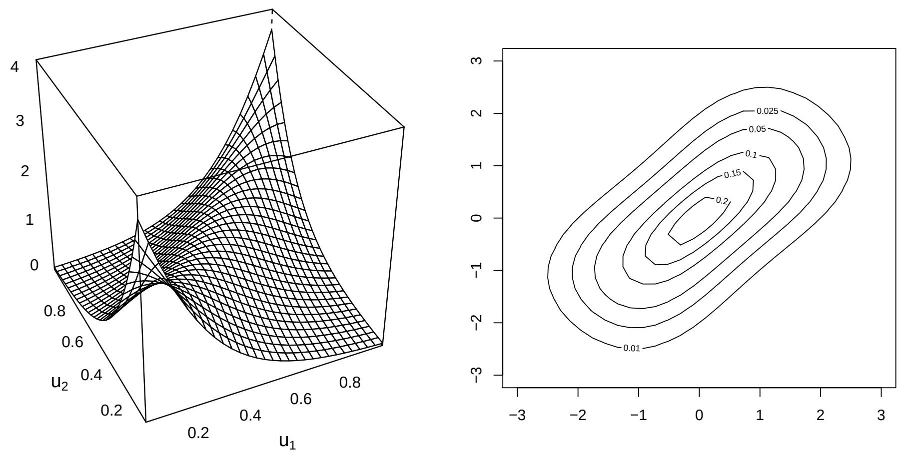
where D_n is the Debye function defined by

$$D_n(x) := \frac{n}{x^n}\int_0^x \frac{t^n}{e^t - 1}\,\mathrm{d}t, \tag{2.103}$$

for any $n \in \mathbb{Z}$. Likewise, the Kendall's tau is given by

$$\tau\Big(C_\alpha^{\textbf{Frank}}\Big) = 1 - \frac{4}{\alpha}\left(1 - D_1(\alpha)\right). \tag{2.104}$$

An example of a bivariate Frank copula is plotted on Figure 2.8.



(a) 3D plot of the copula density *(b) Contour plot of the joint distribution with standard normal margins*

Figure 2.8: *Example of a Frank copula in two dimensions with $\alpha = 5$*

2.6.3 Rotations of Copulas

The archimedian families of copulas described above all have positive dependence, i.e. dependence structures with positive correlation measures, but we would like to have variants of them with negative dependence. The following result from [Durante and Sempi, 2016] is helpful in that regard:

Definition 2.38. A *symmetry* of $\mathbb{I}^d$ is a bijection $\xi : \mathbb{I}^d \to \mathbb{I}^d$ on the form $\xi(u_1,\dots,u_d) = (v_1,\dots,v_d)$, where for each $j = 1,\dots,d$ and each permutation $(k_1,\dots,k_d)$ of $(1,\dots,d)$, either

$$v_j = u_{k_j} \quad \text{or} \quad v_j = 1 - u_{k_j}. \tag{2.105}$$

Theorem 2.39. *Let $\boldsymbol{U}$ be a d-dimensional random vector whose distribution function is given by a copula $C \in \mathscr{C}_d$. Let ξ be a symmetry in $\mathbb{I}^d$ and consider the random vector $\boldsymbol{V} = \xi \circ \boldsymbol{U}$. Then the distribution function C^ξ of $\boldsymbol{V}$ is a copula.*

In other words, one can consider symmetric transformations of marginals, and the resulting joint distribution will still be a copula. As it turns out, one can express such a copula in terms of the original copula. For simplicity in notation, and because we later only use bivariate copulas, we present the rotations of copulas in two dimensions:

Definition 2.40. Rotated copulas Let $C \in \mathscr{C}_2$ be a 2-dimensional copula. Then the *90, 180, and 270 degree rotations* of C are given as, respectively,

$$C^{90}(u_1,u_2) = u_2 - C(1-u_1,u_2) \tag{2.106}$$

$$C^{180}(u_1,u_2) = u_1 + u_2 - 1 + C(1-u_1,1-u_2) \tag{2.107}$$

$$C^{270}(u_1,u_2) = u_1 - C(u_1,1-u_2). \tag{2.108}$$

An example is shown on Figure 2.9.

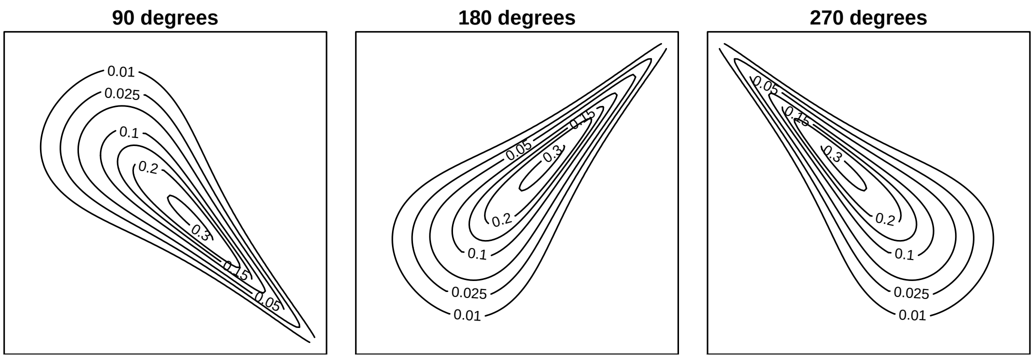


Figure 2.9: *Sketch of contours of rotations of a Clayon copula with $\alpha = 3$*

3 Vines

Bivariate copulas (i.e. 2-copulas) have been studied extensively in the literature, whereas their multivariate versions have seen more limited treatment, due to analytical and computational complexity. As a result, a construction called a *vine copula*, a graphical model which use 2-copulas as building blocks for multivariate ones, was introduced.

Vines were first introduced by [Joe, 1994], in which the author sought to extend bivariate copula families to higher dimensions, and to do this, he introduced what would later be called the *D-vine*. The term *vine*, however, was coined by [Bedford and Cooke, 2002], who were motivated by uncertainty analysis of risk models and who introduced a more systematic decomposition. Vines of Gaussian pair copulas were analysed in [Kurowicka and Cooke, 2006], and maximum likelihood estimation for non-Gaussian vines was first studied in [Aas et al., 2009]. To formally define vine copulas, we need some basic definitions from graph theory:

Definition 3.1 (Graph)**.** A *graph* is an ordered triple $G = (V, E, \lambda)$ where

(a) V is a non-empty set of arbitrary elements called *nodes*

(b) E is a set of arbitrary elements called *edges*

(c) $\lambda : E \to \big\{\{u, v\} : (u, v) \in V \times V, x \neq y\big\}$ is an injective label function that associates each edge in E with a pair of nodes

For a graph $G = \big(V, E, \lambda\big)$, a node $v \in V$ and an edge $e \in E$ are said to be *incident* if $v \in \lambda(e)$, and two nodes $u, v \in V$ are said to be *neighbours* if there is an $e \in E$ such that $\lambda(e) = \big\{u, v\big\}$. A node is said to have *degree k* if it is incident with k edges. If all distinct pairs of nodes in a graph are connected by a unique edge, the graph is said to be *complete*.

Definition 3.2 (Path)**.** Let $G = (V, E, \lambda)$ be a graph and $u_0, \ldots, u_n \in V$ be nodes in G. A *path* from u_0 to u_n of length $n \in \mathbb{N}$ in G is an ordered sequence of edges $\big(e_j\big)_{j=1}^{n}$ such that $\lambda\big(e_j\big) = \{u_{j-1}, u_j\}$ for $j = 1, \ldots, n$.

Definition 3.3 (Tree)**.** A *tree* is a graph $T = (V, E, \lambda)$ such that for any two nodes $u, v \in V$, a path of distinct edges from u to v exists and is uniquely determined.

The *vine* can now be defined.

Definition 3.4 (Regular vine)**.** A sequence of m trees, $\mathcal{V} = \big(T_j\big)_{j=1}^{m}$, is called a d-dimensional *vine* if

(a) For the first tree $T_1 = (V_1, E_1, \lambda_1), |V_1| = d$

(b) For $j = 2, \ldots, m$, the tree $T_j = (V_j, E_j, \lambda_j)$ has nodes $V_j \subseteq V_1 \cup E_1, \cup \cdots \cup E_{j-1}$

A vine $\mathcal{V}$ is furthermore called a *regular vine* (R-vine for short) on n elements if

(c) $m = d$

(d) For $j = 2, \ldots, d$, $V_j = E_{j-1}$ with $|V_j| = d - (j-1)$

(e) For $j = 2, \ldots, d-1$, if $u_j, v_j \in V_j$ are two nodes connected by an edge, i.e. $\exists e \in E_j$ such that $\lambda_j(e) = \{u_j, v_j\}$, then $|\lambda_{j+1}(u_j) \cap \lambda_{j+1}(v_j)| = 1$

In other words, an R-vine on n elements is a sequence of trees nested in such a way that the edges in one tree becomes the nodes of the next. Condition (e) is commonly referred to as the *proximity condition*, and it ensures that two nodes in a tree only share an edge, if the corresponding edges in the previous tree are incident with a common node.

Below are given definitions for some sets needed to characterise R-vines and to express the main result about vine copulas. In all three of them, let $\mathcal{V} = (T_j)_{j=1}^{d}$ be a d-dimensional R-vine, i.e. $T_j = (V_j, E_j, \lambda_j)$ is a tree for $j = 1, \ldots, d$.

Definition 3.5. *The complete union* of an edge $e_j \in E_j, j = 1, \ldots, d$, is the set

$$U_{e_j} := \Big\{ v \in V_1 : \exists e_k \in E_k, k = 1, \ldots, j-1 : v \in \lambda_1(e_1), e_1 \in \lambda_2(e_2), \ldots, e_{j-1} \in \lambda_j(e_j) \Big\}. \tag{3.1}$$

Definition 3.6. For an edge $e_j \in E_j, j = 1, \ldots, d$, with $\lambda_j(e_j) = \{u, v\}$, where $u, v \in V_j$, *the conditioning set* of e_j is $D_{e_j} := U_u \cap U_v$.

Definition 3.7. For an edge e_j as in Definition 3.6, *the conditioned sets* of e_j are

$$C_{e_j,u} := U_u \setminus D_{e_j} \tag{3.2}$$

$$C_{e_j,v} := U_v \setminus D_{e_j} \tag{3.3}$$

$$C_{e_j} := U_u \oplus U_v, \tag{3.4}$$

where $\oplus$ denotes the symmetric difference, i.e. $A \oplus B = (A \setminus B) \cup (B \setminus A)$.

These sets contain the information necessary to fully identify an R-vine. When drawing R-vines, it is convenient to label the nodes and edges by their respective conditioned and conditioning sets. For the purpose of brevity, we will shorten the notation by identifying the members of these sets simply by their indicies, so that, for example, a conditioning set comprising two nodes $v_j, v_k \in V_1$ is denoted as $\{j, k\}$. In this labelling scheme, an edge $e_j \in E_j$ is labelled as "$C_{e_j}|D_{e_j}$". An example of an R-vine with such a labelling scheme is shown on Figure 3.1.

We can now define the *vine copula* as an R-vine in which the elements of the first tree correspond to the d marginal distributions of the variables examined. To construct such a d-dimensional vine copula, one needs to first specify $d-1$ unconditional 2-copulas between variables, indexed by the variables they link, and then recursively specify conditional 2-copulas between variables obtained by transforming variables from the previous tree.

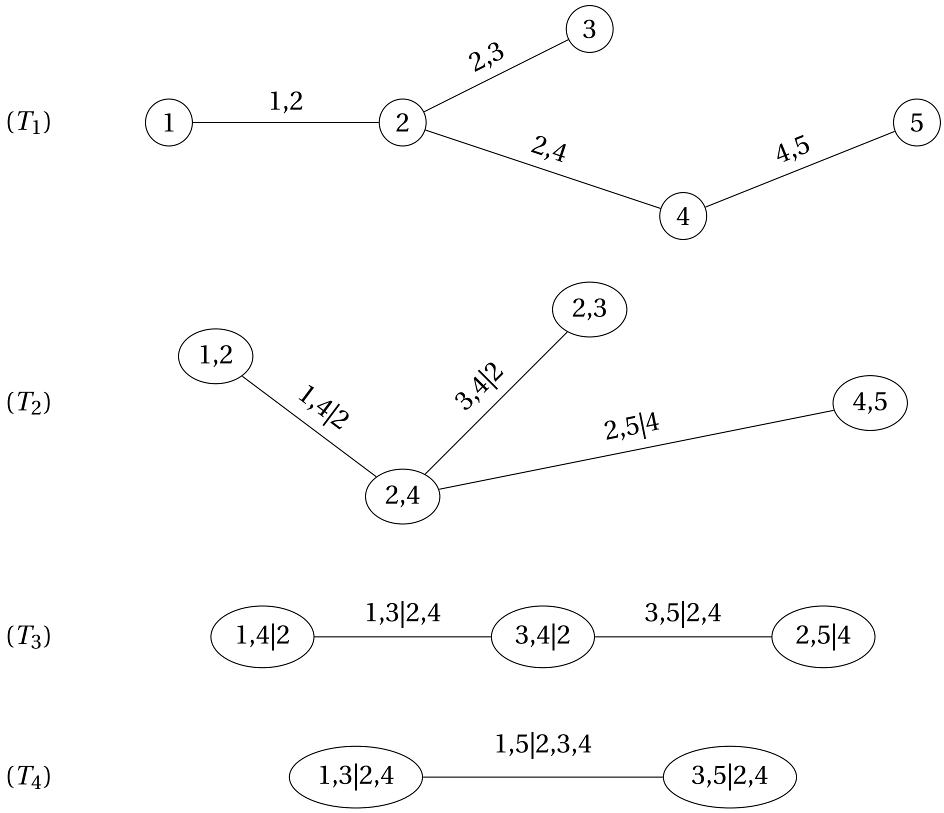


Figure 3.1: *Example of a 5-dimensional R-vine, with node and edge labels given by the indices in their respective conditioned and conditioning sets*

Definition 3.8. An *R-vine copula specification* is a triple $\left(\boldsymbol{F}, \mathcal{V}, B\right)$, where $\boldsymbol{F} = \left(F_1, \ldots, F_d\right)$ is a vector of distribution functions, $\mathcal{V} = \left(\left(V_j, E_j, \lambda_j\right)\right)_{j=1}^{d}$ is a d-dimensional R-vine, and B is a set of pair-copulas B_e, that is $B = \left\{B_e \in \mathscr{C}_2 : e \in E_j, j = 1, \ldots, d-1\right\}$.

With this definition, a joint distribution $H : \mathbb{R}^d \to \mathbb{I}$ of a random vector $\boldsymbol{X}$ is said to realise an R-vine copula specification if for each $e \in E_j$, $j = 1, \ldots, d-1$ with $\lambda_j(e) = \{u, v\}$, B_e is the 2-copula of $X_{C_{e,u}}|\boldsymbol{X}_{D_e}$ and $X_{C_{e,v}}|\boldsymbol{X}_{D_e}$, and such a distribution is called an R-vine distribution. The density of such a distribution was shown to be a product of unconditional and conditional 2-copula densities in [Bedford and Cooke, 2002], and this result is repeated below.

Theorem 3.9. *Let $\left(\boldsymbol{F}, \mathcal{V}, B\right)$ be a d-dimensional R-vine specification, with $\boldsymbol{F} = \left(F_1, \ldots, F_d\right)$, $\mathcal{V} = \left(\left(V_j, E_j, \lambda_j\right)\right)_{j=1}^{d}$, and let $c_{C_{e,u}, C_{e,v}|D_e}$ be the copula density for the 2-copula B_e, $e \in E_j$ with $\lambda_j(e) = \{u, v\}$, $j = 1, \ldots, d-1$. Then there exists a unique, d-dimensional distribution H that realises this R-vine copula specification with the density*

$$f_{1,\ldots,d}(\boldsymbol{x}) = \prod_{k=1}^{d} f_k\left(x_k\right) \prod_{j=1}^{d-1} \prod_{e \in E_j} c_{C_{e,u}, C_{e,v}|D_e}\Big(F_{C_{e,u}|D_e}\left(x_{C_{e,u}}|\boldsymbol{x}_{D_e}\right), F_{C_{e,v}|D_e}\left(x_{C_{e,v}}|\boldsymbol{x}_{D_e}\right)\Big), \tag{3.5}$$

where f_j denotes the density of F_j for $j = 1, \ldots, d$, and $F_{C_{e,u}|D_e}$ is the conditional distribution of $X_{C_{e,u}}|\boldsymbol{X}_{D_e}$.

We skip the formal proof for Theorem 3.9 here (see [Bedford and Cooke, 2002]) and instead provide an intuitive demonstration of such an R-vine copula construction. Let $\boldsymbol{X} = \left(X_1, \ldots, X_d\right)$ be a random vector on the probability space $\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ with the joint density function $f : \mathbb{R}^d \to \mathbb{R}$. As is well known from probability theory, the law of total probability lets one factorise the density into a product of conditional densities,

$$f_{1,\ldots,d}(\boldsymbol{x}) = f_{d|d-1,\ldots,1}\left(x_d|x_{d-1}, \ldots, x_1\right) f_{d-1|d-2,\ldots,1}\left(x_{d-1}|x_{d-2}, \ldots, x_1\right) \cdots f_{2|1}\left(x_2|x_1\right) f_1\left(x_1\right), \tag{3.6}$$

where f_j denotes the marginal density for X_j, $j = 1, \ldots, d$, and $f_{j|\mathcal{J}}$ denotes the conditional density for $X_j|\boldsymbol{X}_{\mathcal{J}}$. Notice that this factorisation is not unique, but is just given as a special case here for illustrative purposes. By Corollary 2.21, the joint density of the subvector $\left(X_1, X_2\right)$ can be expressed of a 2-copula density,

$$f_{1,2}\left(x_1, x_2\right) = c_{1,2}\Big(F_1\left(x_1\right), F_2\left(x_2\right)\Big) f_1\left(x_1\right) f_2\left(x_2\right), \tag{3.7}$$

where F_1 and F_2 are the marginal distribution functions for X_1 and X_2, respectively, and thus, the conditional density can be written as

$$f_{2|1}\left(x_2|x_1\right) = c_{1,2}\Big(F_1\left(x_1\right), F_2\left(x_2\right)\Big) f_2\left(x_2\right). \tag{3.8}$$

Likewise, the conditional density $f_{3|1,2}$ of the random variable $X_3|X_2, X_1$ can be expressed as e.g.

$$f_{3|1,2}\left(x_3|x_1, x_2\right) = c_{1,3|2}\Big(F_{1|2}\left(x_1|x_2\right), F_{3|2}\left(x_3|x_2\right)\Big) f_{3|2}\left(x_3|x_2\right), \tag{3.9}$$

where $F_{1|2}$ and $F_{3|2}$ are the conditional distribution functions of $X_1|X_2$ and $X_3|X_2$, respectively. One can then further write conditional densities appearing as factoris in one expression in terms of pair copulas, until all that is left are marginal densities and pair copulas. This is best illustrated through a low-dimensional example:

Example 3.10. Let $d = 4$. The density $f_{1,2,3,4}$ of $\boldsymbol{X} = \left(X_1, \ldots, X_4\right)$ can be decomposed as

$$f_{1,2,3,4}(\boldsymbol{x}) = \underline{f_{4|1,2,3}\left(x_4|x_1, x_2, x_3\right)}\ \underline{f_{3|1,2}\left(x_3|x_1, x_2\right)}\ \underline{f_{2|1}\left(x_2|x_1\right)}\ f_1\left(x_1\right), \tag{3.10}$$

where the resulting conditional densities appearing as factors (underlined) can in turn be expressed using copulas;

$$f_{2|1}\left(x_2|x_1\right) = c_{1,2}\Big(F_1\left(x_1\right), F_2\left(x_2\right)\Big)\, f_2\left(x_2\right) \tag{3.11}$$

$$f_{3|1,2}\left(x_3|x_1, x_2\right) = c_{1,3|2}\Big(F_{1|2}\left(x_1|x_2\right), F_{3|2}\left(x_3|x_2\right)\Big)\, \underline{f_{3|2}\left(x_3|x_2\right)} \tag{3.12}$$

$$f_{4|1,2,3}\left(x_4|x_1, x_2, x_3\right) = c_{1,4|2,3}\Big(F_{1|2,3}\left(x_1|x_2, x_3\right), F_{4|2,3}\left(x_4|x_2, x_3\right)\Big)\, \underline{f_{4|2,3}\left(x_4|x_2, x_3\right)}, \tag{3.13}$$

and in turn, the conditional densities appearing in those expressions can further be rewritten;

$$f_{3|2}\left(x_3|x_2\right) = c_{2,3}\left(F_1\left(x_2\right), F_3\left(x_3\right)\right) f_3\left(x_3\right) \tag{3.14}$$

$$f_{4|2,3}\left(x_4|x_2, x_3\right) = c_{2,4|3}\left(F_{2|3}\left(x_2|x_3\right), F_{4|3}\left(x_4|x_3\right)\right) \underline{f_{4|3}\left(x_4|x_3\right)}, \tag{3.15}$$

and finally, the conditional density appearing here can be rewritten;

$$f_{4|3}\left(x_4|x_3\right) = c_{3,4}\Big(F_3\left(x_3\right), F_4\left(x_4\right)\Big)\, f_4\left(x_4\right). \tag{3.16}$$

Now, collecting similar terms, one sees

$$\begin{aligned} f_{1,2,3,4}(\boldsymbol{x}) = {} & f_1\left(x_1\right) f_2\left(x_2\right) f_3\left(x_3\right) f_4\left(x_3\right) \\ & \times c_{1,2}\Big(F_1\left(x_1\right), F_2\left(x_2\right)\Big)\, c_{2,3}\left(F_1\left(x_2\right), F_3\left(x_3\right)\right) c_{3,4}\Big(F_3\left(x_3\right), F_4\left(x_4\right)\Big) \\ & \times c_{1,3|2}\Big(F_{1|2}\left(x_1|x_2\right), F_{3|2}\left(x_3|x_2\right)\Big)\, c_{2,4|3}\left(F_{2|3}\left(x_2|x_3\right), F_{4|3}\left(x_4|x_3\right)\right) \\ & \times c_{1,4|2,3}\Big(F_{1|2,3}\left(x_1|x_2, x_3\right), F_{4|2,3}\left(x_4|x_2, x_3\right)\Big), \end{aligned} \tag{3.17}$$

which is exactly the form of the density specified in Theorem 3.9, and on Figure 3.2 the structure is shown (with labels slightly extented to make clear which nodes correspond to marginal densities and which correspond to copulas).

There are two special cases of R-vines that are commonly used: *C-vines* and *D-vines.* A C-vine (short for *canonical vine*) is regular vine in which a single node is selected as the *root node* in each tree, and all pair-copulas in the given tree have that node as an edge, resulting in a star-shaped structures. A D-vine (short for *drawable vine*) is a regular vine in which one chooses a specific order of the variables, leading to a path structure in the trees. For example, the vine on Figure 3.2 is a D-vine, and we will see examples of C-vines arising in the applied part of this project.

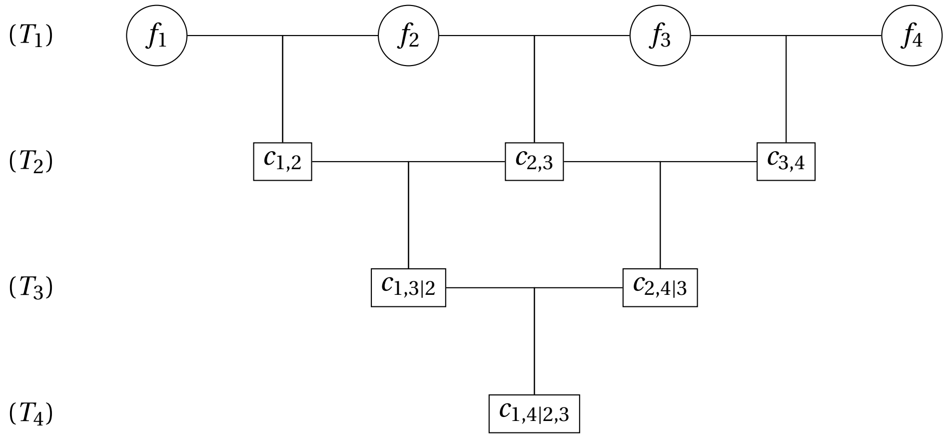


Figure 3.2: *Visualisation of an R-vine distribution on four variables with marginal densities* f_j, $j = 1, \dots, 4$. *Here, an edge going from another edge to a node in the next tree just represents that such nodes are treated as edges in the previous tree.*

3.1 Estimation

The process of fitting an R-vine distribution to a real data set can be split into three separate parts:

(a) Selection of the R-vine structure

(b) Choice of 2-copulas for each edge in this R-vine

(c) Estimation of the parameter(s) of each of these 2-copulas

Unsurprisingly, when moving from single copulas to vine copulas, the process becomes much more complex. The number of possible structures for a d-dimensional vines grows very rapidly in d, with this number being given by $\prod_{j=1}^{d} i^{i-2}$ [Morales-Nápoles et al., 2010], so doing steps (b) and (c) for all such structures quickly become an infeasible endeavour. In light of this, it's also infeasible to choose bivariate copula families by visual interpretation of plots of the data, as such heuristic analyses would have to be made for every pair of every possible R-vine.

If at first, we assume that the R-vine structure and copula families are known beforehand, then the parameters can be estimated using maximum likelihood estimation, if we can evaluate the density of the R-vine distribution. A method for this was developed by [Dißmann et al., 2013], in which an array representation of the R-vine is used to encode information about the conditioned and conditioning sets. This method also makes use of Corollary 2.21, which was used in the context of vines by [Joe, 1996], stating that for a vine copula $(\boldsymbol{F}, \mathcal{V}, B)$, $e \in E_j$ with $\lambda_j(e) = \{u, v\}$, $\lambda_{j-1}(u) = \{u_1, u_2\}$, and $\lambda_{j-1}(v) = \{v_1, v_2\}$,

$$\begin{aligned} F_{C_{e,u}|D_e}\left(x_{C_{e,u}}|\boldsymbol{x}_{D_e}\right) &= \frac{\partial C_{C_u|D_u}\Big(F_{C_{u,u_1}|D_u}\left(x_{C_{u,u_1}}|\boldsymbol{x}_{D_u}\right), F_{C_{u,u_2}|D_u}\left(x_{C_{u,u_2}}|\boldsymbol{x}_{D_u}\right)\Big)}{\partial F_{C_{u,u_2}|D_u}\left(x_{C_{u,u_2}}|\boldsymbol{x}_{D_u}\right)} \\ &=: h\Big(F_{C_{u,u_1}|D_u}\left(x_{C_{u,u_1}}|\boldsymbol{x}_{D_u}\right), F_{C_{u,u_2}|D_u}\left(x_{C_{u,u_2}}|\boldsymbol{x}_{D_u}\right)\Big), \end{aligned} \tag{3.18}$$

and similarly for $F_{C_{e,v}|D_e}$. In other words, the conditional distributions can be obtained recursively using copulas, and in the literature, this function is called the *h-function*, for notation purposes.

3.1.1 Sequential Estimation

A method more typical in the literature, however, is taking advantage of the ordered tree structure of the R-vine in so-called *sequential estimation*, which in short goes as follows

(1) Estimate parameter(s) for each 2-copula in E_1

(2) Compute the transformed variables for V_2 using h-functions

(3) Repeat the steps for $j = 2,\ldots,d-1$, using the transformed variables to estimate copula parameters

Since this method only involves maximum likelihood estimation for 2-copulas, it's rather fast, and in terms of computing strategies, step (1) can be done in parallel. The estimated parameters can then be used as starting values for the full joint maximum likelihood estimation.

Unless one has expert knowledge about how the vine should be structured beforehand, however, we need some method to automatically select the "best" structure, in some sense of the word. One such method was proposed by [Dißmann et al., 2013], in which empirical Kendall's taus are calculated and used for edge weights in a complete graph of the variables, which is then pruned using a *maximum spanning tree* algorithm. This method is described in Algorithm 3.1.

Algorithm 3.1 Algorithm for selecting an R-vine structure

procedure RVINECOPSELECT($x_{n,1},\ldots,x_{n,d}$ for $n = 1,\ldots,N$) ▷ N realisations of a d-dimensional random vector with marginals $F_1,\ldots,F_d$

  $\hat{\tau}_{j,k} \leftarrow \hat{\tau}\big(\boldsymbol{x}_j, \boldsymbol{x}_k\big)$ for all possible pairs $\{j,k\}, 1 \le j < k \le d$

  Select $T_1 = \big(V_1, E_1, \lambda_1\big)$ to be the tree such that

$$T_1 = \arg\max_{T_1} \sum_{\substack{e \in E_1 \\ \lambda_1(e) = \{\boldsymbol{x}_j, \boldsymbol{x}_k\}}} \big|\hat{\tau}_{j,k}\big|$$

  For each edge e with $\lambda_1(e) = \{\boldsymbol{x}_j, \boldsymbol{x}_k\}$, select a 2-copula and estimate the parameter(s)

  $\boldsymbol{x}_{j|k} \leftarrow \widehat{F}_{j|k}\big(x_{n,j}|x_{n,k}\big)$ and $\boldsymbol{x}_{k|j} \leftarrow \widehat{F}_{k|j}\big(x_{n,k}|x_{n,j}\big)$ for $n = 1,\ldots,N$ using the fitted copula

  for $j = 2,\ldots,d-1$ **do**

    $\hat{\tau}_{j,k|D} \leftarrow \hat{\tau}\big(\boldsymbol{x}_{j|D}, \boldsymbol{x}_{k|D}\big)$ for all conditional variable pairs $\{j,k|D\}$ such that all edges fulfil the proximity condition

    For each edge e with $\lambda_j(e) = \{\boldsymbol{x}_{j|D}, \boldsymbol{x}_{k|D}\}$, select and estimate a conditional 2-copula

    $\boldsymbol{x}_{j|k,D} \leftarrow \widehat{F}_{j|k,D}\big(x_{n,j}|x_{n,k}, \boldsymbol{x}_{n,D}\big)$ and $\boldsymbol{x}_{k|j,D} \leftarrow \widehat{F}_{k|j,D}\big(x_{n,k}|x_{n,j}, \boldsymbol{x}_{n,D}\big)$ using the fitted copula

  return R-vine copula specifiction, $\big(\boldsymbol{F}, \mathcal{V}, B\big)$

Note that this method still requires one to choose the copula families used for each edge. A commonly used method is to have a set of candidate families and choose between them for each edge using a selection criterion. The Akaike's Information Criterion [Akaike, 1973] (AIC) is commonly used for this, and has been demonstrated to be a reliable criterion for copulas by [Brechmann, 2010] and [Manner, 2007].

3.2 Simulation

Parts of the analysis in later sections will involve estimating quantities using simulation, and as part of the pipeline, we need to be able to simulate observations coming from an R-vine copula distribution. Sklar's theorem also proves useful here, as it implies that in order to simulate a random sample from a d-copula C, one just needs to

1. Simulate $\boldsymbol{u} = (u_1, \dots, u_d)$ from C
2. Transform each u_j by the inverse of the respective marginal distribution functions, i.e. set $\boldsymbol{x} = \left(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\right)$

Step 2 is straightforward, given the marginal distributions $F_1, \dots, F_d$ and a sample from the copula from step 1, but drawing this copula sample requires doing the following:

1. Draw d independent samples, $v_1, \dots, v_d$, from Unif(0, 1)
2. Transform these samples as follows

$$
\begin{aligned}
u_1 &= v_1 \\
u_2 &= F_{2|1}^{-1}(v_2|u_1) \\
u_3 &= F_{3|1,2}^{-1}(v_3|u_1, u_2) \\
&\vdots \\
u_d &= F_{d|1,\dots,d-1}^{-1}(v_d|u_1, \dots, u_{d-1})
\end{aligned}
$$

This idea extends to vine copulas by using inverses of h functions, which is detailed in [Brechmann, 2010].

Part II
Application in Energy Markets

4 European Power Market Data

In this second part of the project, we turn our attention to the German electricity market, and employ a model involving a vine copula as described in the previous part in order to describe and simulate the joint behaviour of certain market variables.

The data used consists of a set of time series downloaded from the European Network of Transmission System Operators for Electricity's Transparency Platform, extracted from the following regulation articles (details about these articles can be found on [ENTSO-E, 2019]):

Day-ahead Prices (12.1.D) consists of hourly day-ahead electricity prices in EUR/MW for given control areas, bidding zones, or countries. Germany is part of a bidding zone together with Luxembourg, and prior to 2018-09-30, Austria was also part of that bidding zone, so to obtain historical data for German day-ahead prices, DE-AT-LU bidding zone prices have been downloaded for 2015-2018 and DE-LU bidding zone prices have been downloaded for 2018-2019.

Actual Total Load (6.1.A) consists of quarterly total load (power demand) in MW for given control areas, bidding zones, or countries. Historical data for Germany has been downloaded from this article for 2015-2019.

Actual Generation per Production Type (16.1.B&C) consists of hourly, aggregated generation in MW for each type of power production: biomass, lignite, coal-derived gas, gas, hard coal, oil, oil shale, peat, geothermal, pumped storage, run-of-river, water reservoir, marine, nuclear, waste, solar, onshore wind, offshore wind, and other renewables. Historical data for German solar, onshore wind, and offshore wind production has been downloaded from this article for 2015-2019.

4.1 Preprocessing and Aggregation

The data is preprocessed and aggregated into four daily time series representing prices, load, solar production, and wind production, respectively.

The series of prices consists of DE-AT-LU bidding zone prices starting from 2015-01-06, and ending on 2018-09-30, and then DE-LU bidding zone prices starting from 2018-10-01 and ending on 2019-07-31. These hourly day-ahead electricity prices are then grouped together by date and averaged, yielding average daily prices, which are also called *base* prices on the market.

The series of load consists of the German actual load data starting from 2015-01-06 and ending on 2019-07-31, grouped by data and averaged into daily average load.

The wind production consists of both German onshore and offshore wind, in the same date range as above. The two series are summed together hourwise, and the resulting series is grouped by date and averaged to form daily average wind production. For observations where the value for either onshore or offshore was missing, the sum is set to simply be the non-missing value. For observations where both onshore and offshore had missing values, the sum is set to also be missing. Missing values are removed when averaging the observations, and the resulting daily series has itself has two missing values: observations for 2016-10-28 and 2016-12-09. These are estimated from surrounding values using simple linear interpolation and added in place of the missing values.

Solar production is likewise averaged over daily groups, and missing observations are ignored when calculating this average. For the same dates as for wind production, 2016-10-28 and 2016-12-09, all hourly observations are missing, leading to two missing values in the averages. These two values are also estimated with linear interpolations. In addition, for the dates 2015-02-28 and 2016-06-01, the observations for all but one night hour are missing, leading to zero values for the averages. These zero values are also replaced by linear interpolations of surrounding values, as they are only zero due to missingness, and the solar data is otherwise quite regular.

The four time series are plotted on Figure 4.1, and some summary statistics are shown in Table 4.1. In the following sections, when estimating model parameters from the data, we restrict all four variables to the date range 2015-06-01 – 2018-12-31. The remaining observations, 2019-01-01 – 2019-07-31, are used for comparison with simulated results later on.

	Spot	Load	Solar	Wind
Mean	34.87	55948.74	4186.46	10392.41
Median	33.56	57079.17	4150.41	8263.86
Std. Error	13.30	6506.29	2696.19	7556.22
Min	-52.11	38398.25	207.93	715.57
Q1	27.70	51253.14	1592.87	4577.15
Q3	41.46	60718.84	6414.03	14094.18
Max	101.92	68944.61	10737.04	41825.95

Table 4.1: *Summary statistics of the four aggregated ENTSO-E time series*

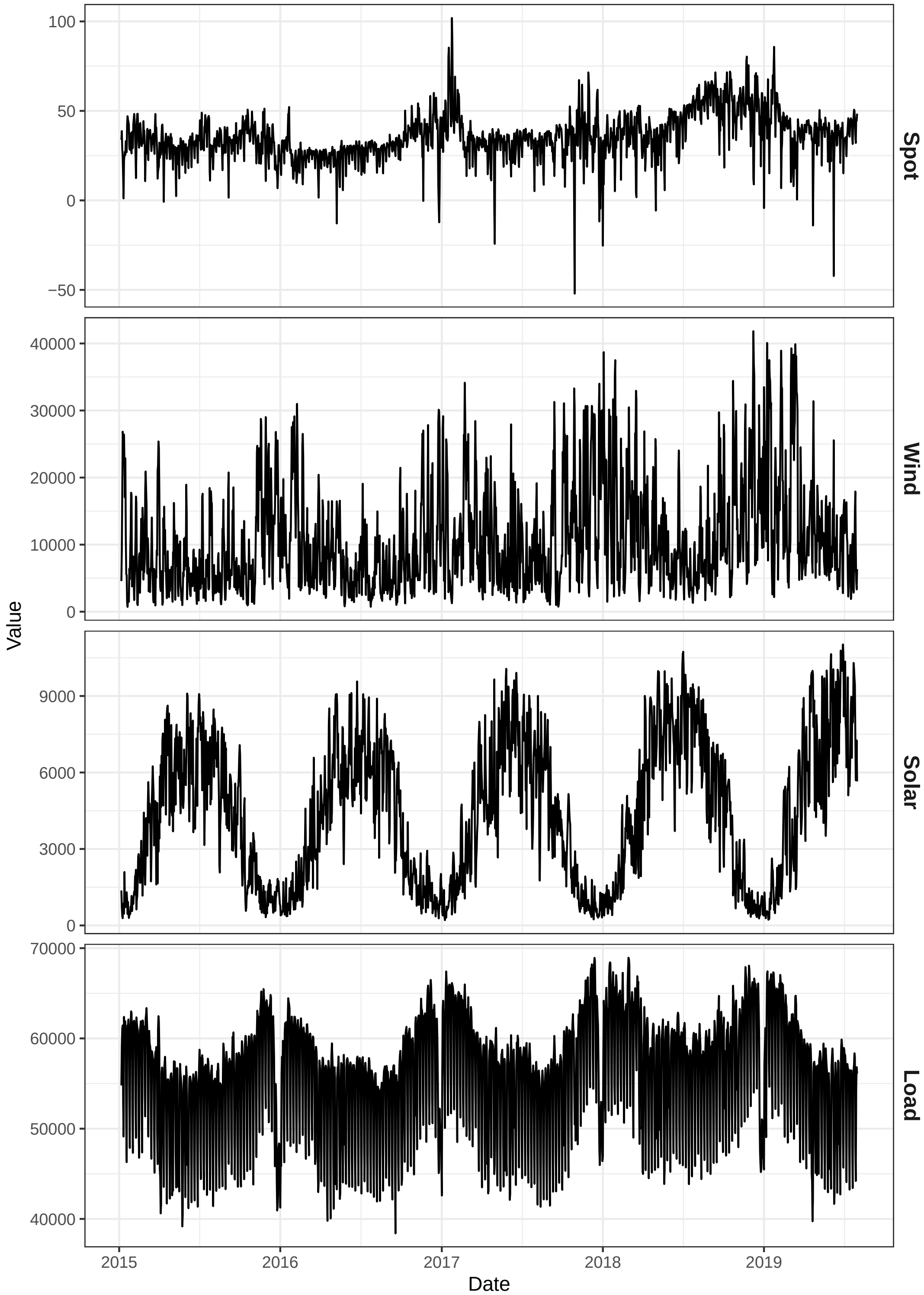


Figure 4.1: *Daily aggregated time series from ENTSO-E*

5 Modelling

In this chapter we go through the steps and results involved in modelling the joint distribution for the day-ahead base price, the load, and the solar and wind production. Powered by Sklar's Theorem, we fit marginal models for each of the four variables individually in the first section, and then link them together with a vine copula in the second.

5.1 Marginal Analysis

For the marginal models of each variable, we apply a two-step filter involving two models. Denoting the modelled variable at time t as Y_t, we consider models on the form

$$Y_t = s_t^Y + X_t^Y, \tag{5.1}$$

where s_t^Y is the deterministic part of Y, and X_t^Y is the stochastic part. The two parts will be modelled as follows:

Seasonal model: to account for deterministic variations in the data due to seasons, first a linear model is fitted using sine-cosine pairs of various frequencies and indicators, and the residuals are extracted to continue the filtering process in the serial model. This model also filters out a linear time trend and level.

Serial model: to account for serial correlation and heteroskedasticity in volatility, an ARMA-GARCH model is fitted on the residuals of the seasonal model, and the standardised residuals are extracted and transformed to uniform variables using the estimated ARMA-GARCH distribution.

For the seasonal part, which depends only on the time index $t = 1, \dots, T$, the term is given by

$$s_t^Y = \alpha_0 + \alpha_1 \cdot t + \sum_{\phi \in \Phi} \Big(\beta_{1,\phi} \sin\big(2\pi t\phi\big) + \beta_{2,\phi} \cos\big(2\pi t\phi\big) \Big) + \sum_{D \in \boldsymbol{D}} \gamma_D D(t), \tag{5.2}$$

where $\Phi \subseteq \mathbb{I}$ is a set of frequencies, and $\boldsymbol{D}$ is a set of *dummy variables*, here denoted as functions $D(t)$, such that $\operatorname{ran} D = \{0, 1\}$ for all $D \in \boldsymbol{D}$. We consider the frequency set $\Phi = \{1/365, 2/365\}$, corresponding to annual and semiannual cycles, and for dummy variables, we consider weekday indicators and holiday/workday indicators as well as their interaction terms. A day is counted as a holiday if either

(a) it's an official German holiday,

(b) it's an official German non-working day, or

(c) it's between Boxing Day and New Years Day, i.e. December 27–31

The reason for including holidays—and furthermore to count the Christmas week as such—is exactly that we are trying to capture predictably differering levels due to human activity, especially in consumption. On holidays, fewer companies will be consuming at their usual level, due to employees having those days off, and as such their production—and therefore also power consumption—will be expected to be at a lower level. This can be seen quite clearly on **??**, where the consumption curve has notable dips every year between Christmas and New Years Eve. The interaction terms between the two sets of dummy variables are included to account for holidays that fall on weekends; the effects of weekdays and holidays are counted separately, but we do not expect a holiday falling on a weekend to have the same effect as a holiday falling on a weekday. Including the interaction terms takes this into account.

Note that even with the included variables, the seasonal model is a simplification. It does not, for example, take into account that single holidays falling on Thursdays might be followed by a pseudo-holiday the next day, as employees will be likely to spend a vacation day that Friday to get an extended weekend. It also disregards days that are not official holidays, but which would still see different activity levels, for example vacation weeks in schools. We discuss this a bit more when modelling the consumption.

The serial model is specified by an ARMA (autoregressive moving average) model in the mean,

$$X_t^Y = \sum_{j=1}^{p} \phi_j X_{t-1}^Y + \sum_{k=1}^{q} \theta_k \epsilon_{t-1} + \epsilon_t, \tag{5.3}$$

where ϵ_t follows a GARCH (generalised autoregressive conditional heteroskedasticity) model, i.e. $\epsilon_t = \sigma_t z_t$ where σ_t depend on the type of GARCH model and z_t are realisations of i.i.d. random variables with zero mean and unit variance. A lot of different ARMA-GARCH models are available to choose from, due to model order of the ARMA component, model type and order of the GARCH component, and distribution of z_t. Below are given short definitions of a few types of GARCH models:

Definition 5.1 (GARCH)**.** The GARCH(q,p) model specifies the conditional variance as

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^{p} \beta_k \sigma_{t-k}^2. \tag{5.4}$$

Definition 5.2 (GJR-GARCH)**.** The GJR-GARCH(q,p) model specifies the conditional variance as

$$\sigma_t^2 = \sum_{j=1}^{q} \left(\alpha_j + \gamma_j \mathbb{1}(\epsilon_{t-j} < 0)\right) \epsilon_{t-j}^2 + \sum_{k=1}^{p} \beta_k \sigma_{t-k}^2. \tag{5.5}$$

Definition 5.3 (E-GARCH)**.** The E-GARCH(q,p) model specifies the conditional variance as

$$\log \sigma_t^2 = \omega + \sum_{j=1}^{q} \left(\alpha_j z_{t-j} - \gamma_j \left(|z_{t-j}| - \mathbb{E}|z_{t-j}|\right)\right) + \sum_{k=1}^{p} \beta_k \log(\sigma_{t-k}^2). \tag{5.6}$$

Note that in the specification for the E-GARCH model the mean of the absolute value of the conditional residuals appear. The values of these for the distributions we are going to use are given as follows:

$$Z \sim N(0,1) \implies \mathbb{E}|Z| = \sqrt{\frac{2}{\pi}}, \tag{5.7}$$

$$Z \sim t(\nu) \implies \mathbb{E}|Z| = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\sqrt{\frac{\nu-2}{\pi}}, \tag{5.8}$$

where N and t denote the distribution functions of the normal and Student's t-distribution, respectively.

The latter two specifications model shocks to the conditional variance asymmetrically, which is known as the *leverage effect*. Note that, like for the ARMA model, the model order of these are also denoted with p and q, since this is the standard notation in the literature, and since we won't be using general orders of the models beyond this point. We consider the following model specifications:

- ARMA orders: $p, q \in \{0,1\}$
- GARCH orders: $p, q \in \{0,1\}$, such that at least one is nonzero
- GARCH types: standard GARCH, E-GARCH, GJR-GARCH
- Conditional distributions: Gaussian, Student's t

for a grand total of 72 different models per series. These are fitted with numerical maximum likelihood methods using the R package `rugarch` [Ghalanos, 2019]. For choosing a single "best" model among the model specifications, we use AIC, and in case this quantity can't be calculated—e.g. in case of non-convergence—that model specification is dropped from further considerations.

5.1.1 Electricity Prices

In financial mathematics, prices of assets are commonly considered through log-returns—for a sequence of prices, $(P_t)_{t\in\mathcal{J}}$, the log-returns are given by

$$r_t := \log\frac{P_t - P_{t-1}}{P_{t-1}} = \log P_t - \log P_{t-1}. \tag{5.9}$$

The returns represent how much an investor would earn or lose by buying one unit of the asset at time $t-1$ and selling it again at time t—this is convenient when considering the logarithms of the returns, since the total profit or loss incurred by buying at time t_0 and selling at time $t_0 + T$ is simply $\sum_{t=t_0}^{t_0+T} r_t$. While log-returns are therefore natural to consider for the prices of assets such as stocks or commidities like oil and gold, it makes little sense to talk about the "returns" of electricity, since it is a non-storable good.

In the literature, when modelling electricity spot prices, it is common to use log-prices rather than log returns for very spiky and non-negative prices, since it makes the spot price distribution more symmetric and less leptokurtic [Manner et al., 2019]. However, since the logarithm function is only defined on the positive real half-line, it cannot be used to transform negative prices, and spot prices on the German day-ahead electricity market exhibit occasional, negative spikes. As such, before applying a logarithmic transformation, all prices in the data set are offset by a constant $K > 0$ such that the "log-prices" considered become

$$p_t := \log\big(P_t + K\big), \quad \forall\, t \in \mathcal{J}. \tag{5.10}$$

Note that this constant is only really constant in the sense that the same value is used for all $t \in \mathcal{J}$, but this value will in fact depend on the data, as it must be at least large enough to offset the most negative prices.

Seasonal model Fitting the seasonal component is sensitive to outliers in the data, so as a prior step to fitting the model described in (5.2), we filter them out temporarily. Following [Benth et al., 2008], since the data is clearly not normally distributed, we use the following summary statistic to identify outliers: Given the lower and upper quartiles of $\{p_t\}_{t\in\mathcal{J}}$, q_1 and q_3, respectively, and the interquartile range $q_R = q_3 - q_1$, an observation p_t is classified to be an outlier, if

$$p_t < q_1 - 3q_R \quad \text{or} \quad p_t > q_3 + 3q_R. \tag{5.11}$$

The quartiles for the log-transformed prices yield the admissible interval $\big[6.19, 6.37\big]$ for the observations, making 9 of them outliers. These observations are removed from the seasonal fit only, thus resulting in a deterministic component visualised on Figure 5.1 and residuals used as data for the serial model on Figure 5.2.

Serial model Using the residuals from the seasonal model, we now fit a panel of ARMA-GARCH models as described above. An overview of the results can be seen in Table A.1 and A.2, where the model names are given such that the model type, model orders, and conditional distributions are specified, with N representing the normal distribution, and t representing Student's t-distribution.

The model with the lowest AIC is the ARMA(1,1)-E-GARCH(1,1)-t (model #72) with a value of -6.037. Other models have AIC values close to this, e.g. the AR(1)-E-GARCH(1,1)-t (model #70) with -6.034. The fact that the models providing the best fits are found among the models that can account for leverage fit well into our intuition about the behaviour of electricity prices—the conditional variance is more sensitive to negative shocks than positive ones. Moreover, the Student's t-distributed conditional error terms seem to be preferred, indicating further that the prices do in fact have heavier tail distributions.

For the purpose of this project, we need to choose a single model to continue with, and as such we select model #72 to be the "best" model. The coefficients for this model are given in Table 5.1 together with the robust (i.e. heteroskedasticity-consistent) standard errors of [White, 1982]. On Figure 5.3 the residuals from the seasonal model, used as input for this step, are plotted together with the fitted values, and on Figure 5.4, the standardised residuals from the model are plotted.

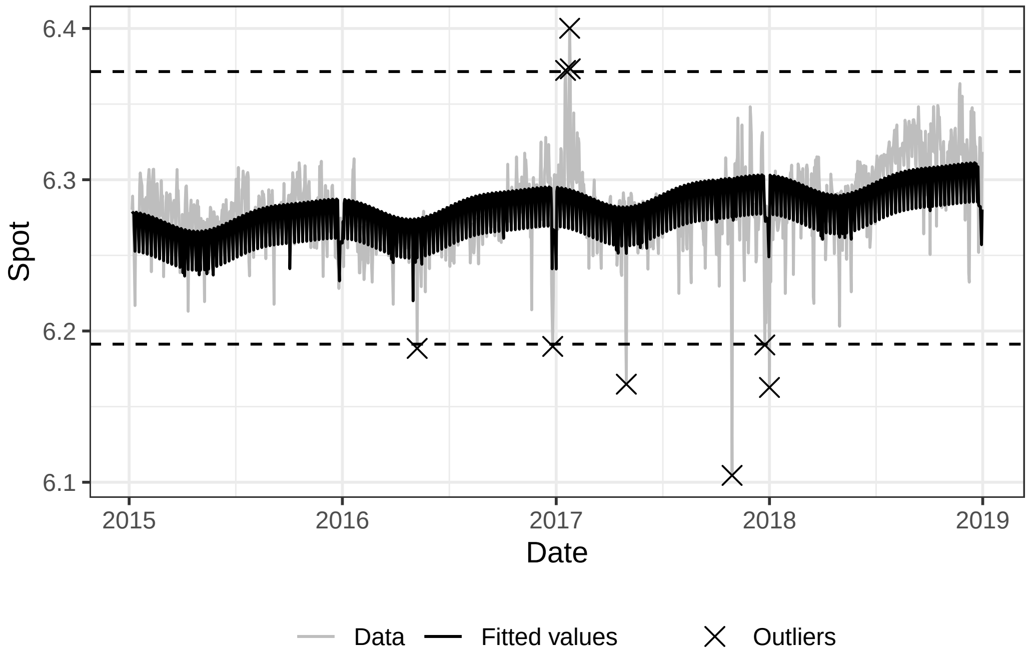


Figure 5.1: *Plot of the fitted seasonal model for the spot price data. The dashed lines represent the bounds used to identify outliers, which are marked with* ×.

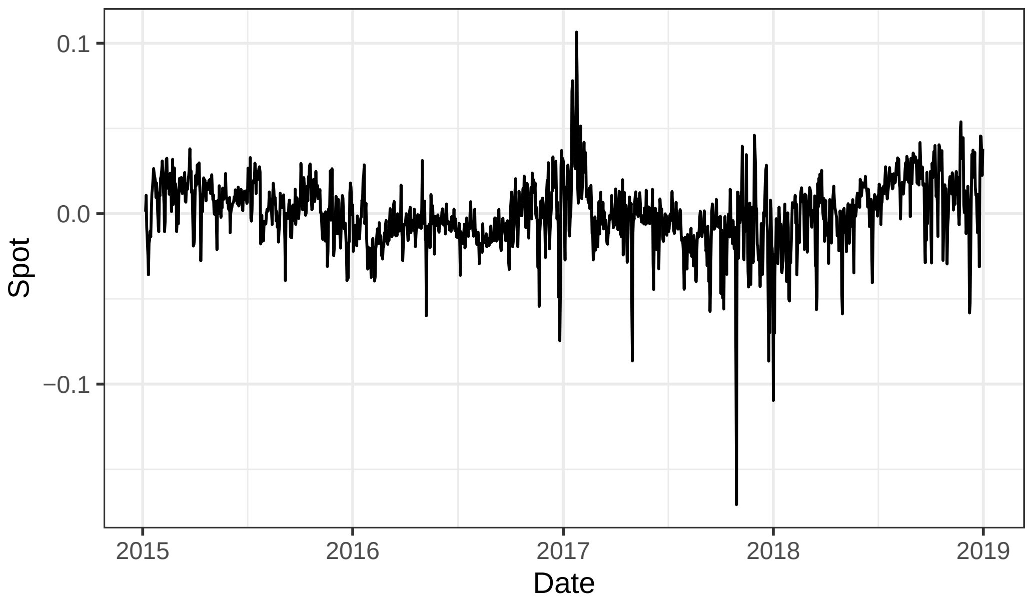


Figure 5.2: *Plot of the residuals of the seasonal spot price model*

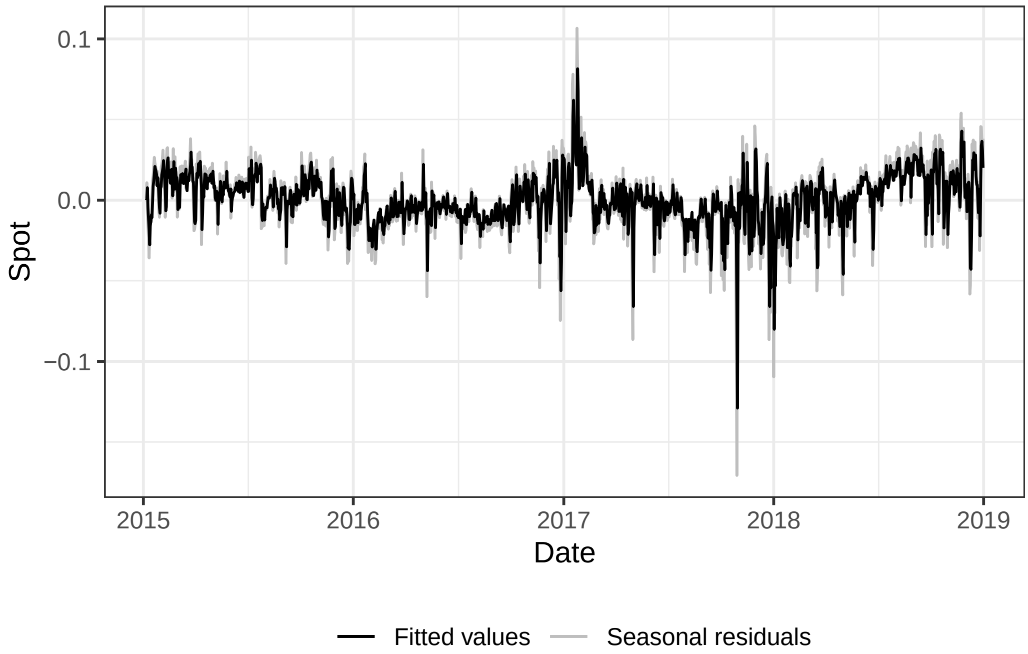


Figure 5.3: *Fitted values of the chosen model plotted together with the residuals of the seasonal model, on which it is fitted*

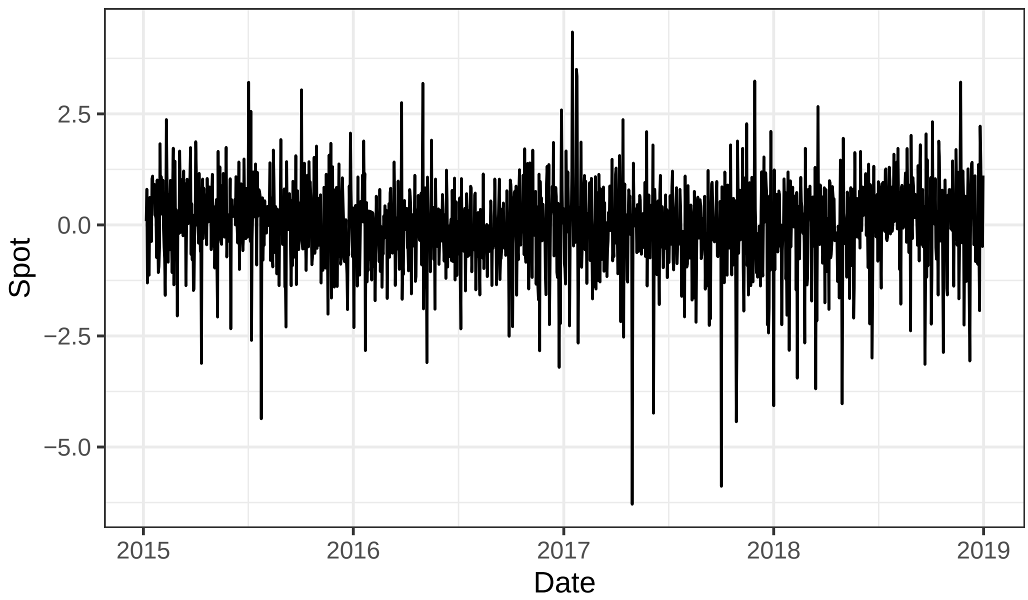


Figure 5.4: *Standardised residuals of the chosen model*

	Estimate	Std. Error	t-value	$\mathbb{P}(>\|t\|)$
ϕ	0.8395	0.0256	32.8571	0.0000
θ	-0.1371	0.0461	-2.9729	0.0030
ω	-0.7864	0.2134	-3.6847	0.0002
α	-0.0912	0.0397	-2.2997	0.0215
β	0.9106	0.0243	37.4030	0.0000
γ	0.3712	0.0507	7.3264	0.0000
ν	4.7286	0.5655	8.3614	0.0000

Table 5.1: *Coefficents of the chosen ARMA-GARCH model for the spot data*

5.1.2 Consumption

While the prices of electricity is driven by many factors reflected in trading activity, and production of renewable energy is subject to the uncertainty of the weather, one would intuitively expect consumption of electricity to be decently deterministic. Given a bit of expert knowledge about the schedule of business operations in Germany, it seems reasonable to assume that one could account for must of the structure in the data and be left with a white noise process. Unfortunately, aquiring such knowledge is out of scope for this project—there are whole companies in the business of providing this kind of data—so we instead apply the same seasonal filter as for the price data and try to capture the rest of the structure with a time series model.

As noted in Table 4.1, the load data is always positive, so we don't have to add any constants to be able to perform a logarithmic transformation on it.

Seasonal model Following the same method for fitting the seasonal model as for the spot price data, we obtain the model fit described on Figure 5.5 and 5.6. Note that the aforementioned notable dips in consumption around Christmas each year are somewhat captured in the modelled path for s_t^C, with Christmas 2015 and 2018 not being fully explained by the holiday dummies.

Serial model As before, we use the residuals from the model as input data in the serial model. We fit the same panel of ARMA-GARCH models as we did for the spot price data, with an overview of these summarised in Table A.3 and A.4.

With an AIC of -4.713, the model we choose to continue with is model #60, which is an ARMA(1,1)-GJR-GARCH(1,1)-t. Its coefficients and standard errors are given in Table 5.2.

The residuals from the seasonal model are plotted together with the fitted values from the serial model on Figure 5.7, and the standardised residuals are plotted on Figure 5.8.

5.1.3 Solar Power Generation

Like with the load data, the average daily solar power generation is strictly positive, and as such we can use a logarithmic transformation to dampen the variance in the data.

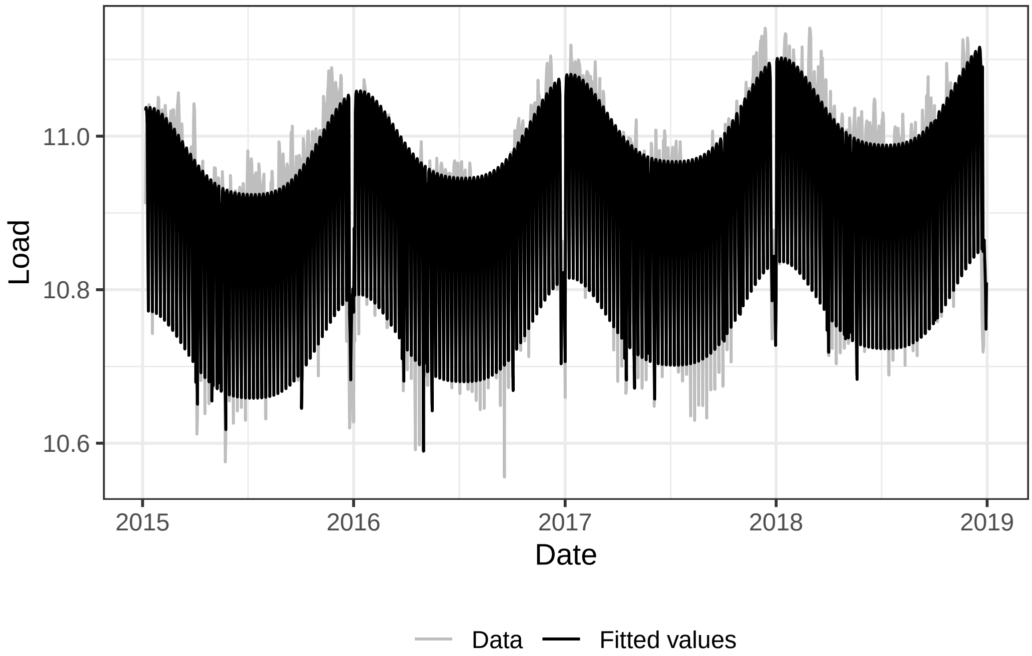


Figure 5.5: *Plot of the fitted seasonal model for the load data*

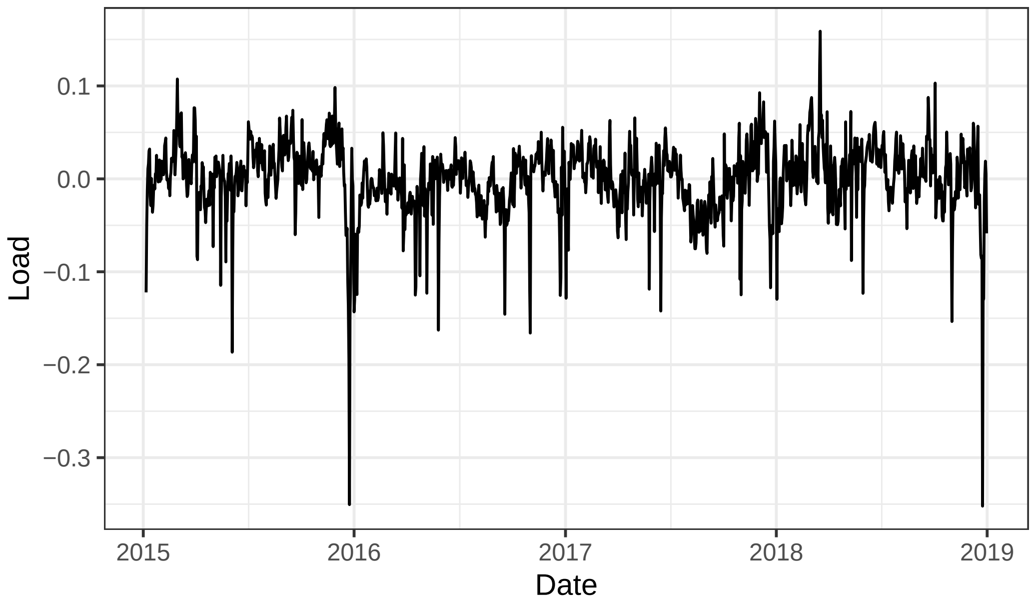


Figure 5.6: *Plot of the residuals of the seasonal load model*

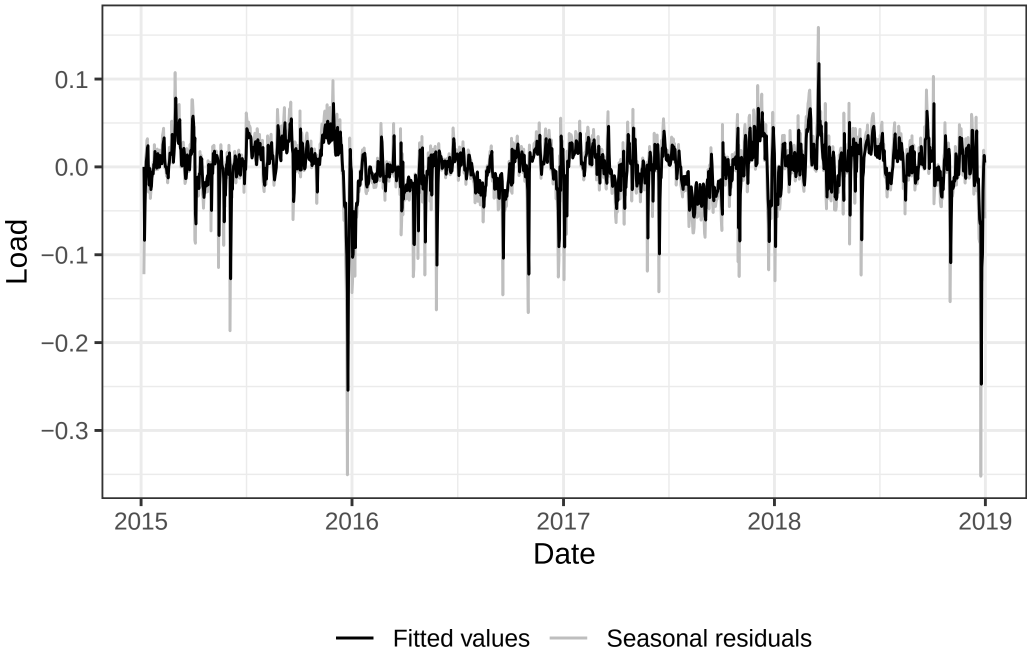


Figure 5.7: *Fitted values from the selected model plotted together with the residuals from the seasonal model*

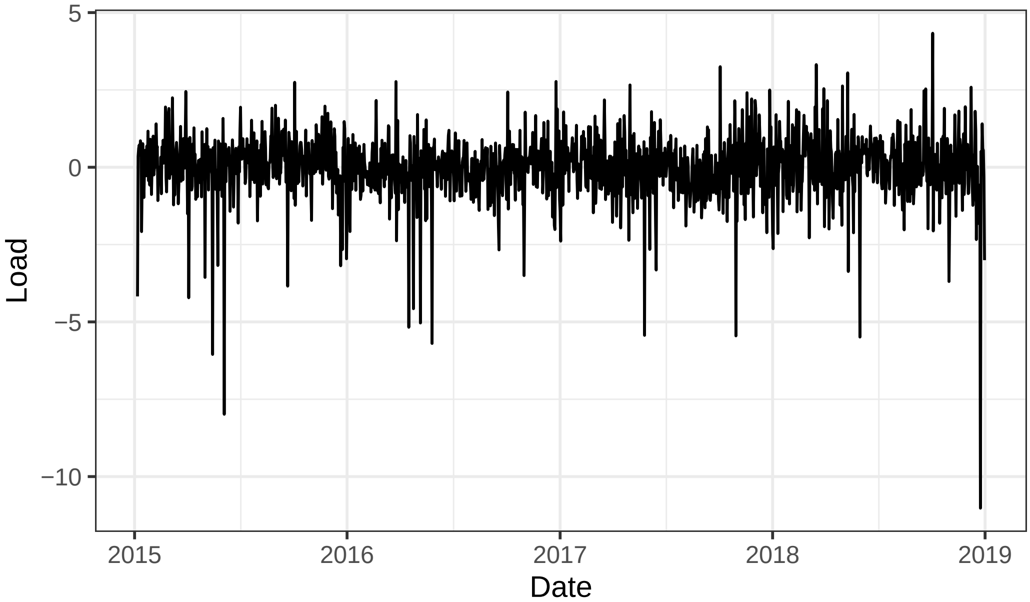


Figure 5.8: *Plot of the standardised residuals from the serial model for the load data*

	Estimate	Std. Error	t-value	$\mathbb{P}(>\lvert t\rvert)$
ϕ	0.7854	0.0258	30.4282	0.0000
θ	-0.1013	0.0447	-2.2666	0.0234
ω	0.0003	0.0000	6.9601	0.0000
α	0.3484	0.0775	4.4979	0.0000
β	0.1753	0.0508	3.4543	0.0006
γ	0.2652	0.1385	1.9146	0.0555
ν	3.9361	0.4330	9.0903	0.0000

Table 5.2: *Coefficents of the chosen ARMA-GARCH model for the consumption data*

Seasonal model While electricity prices and power consumption are subject to variance due to human behaviour and schedules, solar power generation is subject only to the natural phenomenon of sunlight, and as such, it makes little sense to include the weekday or holiday variables in the seasonal model for solar power generation. Hence, the seasonal term filtered from the log-series is

$$s_t^S = \alpha_0 + \alpha_1 \cdot t + \sum_{\phi\in\Phi} \Big(\beta_{1,\phi} \sin\big(2\pi t\phi\big) + \beta_{2,\phi} \cos\big(2\pi t\phi\big)\Big), \tag{5.12}$$

still with $\Phi = \{1/365, 2/365\}$. The fitted values and residuals are plotted on Figure 5.9 and 5.10. The seasonal variation in solar production is fairly regular on an annual cycle, with production unsurprisingly peaking in the summer months and slumping in the winter months, with little to no trend over the years.

Serial model One can argue that modelling volatility of solar power production with a GARCH process doesn't intuitively make sense. The usual interpretation of such a model is that shocks are followed by an increase in volatility that persists for some time. Looking at the residuals on Figure 5.10, they appear to exhibit higher volatility during winter months and lower volatility during summer months. This fits our expectation, since the longer windows of strong sunshine on summer days makes for a fairly high baseline of production, while during winter, there are fewer hours of sunlight in a day, so if the sky is overcast for a few hours, this has a large impact on the total production for that day.

It might therefore make better sense to consider a model with deterministic—but still time-varying—volatility for the solar power production, possibly following some sinusoidal curve with peaks during winters. In this project, however, we continue with a GARCH process for the volatility, simply because established software for estimating such models already exist, and because it will likely still be a good enough fit.

The usual panel of ARMA-GARCH models is therefore fitted on the residuals from the seasonal fit, and an overview of each fit is summarised in Table A.5 and A.6. The model with the lowest AIC (0.5616) is model #58, an AR(1)-GJR-GARCH(1,1)-t, with coefficients given in Table 5.3. Worth noting here is that the leverage parameter γ is both negative and bordering on insignificant. The sign implies that volatility actually decreases following a negative shock.

The series of fitted values are plotted together with the residuals of the seasonal model on Figure 5.11 and the residuals are plotted on Figure 5.12.

	Estimate	Std. Error	t-value	$\mathbb{P}(>\|t\|)$
ϕ	0.4975	0.0282	17.6228	0.0000
ω	0.0014	0.0006	2.3965	0.0166
α	0.1009	0.0199	5.0838	0.0000
β	0.9174	0.0153	60.1446	0.0000
γ	-0.0501	0.0256	-1.9569	0.0504
ν	13.4136	4.1731	3.2143	0.0013

Table 5.3: *Coefficents of the chosen ARMA-GARCH model for the solar data*

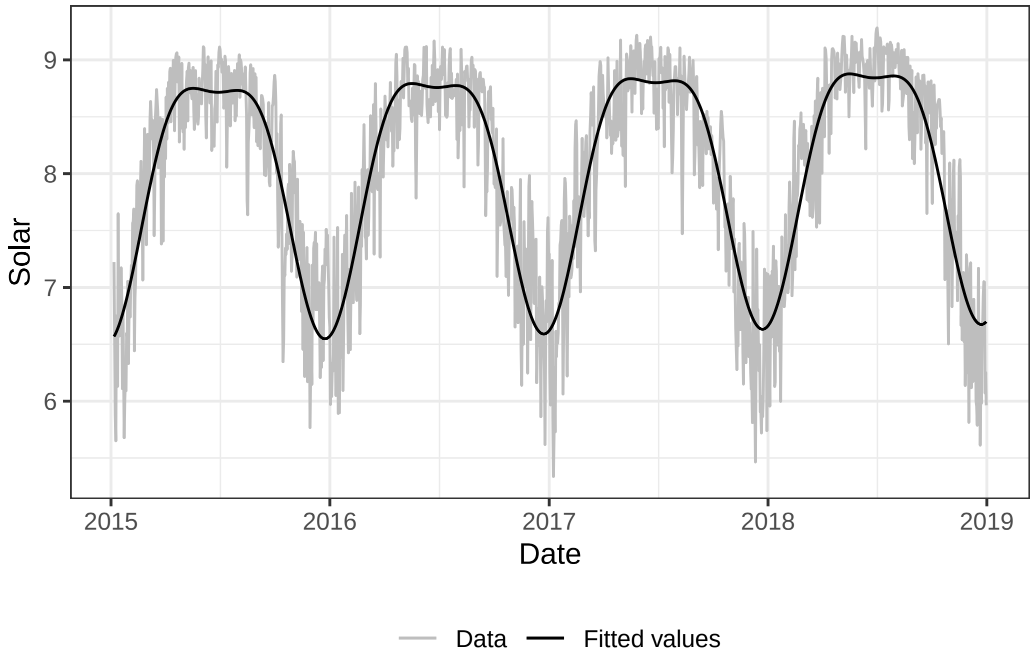


Figure 5.9: *Plot of the fitted seasonal model for the solar power generation data*

5.1.4 Wind Power Production

As with the load and solar power production data, the wind power production data is strictly positive and can be logarithmically transformed.

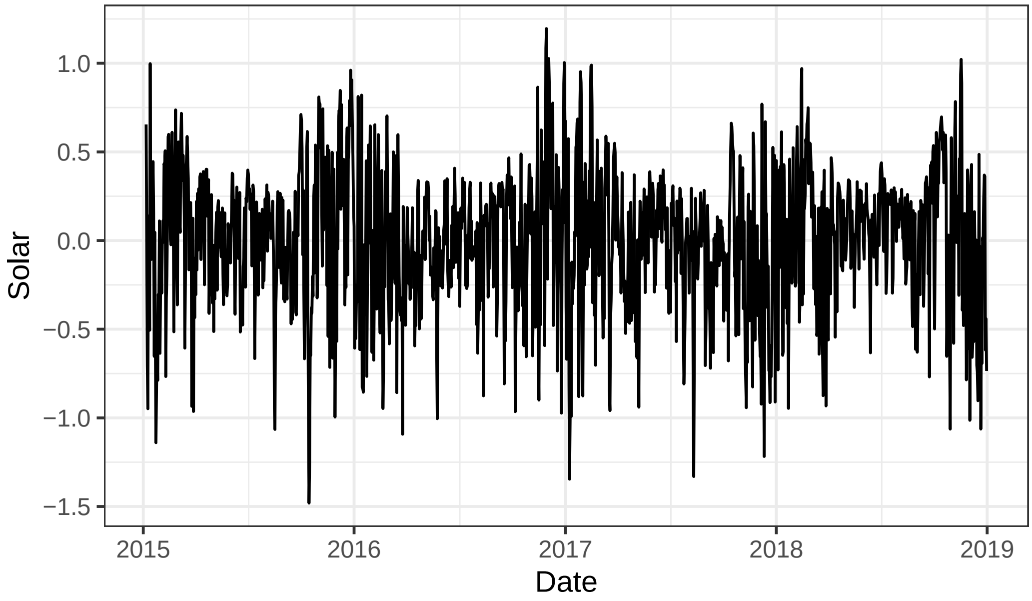


Figure 5.10: Plot of the residuals of the seasonal solar model

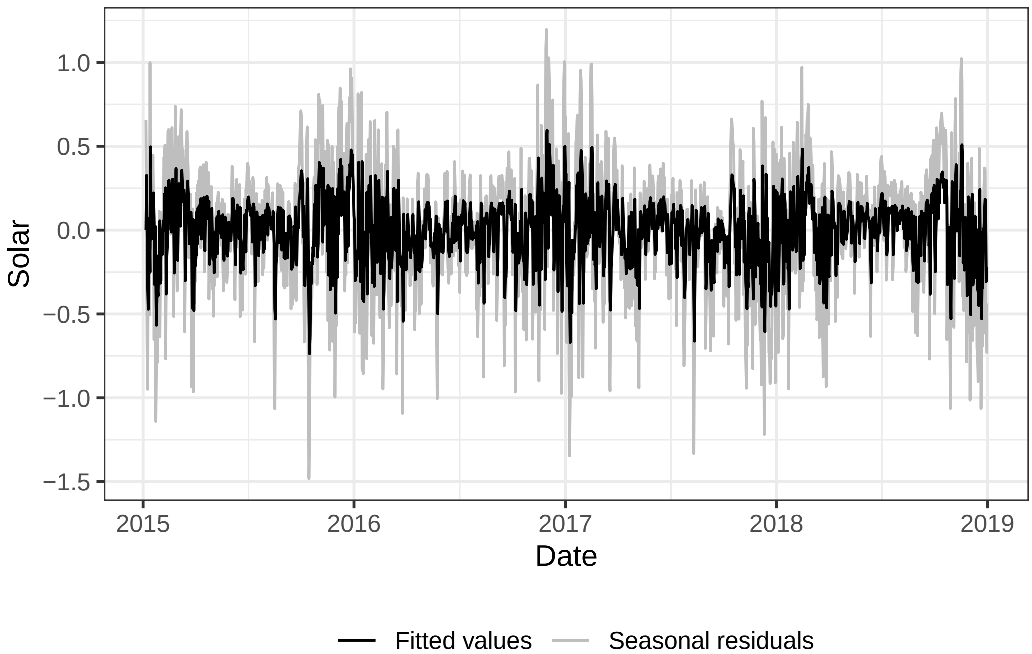


Figure 5.11: *Fitted values from the selected model plotted together with the residuals from the seasonal model*

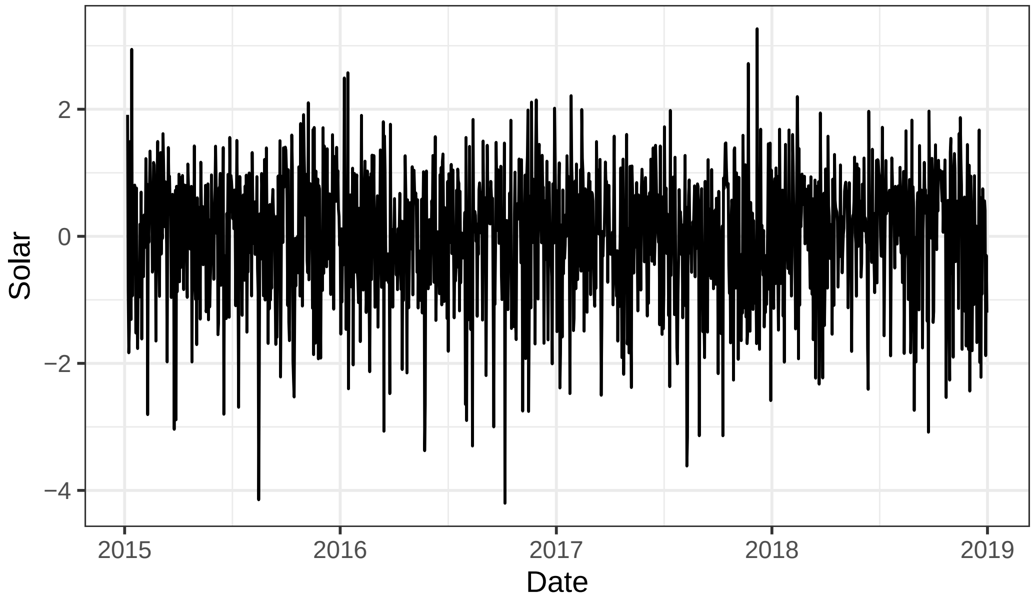


Figure 5.12: *Plot of the standardised residuals from the serial model for the solar data*

Seasonal model Wind power production, like solar power production, is not subject to human scheduling in the same way that prices and consumption is, so we also omit the weekday and holiday dummy variables from the seasonal wind power model. The fitted values and residuals are plotted on Figure 5.13 and 5.14, respectively.

Serial model After fitting the panel of ARMA-GARCH models to the residuals of the seasonal model, we obtain the results summarised in Table A.7 and A.8. Model #28, an ARMA(1,1)-E-GARCH(1,0)-N, is chosen as the best model with an AIC of 1.7029, and the parameters are listed in Table 5.4.

	Estimate	Std. Error	t-value	$\mathbb{P}(>\lvert t\rvert)$
ϕ	0.4915	0.0298	16.5040	0.0000
θ	0.1836	0.0315	5.8243	0.0000
ω	-1.1383	0.0342	-33.2343	0.0000
α	-0.3307	0.0413	-7.9995	0.0000
γ	-0.2327	0.0557	-4.1769	0.0000

Table 5.4: *Coefficents of the chosen ARMA-GARCH model for the wind data*

Unlike the three other variables, the chosen model for this variable has a Gaussian conditional distribution, and in all the fitted models with Student's t-distributed conditional residuals, the degrees of freedom are estimated to be very high (close to or equal to 100, which is upper bound for the parameter during the optimisation process). Its fitted values are plotted on Figure 5.15 and its residuals on Figure 5.16.

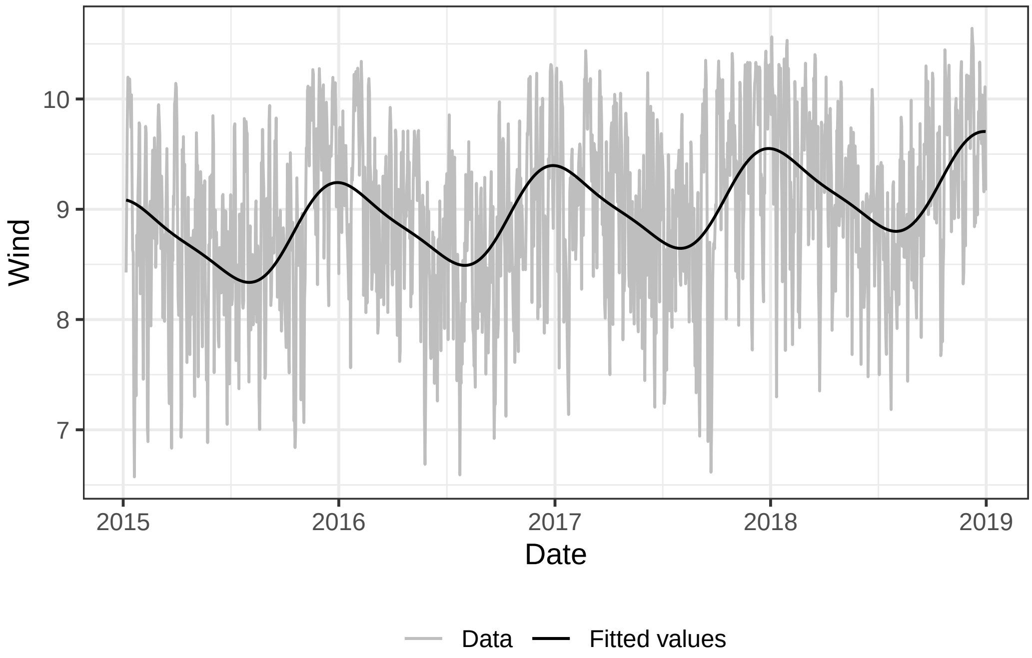


Figure 5.13: *Plot of the fitted seasonal model for the wind power generation data*

5.2 Joint Analysis

Now that we have estimated marginal models for each of the four variables, we have all we need to start linking them together with a copula model. We estimate an R-vine model for the joint distribution as follows:

1. Estimate sequentially the tree structure with a maximum spanning tree algorithm using empirical Kendall's τ's as edge weights in the complete graph

2. Test each edge for independence

3. If not independent, fit a panel of pair-copulas consisting of Gaussian, Student's t, Clayton, Gumbel, and Frank copulas, as well as their rotations

4. Once all the tree structures and parameters have been estimated, assume the structure is correct and reestimate the parameters with maximum likelihood using the sequential estimates as starting values

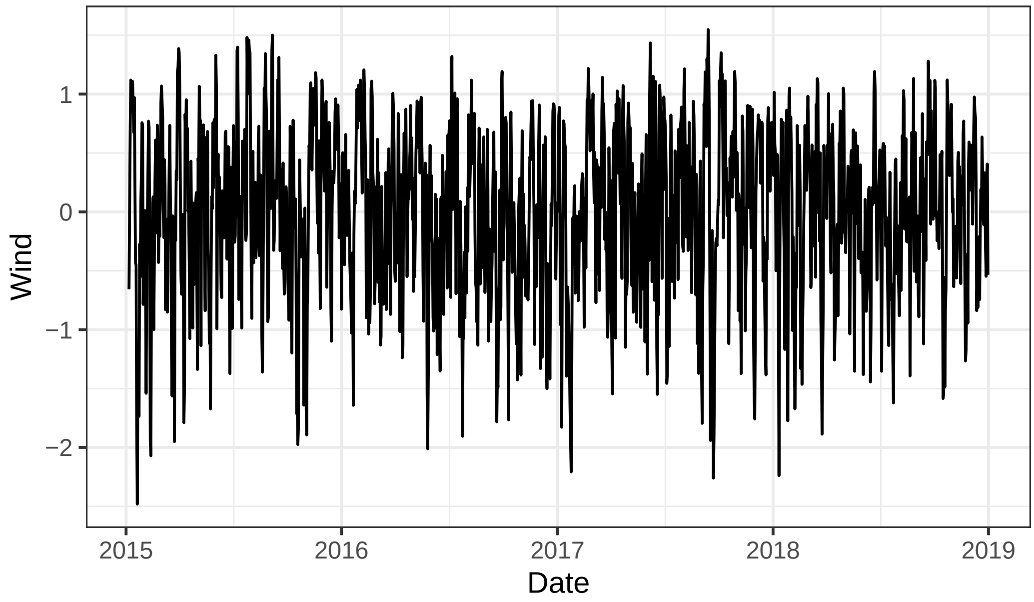


Figure 5.14: *Plot of the residuals of the seasonal wind model*

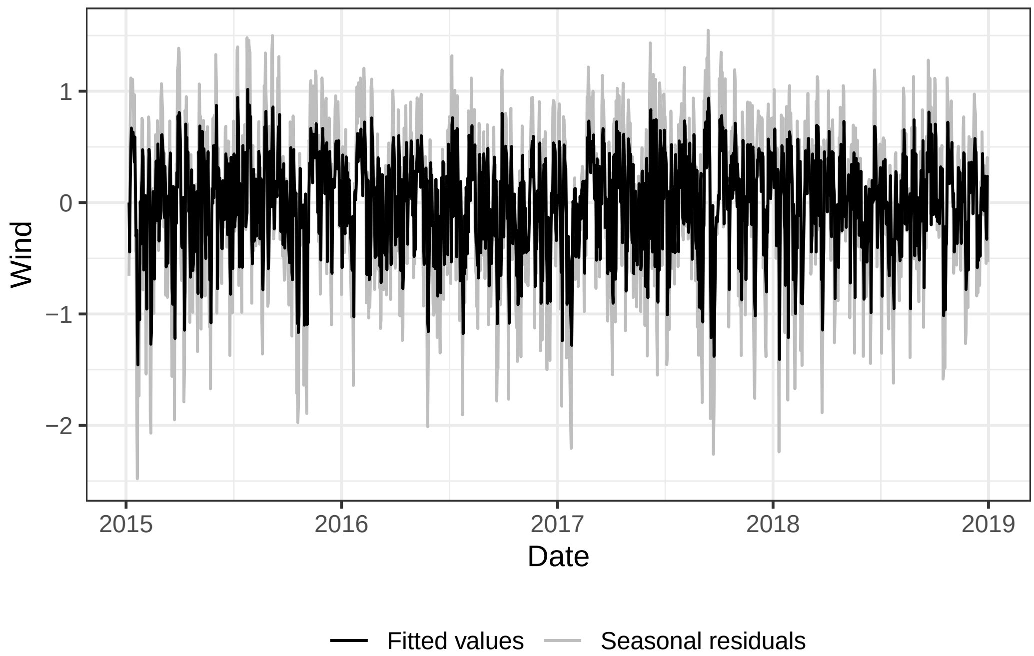


Figure 5.15: *Fitted values from the selected model plotted together with the residuals from the seasonal model*

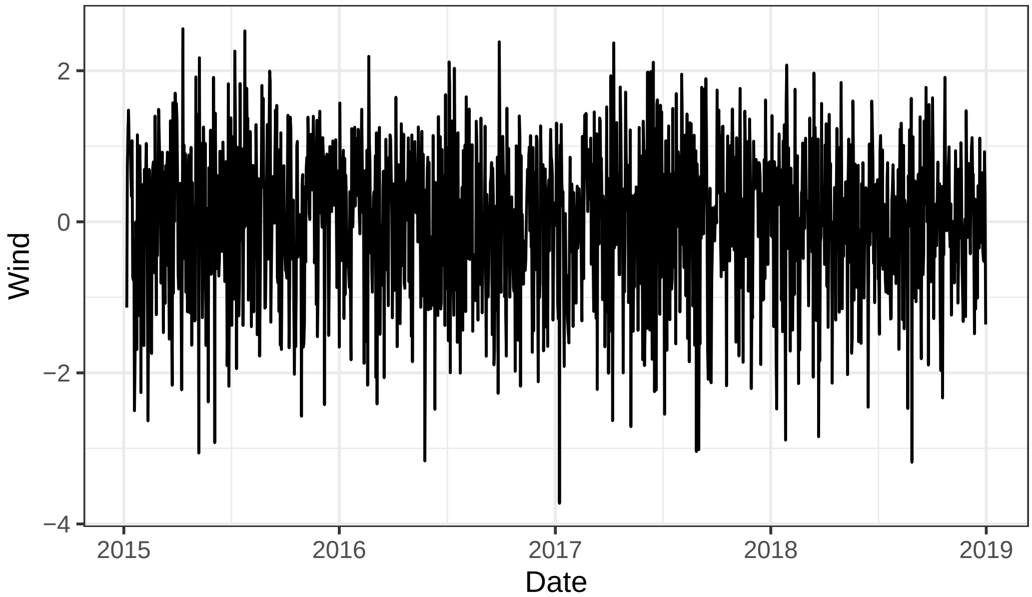


Figure 5.16: *Plot of the standardised residuals from the serial model for the wind data*

The resulting structure is summarised on Figure 5.17, with the selected copulas and estimated parameters indicated on each edge, and the contours of the pair-copula densities are plotted on Figure 5.18.

This structure corresponds to a C-vine with the wind power production (here enumerated as variable 4) as the root node. As expected, the dependence between the spot prices and the wind production is negatively sloped, as is the dependence between spot prices and solar production (conditional on wind production), which fits well with our intuition about how the prices are set.

Rotated copulas are chosen for two of the edges due to negative Kendall's τ's. Because these are 90 and 270 degree rotations, these naturally have no tail dependence coefficents, as defined in Definition 2.28, but one can consider similar quantities, e.g.

$$\lambda_{U,L}(X,Y) := \lim_{t\downarrow 0} \mathbb{P}\left(Y \leq F_Y^{(-1)}(t) \,\middle|\, X > F_X^{(-1)}(1-t)\right), \tag{5.13}$$

which compares the upper tail of X with the lower tail of Y—let's call it *right rotated tail dependence*, as it describes the mass concentrated in the lower right corner of the contours of (X,Y). One can similarly define a *left rotated tail dependence*, $\lambda_{L,U}$, which describes the upper left corner mass of the contours.

The actual values of these can be derived from the standard versions of the copulas they're describing. For the edge $(4,1)$ in T_1, we have a 270-degree rotated Gumbel copula with parameter $\alpha^{270} = -1.76$, which has mass in the lower right of the contours. The right rotated tail dependence coefficent, $\lambda_{U,L}$, of this copula then corresponds to the upper tail

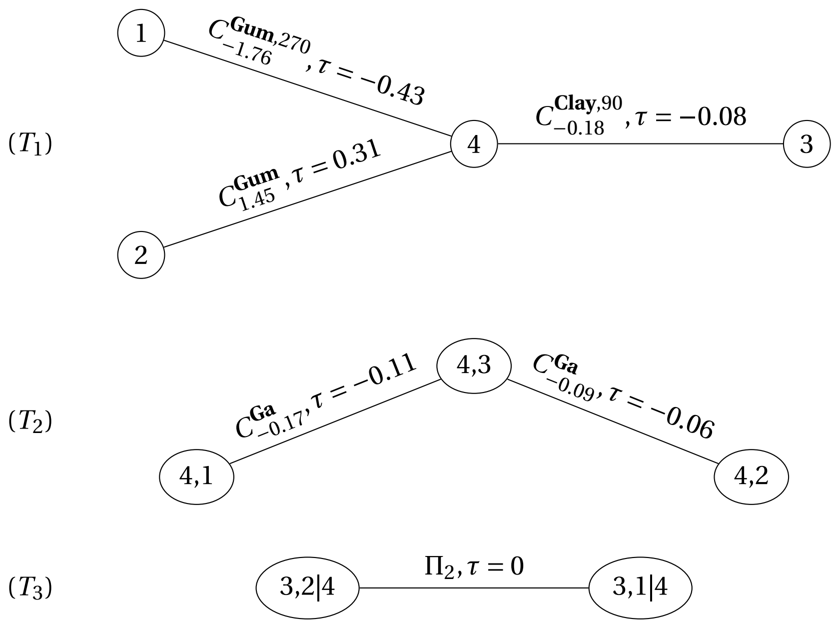


Figure 5.17: *Estimated vine structure, with copulas and Kendall's τ's given as edge labels. Variables are labelled by numbers such that Spot = 1, Load = 2, Solar = 3, and Wind = 4.*

dependence coefficent, λ_U, of the corresponding standard Gumbel copula with parameter $\alpha = -\alpha^{270} = 1.76$, i.e.

$$\lambda_{U,L}\left(C^{\mathbf{Gum},270}_{-1.76}\right) = \lambda_U\left(C^{\mathbf{Gum}}_{1.76}\right) = 2 - 2^{1/1.76} = 0.517. \tag{5.14}$$

The interpretation of this coefficent fits well with our understanding of the market: when wind production (variable 4) is in the right tail end of its distribution, the spot prices (variable 1) is very likely to be in the left tail end of its distribution. In other words, high levels of wind production pushes the spot prices down.

The other rotated copula, found on the edge $(4,3)$ (wind, solar) in T_1, is a 90-degree rotated Clayton copula with parameter $\alpha^{90} = -0.18$, which like the 270 degree rotated Gumbel copula also has mass in the lower right, since the standard Clayton copula has lower tail dependence. Its right rotated tail dependence coefficent therefore corresponds to the lower tail dependence coefficent, λ_L, for the unrotated version with parameter $\alpha = -\alpha^{90}$, i.e.

$$\lambda_{U,L}\left(C^{\mathbf{Clay},90}_{-0.18}\right) = \lambda_L\left(C^{\mathbf{Clay}}_{0.18}\right) = 2^{-1/0.18} = 0.021. \tag{5.15}$$

The interpretation is also similar to the one above, albeit with a much smaller size of coefficient: when the wind production is very high, there is a small probability that the solar production is very low. The explanation for this effect could be that windy weather is slightly correlated with weather conditions that block out the sun, and perhaps if the productions we were considering were more local and not averaged over days, this effect would be more pronounced.

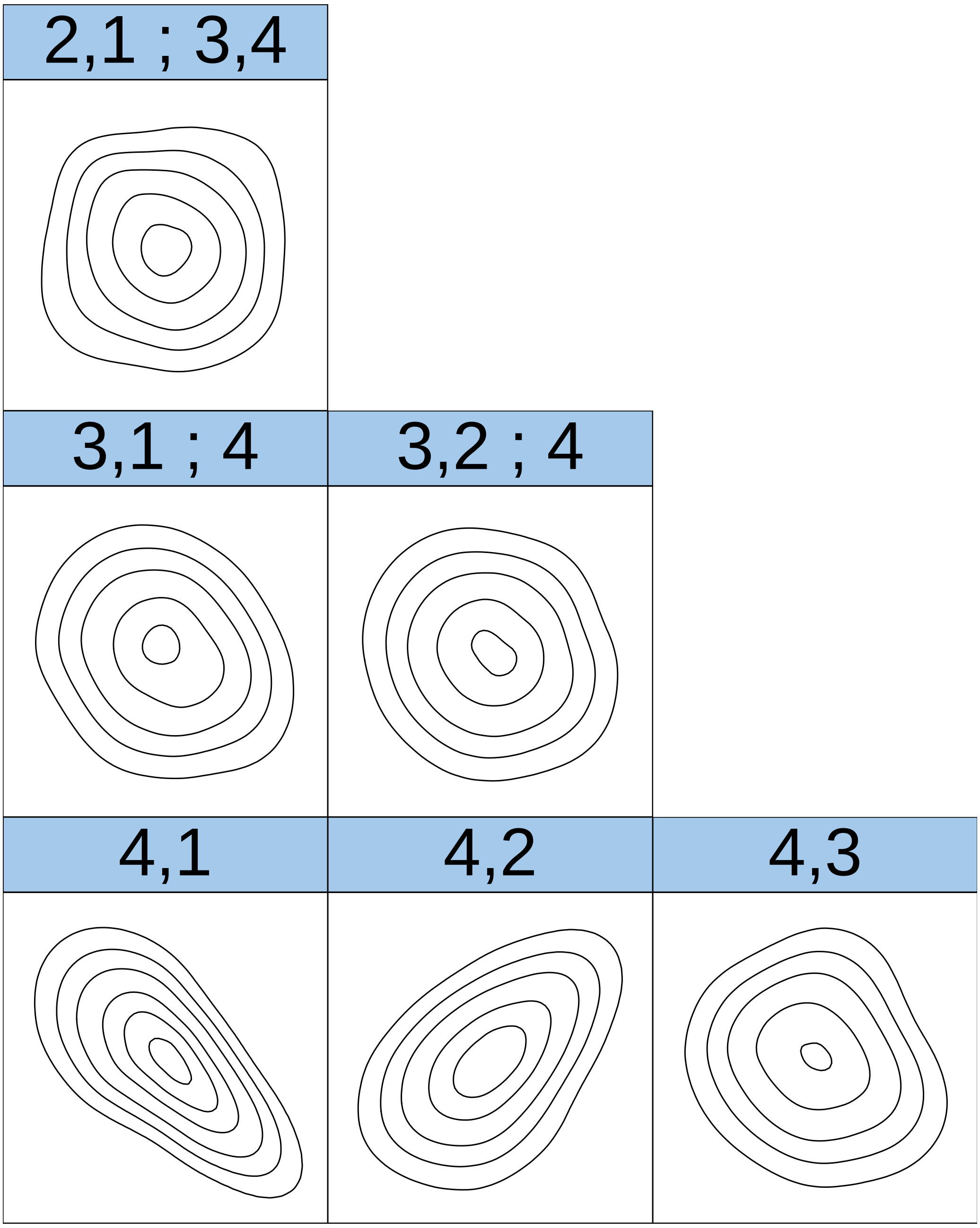


__Figure 5.18:__ Kernel density contours of each copula, given the data

A rather surprising finding in this model is in the edge $(4,2)$, which describes the dependence between wind production and consumption. Here we obtain a Gumbel copula with $\alpha = 1.45$, which implies an upper tail dependence coefficent of

$$\lambda_U\left(C_{1.45}^{\mathbf{Gum}}\right) = 2 - 2^{1/1.45} = 0.387. \tag{5.16}$$

That is, according to the model, in very high wind production scenarios, we are also about 39% likely to also be in a very high consumption scenario. This goes a bit against our intuition—how can extreme wind production levels affect the consumption levels? One possible explanation is that there are industrial consumers on the market who are somehow sensitive to the electricity spot price, and therefore ramps up their production of goods—and thus consumes more electricity—when the spot price is low.

An example of this could be owners of *pumped storage* hydroelectricity plants. These facilities have large reservoirs of water at different elevations, which can either produce electricity by letting the water run from the high reservoir to the lower or consume electricity by pumping the water up the opposite direction, effectively storing the power as potential energy. In a scenario with high wind production, the spot prices are simultaneously likely to be very low, or even negative, which acts as an incentive for the pumped storage owners to consume electricity at their plant and hope for prices to go up on a later day. Given enough of such plants or similar facilities, this effect could potentially be pronounced enough to produce this positive dependence, and according the [ENTSO-E, 2019], the total installed capacity for pumped storage is 9422 MW, which—if the effect is assumed to be symmetric and all of it is set to consume—would be enough to lift the consumption into a high level.

In T_2 in the vine, both copulas chosen for the edges are Gaussian, with parameters -0.17 and -0.09, respectively. Conditional on wind production, the solar production is negatively correlated with spot prices and consumption, but not very strongly, and with no tail dependence. This would indicate that out of the two sources of renewable energy, it's the wind production that's the strongest driving force in determining the spot prices. And moving to T_3, we get truncation, i.e. the independence copula Π_2 is chosen for the edge $(2,1|3,4)$, meaning that the spot prices and the consumption are conditionally independent given wind and solar production. This is again in line with our intuition, as we know the demand curve for electricity to be close to vertical, which corresponds to the consumption being independent of the prices, and whatever actual dependence there might have been is already captured by the dependence with wind and solar.

6 Portfolios of PPAs

We now present a scenario in which one can make use of being able to simulate the joint distribution of day-ahead prices, consumption, and solar and wind production. Imagine that we trade electricity in such a way that we do not own any generators ourselves, but we buy from e.g. power plant owners and sell to industrial consumers. The service we provide is two-fold:

(a) Buying power production from plant owners at a fixed price

(b) Selling power to consumers at a fixed price

The price is fixed through an instrument called a *power purchase agreement*, or PPA for short. Such contracts can have deliveries that are several years out in the future, but for the purpose of this example, we will consider PPAs on shorter windows of time.

6.1 Portfolio Setup

Assume the following setup. Two power providers—call them S and W, respectively—wish to enter into a PPA with us to get a fixed price on the electricity they produce. Provider S owns a large solar farm that accounts for 2% of all of Germany's solar power production, and W owns a wind turbine park that accounts for 5% of Germany's total wind power production, and we assume that their production is representative in such a way that those percentages are constant.[1] On the other hand, we have a large industrial consumer—call them C—who wish to enter into a PPA to pay a fixed price for electricity consumed. Consumer C accounts for 1% of all of Germany's consumption, and as with S and W, we assume this ratio to always be constant.[2] All three PPAs are signed on New Year's Eve 2018 with delivery in May through July 2019, i.e. it can be considered forward prices $F_c\left(t, T_1, T_2\right)$ with $T_1 = t + 121$ and $T_2 = 212$, representing delivery starting 121 days into the future and ending 91 days later (212 days into the future), for each of the counterparts $c \in \{C, S, W\}$.

[1] In a real scenario, regional differences would mean that this assumption could not possibly hold, since weather conditions (wind flow velocity, whether it's cloudy, etc.) are far from homogeneous over an area as large as the country of Germany. For now, one can imagine that these counterparts are not actually single counterparts, but comprise many smaller counterparts, who overall are assumed to be representative of the country.

[2] The three given sizes of our counterparts are arbitrarily chosen, but later we will investigate how to "choose" counterparts in such a way that the resulting payoff distribution is "best" in some sense.

6.2 Simulated Payoff Distributions

In order to answer questions about these PPAs, we use our model to draw $N = 1000$ samples of joint paths of length 212 of the four variables, denoting $\widehat{P}^i_t$, $\widehat{C}^i_t$, $\widehat{S}^i_t$, and $\widehat{W}^i_t$ to be the i'th time t simulated value for spot, consumption, solar, and wind, respectively, for $i = 1, \ldots, N$ and $t = T_1, \ldots, T_2$. Throughout this section, the simulation used to calculate quantities is the same.

The first obvious problem we need to answer is: what should F_C, F_S and F_W be? A first suggestion could simply be the unconditional mean of the day-ahead electricity prices for the period of delivery. Given our simulations, this price is

$$F_C = F_S = F_W = \frac{1}{N(T_2 - T_1)} \sum_{i=1}^{N} \sum_{t=T_1}^{T_2} \widehat{P}^i_t = €42.7/\mathrm{MW}. \tag{6.1}$$

However, this price does not account for the correlation between day-ahead prices and the quantities produced or consumed. Since we have simulations from the joint distribution, we can reflect this correlation in the forward prices by weighting each price by its corresponding produced quantity, i.e. set the forward prices to the weighted means

$$F_S = \frac{\sum_{i=1}^{N} \sum_{t=T_1}^{T_2} \widehat{S}^i_t \widehat{P}^i_t}{\sum_{i=1}^{N} \sum_{t=T_1}^{T_2} \widehat{S}^i_t} = €42.3/\mathrm{MW} \tag{6.2}$$

$$F_W = \frac{\sum_{i=1}^{N} \sum_{t=T_1}^{T_2} \widehat{W}^i_t \widehat{P}^i_t}{\sum_{i=1}^{N} \sum_{t=T_1}^{T_2} \widehat{W}^i_t} = €37.9/\mathrm{MW}, \tag{6.3}$$

which as expected are lower than the unweighted mean, due to the negative correlation between the day-ahead price and the solar and wind production. A similar argument can be made for the consumer C, yielding the forward price

$$F_C = \frac{\sum_{i=1}^{N} \sum_{t=T_1}^{T_2} \widehat{C}^i_t \widehat{P}^i_t}{\sum_{i=1}^{N} \sum_{t=T_1}^{T_2} \widehat{C}^i_t} = €43.1/\mathrm{MW}, \tag{6.4}$$

which is higher than the unweighted mean due to the positive correlation between the day-ahead price and the consumption level.

The payoff of a portfolio of three such PPAs can now be calculated as

$$\mathcal{P}_{\mathrm{PPAP}} = 24 \sum_{t=T_1}^{T_2} \Big(C_t F_C - S_t F_S - W_t F_W + P_t \left(S_t + W_t - C_t \right) \Big) \tag{6.5}$$

$$= 24 \sum_{t=T_1}^{T_2} \Big(S_t \left(P_t - F_S \right) + W_t \left(P_t - F_W \right) - C_t \left(P_t - F_C \right) \Big), \tag{6.6}$$

that is, the quantities sold or bought, at the price specified in the respective PPA, plus the surplus generated power sold at the spot market (if S and W produce more power than C consumes) or the deficit power bought at the spot market (if C consumes more than S and W produce). This can also be represented in terms of of price-forward spreads weighted by the actualised production or consumption, with the signs of the terms representing positions (positive = bought quantities, negative = sold quantities). Exchanging the actualised quantities, P_t, C_t, S_t, and W_t, by the simulated quantities, we get N simulated payoffs $\mathcal{P}^i_{\text{PPAP}}$, $i = 1, \ldots, N$, representing a *payoff distribution*. This distribution is visualised on Figure 6.1.

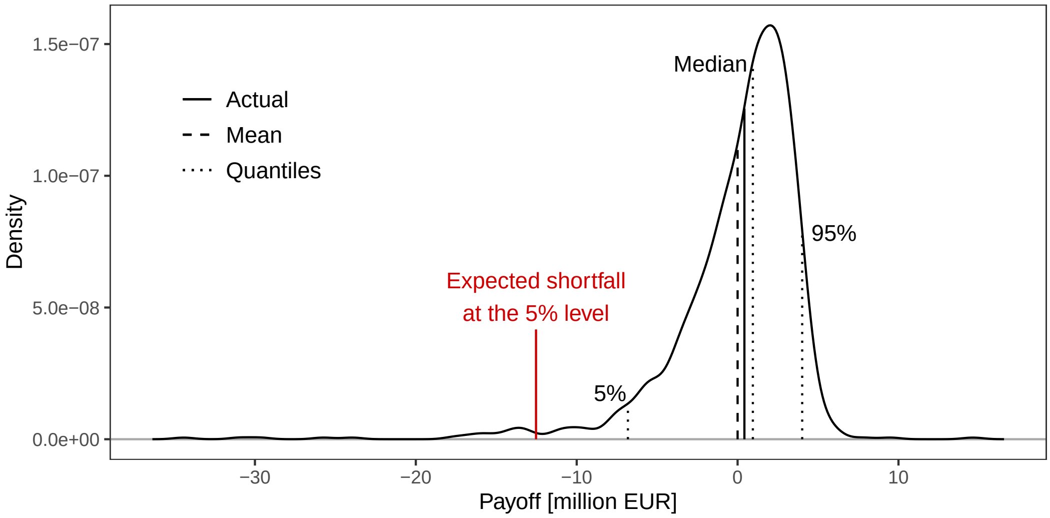


Figure 6.1: *Simulated distribution for $\mathcal{P}_{\text{PPAP}}$, with lines inserted to indicate, in order from left to right: the 5% expected shortfall, the 5% quantile/Value-at-Risk, the mean payoff, the actualised payoff, the median, and the 95% quantile*

There are several things to notice here. The mean payoff is zero by construction, but the median payoff is positive, about €0.95 million. However, while the distribution is skewed to the right, it's left tail is almost three times longer than the right tail. Indeed, inspecting the tail quantiles $q_{0.05}(\mathcal{P}_{\text{PPAP}})$ and $q_{0.95}(\mathcal{P}_{\text{PPAP}})$, where $q_p(X) = F_X^{(-1)}(p)$, we see that

$$q_{0.05}(\mathcal{P}_{\text{PPAP}}) = €-6,816,790.40 \quad \text{and} \quad q_{0.95}(\mathcal{P}_{\text{PPAP}}) = €4,015,347.90 \tag{6.7}$$

In other words, in the best 5% of cases, we only earn about 60% as much as we lose in the worst 5% of cases. The latter is also called the *5% Value-at-Risk* (or VaR for short). This skew is further accentuated when considering the *expected shortfall* of the portfolio, defined as

$$\text{ES}_p(\mathcal{P}_{\text{PPAP}}) = \mathbb{E}\Big(\mathcal{P}_{\text{PPAP}} \big| \mathcal{P}_{\text{PPAP}} \leq q_p(\mathcal{P}_{\text{PPAP}})\Big), \tag{6.8}$$

i.e. the expected loss conditional on being in the left tail of the distribution. At the 5% level, the expected shortfall for the portfolio is €12.5 million. If we define a similar quantity for the right tail, i.e. where the inequality is flipped and $p = 0.95$, we obtain an expected right-tail profit of only €4.97 million, only about 40% of the expected shortfall.

6.3 Towards An Optimal Portfolio

The nature of the payoff distribution is intuitively clear: variations in consumption notwithstanding, high production of renewables gives us a surplus of power that we can sell on the spot market, but the spot price is pushed down, while low production instead pushes the spot price up, but the lower quantity also means we have to buy the deficit on the spot market to serve our consumer counterparts. Such a distribution, where the majority of the mass may be on the positive halfline, but the left tail is slowly decaying, is not attractive to a trader.

For illustrative purposes, imagine that the sizes of our counterparts are different, say, such that the consumer C comprises 10% of German consumption instead of 1%. The payoff distribution for such a contract is plotted on Figure 6.2.

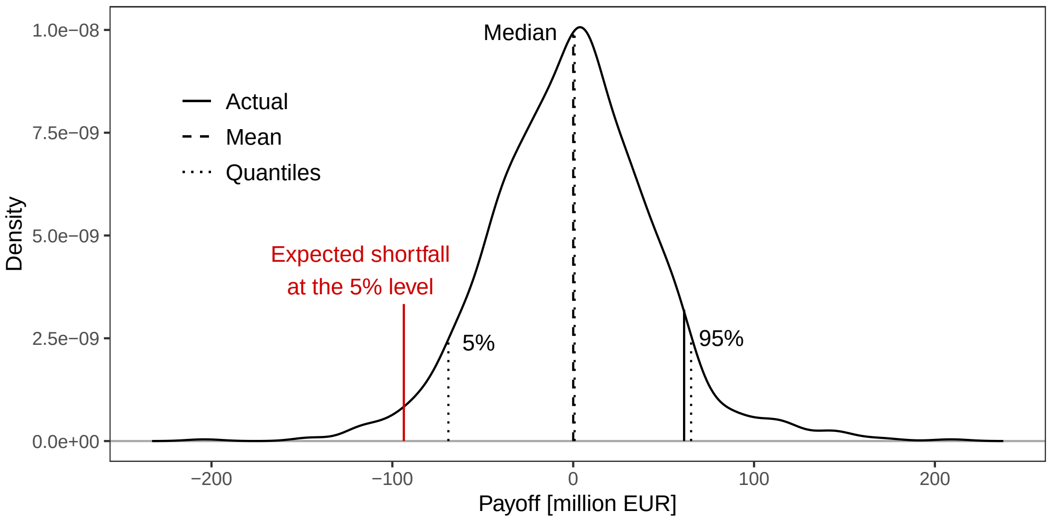


Figure 6.2: *Simulated distribution for $\mathcal{P}_{\text{PPAP}}$ with C comprising 10% of the consumption instead of 1%*

This distribution is more symmetric, but in return the expected shortfall is far higher than before—at €93.7 million, it's almost an order of magnitude larger. The reason for both is that as C grows in size, the consumption dominates the production in our portfolio, and thus the effect on the payoff from the PPAs with the wind and solar producers vanish, since we have to buy the majority of the consumed quantity on the spot market, making us more exposed to variations in day-ahead prices.

If on the other hand, if we imagine that W is a larger provider, say, comprising 50% of the total German wind power production instead of 5%, we obtain a payoff distribution as plotted on Figure 6.3.

This distribution is similar to the first in shape, but the expected shortfall is even higher than in the high-C scenario; at €181 million it's almost double the value. The higher position in wind production means we are less likely to have a deficit against C, but we still become

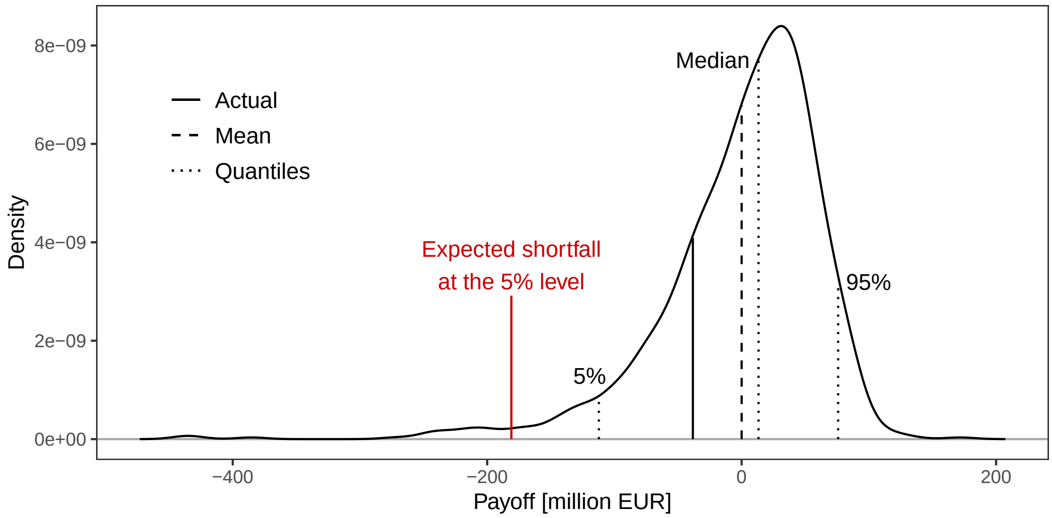


Figure 6.3: *Simulated distribution for $\mathcal{P}_{\text{PPAP}}$ with W comprising 50% of the wind power production instead of 5%*

exposed to the day-ahead price, as we now sell surplus power on the spot market. High wind productions imply low spot prices, which nets us a loss on the surplus.

Clearly, the shape of the payoff distribution depend greatly on the size of each counterparts relative to each other, and the question is whether we can "choose", in some sense, a portfolio with the most desirable possible payoff distribution. To this end, assume that each counterpart C, S, and W is a basket of smaller counterparts, and we can control the size of the basket by entering into contracts with more or fewer of these smaller counterparts. For the purpose of this example, assume furthermore that we can choose any size of counterpart this way, up to 100% of the total German consumption/production, and denote it as

$$\mathcal{P}_{\text{PPAP}}\big(\alpha_C,\alpha_S,\alpha_W\big) = 24\sum_{t=T_1}^{T_2}\Big(\alpha_S S_t^*\left(P_t - F_S\right) + \alpha_W W_t^*\left(P_t - F_W\right) - \alpha_C C_t^*\left(P_t - F_C\right)\Big), \tag{6.9}$$

where $\alpha_C, \alpha_S, \alpha_W \in (0,1)$, and S_t^*, W_t^*, and C_t^* denote the total German solar power production, wind power production, and consumption, respectively, at time t. Our goal is now to find values of α_C, α_S, and α_W, such that the resulting distribution is "optimal". For the measure of optimality, we consider the expected shortfall at the 5% level under the given distribution, and the problem then becomes a maximisation problem:[3]

$$\max_{\alpha_C,\alpha_S,\alpha_W} \text{ES}_{0.05}\Big(\mathcal{P}_{\text{PPAP}}\big(\alpha_C,\alpha_S,\alpha_W\big)\Big) \tag{6.10}$$

[3] Note that we have previously listed expected shortfalls without signs, since the interpretation is the amount we lose, but for these distributions the values of those expectations are negative, so maximising expected shortfall corresponds to minimising the interpreted loss.

It's easy to see that allowing the variables to vary freely over $(0,1)$ would push them all towards 0 because of the quantity we're optimising over. We conjecture that the distribution shape depends only on the relative size of the counterparts, as long as the size of one does not dominate the others. Therefore, fix the size of C at 1% of the total German consumption, i.e. set $\alpha_C = 0.01$, and let α_S and α_W be free variables in the maximisation problem. The values obtained at the maximum are

$$\alpha_S = 0.015 \quad \text{and} \quad \alpha_W = 0.025, \tag{6.11}$$

that is, when the consumer position comprise 1% of the total consumption, the provider positions yielding the smallest expected shortfall are 1.5% of the total solar power production and 2.5% of the total wind power production. The resulting distribution is shown on Figure 6.4

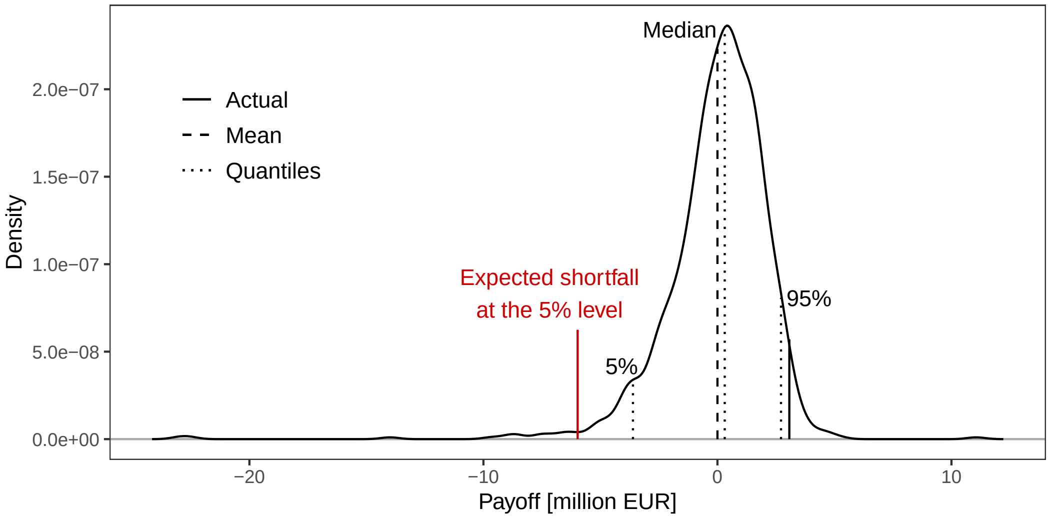


Figure 6.4: *Simulated payoff distribution for $\mathcal{P}_{\text{PPAP}}$ with optimised portfolio allocation*

This distribution, like the first arbitrarily chosen one, is skewed to the right with a long left tail, but the key difference is in the mass of said tail, which has fewer bumps and is "only" twice as long as the right tail. The expected shortfall is €5.97 million, about half of the first portfolio, but the expected right-tail profit is also smaller, at €3.42 million it's about 57% of the expected shortfall.

6.4 Diagnostics and Possible Extensions

The chosen "optimal" payoff distribution above still exhibits some undesirable traits, namely the long left tail, and an entry point for further research might lie in extensions of the underlying model. In this section we take a look at the specific simulations associated with the extreme losses.

As can gleaned from Figure 6.4, the mass in the tail is caused by only a few simulated scenarios—in numbers, out of the tail 5% of values, about 20% of the mass is ascribed to just 6% of the scenarios. These scenarios correspond to the losses that are larger than the largest profit in the right tail, which for this given simulation comprise three values.

The path simulations for the underlying data yielding these losses are plotted on Figure 6.5. The cause of each loss is readily apparent:

- The loss of €14 million (for data plotted with solid lines) is caused by simultaneous spikes in spot prices and wind production happening in June, where the wind production hits 100,000 MW while the spot price dips to €-119/MW, followed by very large fluctuations in the spot price. The high wind production dominates the consumption at that time, so we "sell" a large quantity of power at a negative price, effectively meaning we pay to get rid of it.

- The loss of €22.5 million (for data plotted with dotted lines) is caused by an absurd spike in solar power production towards the end of June, hitting over 10 million MW. The simulated spot price on this date is €36.2/MW, which is lower than the forward price for solar power production, so we face a loss on the surplus production, and since the simulated consumption is only around 50,000 MW, the vast majority of this production is a surplus.

- The loss of €23 million (for data plotted with dashed lines) is likewise caused by abnormally large values for solar power production—in the millions of MW—and this time for a wider range of observations simulated in the first half of July. The simulated spot price is again low, and the consumption is easily dominated by the production, yielding a huge loss.

The huge values for solar and wind productions causing the losses are most likely artifacts of assumptions implicit in the model, that—in light of these results—are too liberal. The production variables are effectively modelled to have the entire positive halfline as their codomains, but this does not reflect the reality that theoretically possible power production is bounded from above.

According to [ENTSO-E, 2019], the total installed capacity for renewables in Germany anno 2019 is around 45,000 MW and 59,000 MW for solar and wind, respectively. This is a far cry from the millions of MWs simulated in these extreme cases, and should be cause for revisiting the base assumptions of the model. Possible ways to incorporate the physical constraints could be to model the production variables with a truncated distribution, i.e. a probability distribution with hard limits on its range, or more simply to consider the produced quantities as ratios of the installed capacity.

Due to time constraints, we do not carry out such corrections in this project, but we conjecture that implementing these physical limits would yield more symmetric payoff distributions on PPA portfolios and let it be a topic for further research in later projects.

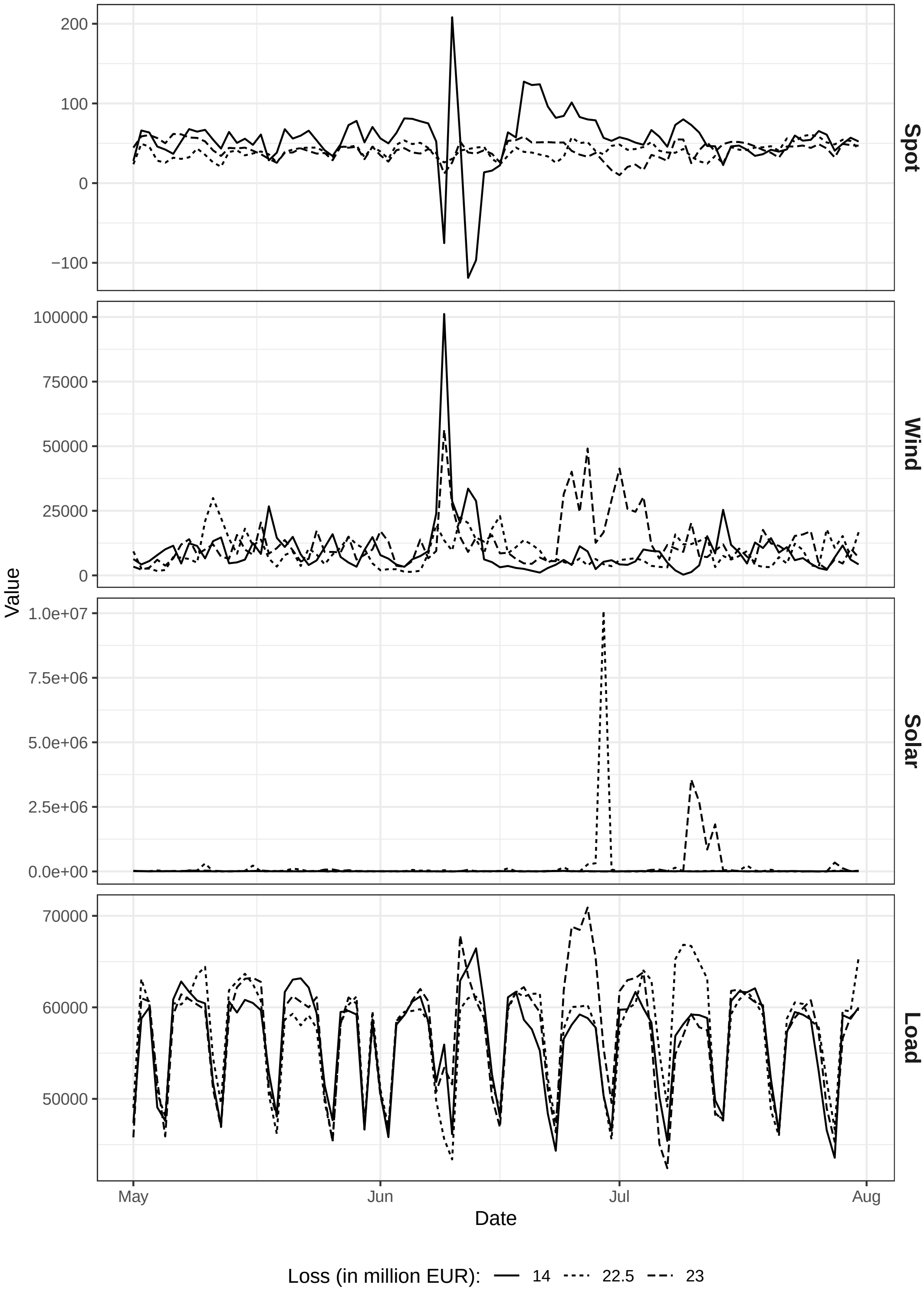


Figure 6.5: *Simulated paths corresponding to the three biggest losses on the optimal portfolio*

7 Conclusions and Outlooks

In this project, we present the theory of copulas, starting from basic definitions from probability theory, and including all results and definitions necessary to prove the main result, Sklar's Theorem (2.20), which lets us separate marginal and joint modelling of variables. With this result, we present the ideas behind vine copulas and how one can construct a d-dimensional copula distribution from a set of 2-dimensional ones. These pair copulas need not be the same or even belong to the same family of copulas, and thus the vine acts as a flexible extension to copula modelling.

Armed with this knowledge, we set out to describe the joint behaviour of daily base prices together with consumption, solar power production, and wind power production, choosing Germany as the area of interest. Based on freely available historical market data, we considered the four variables as time series on a daily frequency by averaging hourly or quarterly observations grouped by days.

We first filtered out seasonal variations in the data due to the time of year via a simple linear model with the log-transformed variables as responses. These models were constructed such that they only depended on time indices, making them completely deterministic, and the residuals were then further modelled as an ARMA-GARCH process. Each model exhibited clear autoregressive dependence, and the leverage-adjusting variants of the GARCH models were preferred to the symmetric standard GARCH. Student's t-distributed conditional errors were chosen for spot prices, consumption, and solar power production, while Gaussian errors were selected for wind power production, indicating lighter tails than the other variables.

Using the residuals transformed by their estimated distributions, we then obtain uniformly distributed variables that we model together with a 4-dimensional vine copula. The estimated structure is a C-vine with the wind power production as the root node, which indicates that among all possible pairs, the strongest rank correlations involved wind for all variables. This is accentuated by the estimated copulas in the first tree of the C-vine, as all three are types with tail dependence or rotated versions of such types. Strong negative dependence was found between wind and spot prices, and weak negative dependence was present between wind and solar. Curiously, wind and consumption exhibited a surprising amount of positive dependence, complete with upper tail dependence. We conjectured that this effect could be due to big market players who are both exposed to the spot price while being able to take on a consumer role. Condtional on wind production, the dependence between the other variables were Gaussian and not very strong, further indicating that wind is the most important contributing factor to the prices.

From this joint model for the four variables, we draw samples for simulated joint paths for the time window 2019-01-01 – 2019-07-31 and use them to simulate the distribution of the payoff for a portfolio of power purchase agreements with delivery during 2019-05-01 – 2019-07-31. Here, the effect of the wind power production became readily apparent, as the payoff distribution exhibited long left tails for portfolio allocations in which the size of consumers and producers did not match. For allocations with a large wind power provider, we became exposed to the price due to the negative dependence between wind and price, and for allocations with a large consumer, we became exposed due to the vanishing effect from the providers. We assume a fixed size consumer and find optimal sizes of providers based on the expected shortfall of the resulting portfolio.

7.1 Future Considerations

In our model, we made several simplifications and empirical findings that are prime candidates for further research in the future.

When applying the model to create portfolios of power purchase agreements, we observed some very extreme left tail outcomes that were contrary to the physical limits of the market. A natural modification of the model would therefore be to consider how to adequately take these physical limits into account, and see whether the resulting portfolio payoff distributions are more desirable. In the same vein, some of the marginal models could be reconsidered altogether. We used ARMA-GARCH models wholesale on the variables, when other, more specialised models may have been appropriate. As already mentioned earlier, consumption could possibly be modelled as simply

$$C_t = s_t^C + \epsilon_t, \tag{7.1}$$

i.e. a deterministic seasonal component s_t and a white noise error term ϵ_t, where s_t would include a lot more terms than the one used in this project. Solar power production seemed to exhibit patterns in the volatility, and a way to express this could be with it's stochastic part being a diffusion process with deterministic variance, that is, $S_t = s_t^S + X_t$ with e.g.

$$\mathrm{d}X_t^S = \sigma_t^S \,\mathrm{d}B_t, \tag{7.2}$$

$$\sigma_t^S = \beta_1 \sin\left(\frac{2\pi t}{365}\right) + \beta_2 \cos\left(\frac{2\pi t}{365}\right), \tag{7.3}$$

where B_t is a standard Brownian motion. The above would correspond to an annual cycle in volatility, which for suitable choices of β_1 and β_2 would peak during winters.

Seeing how exposed we were to the day-ahead prices through the wind power production, it would be a natural question to ask whether the risk associated with wind can be hedged. Indeed, a opportunity for this exists specifically on the German market: since 2017, so-called *wind power futures*, a derivative instrument written on a wind production index by EuroWind (an accredited German wind analysis company), have been available to trade on the European Energy Exchange. Investigating the effect of incorporating such contracts into our portfolio would be an interesting extension to consider.

Another natural extension of the model is found in the dimensions. We aggregated the data to daily observations, both because it's common to consider daily prices in the literature, but also because it dampens shocks a bit. If one instead models the hourly day-ahead electricity prices directly, one faces some interesting considerations. Due to the nature of how the prices are determined, all 24 hourly prices for a given day are based on the same information set, making it inappropriate to consider them as a univariate, hourly time series, and instead they should be considered as a 24-dimensional, daily time series. Here, one could again employ the flexible copula framework, and model the prices of each hour of the day individually as a univariate, daily time series, and link them together afterwards with a suitable copula model. Such a model would account for intraday variation and would allow for more flexible contracts to be written on the prices.

Appendices

A Overview of ARMA-GARCH Model Fits

	Model	ω	α	ϕ	θ	β	γ	ν	LLH	AIC
1	ARMA(0,0)-GARCH(1,0)-N	3.19E-05	1.000	-	-	-	-	-	3720.50	-5.108
2	ARMA(1,0)-GARCH(1,0)-N	1.22E-04	0.393	0.752	-	-	-	-	4264.32	-5.853
3	ARMA(0,1)-GARCH(1,0)-N	1.54E-04	0.383	-	0.527	-	-	-	4087.55	-5.611
4	ARMA(1,1)-GARCH(1,0)-N	1.20E-04	0.416	0.762	-0.027	-	-	-	4264.39	-5.852
5	ARMA(0,0)-GARCH(0,1)-N	1.21E-06	-	-	-	0.997	-	-	3668.48	-5.036
6	ARMA(1,0)-GARCH(0,1)-N	6.46E-07	-	0.695	-	0.997	-	-	4147.61	-5.693
7	ARMA(0,1)-GARCH(0,1)-N	6.51E-07	-	-	0.580	0.998	-	-	4002.63	-5.494
8	ARMA(1,1)-GARCH(0,1)-N	1.25E-06	-	0.703	-0.016	0.994	-	-	4145.28	-5.689
9	ARMA(0,0)-GARCH(1,1)-N	1.27E-06	0.100	-	-	0.900	-	-	3822.43	-5.246
10	ARMA(1,0)-GARCH(1,1)-N	7.43E-07	0.087	0.749	-	0.913	-	-	4304.15	-5.907
11	ARMA(0,1)-GARCH(1,1)-N	7.67E-07	0.086	-	0.590	0.914	-	-	4124.38	-5.660
12	ARMA(1,1)-GARCH(1,1)-N	2.98E-06	0.116	0.799	-0.109	0.884	-	-	4318.25	-5.925
13	ARMA(0,0)-GJR-GARCH(1,0)-N	1.19E-04	0.722	-	-	-	0.134	-	3914.13	-5.372
14	ARMA(1,0)-GJR-GARCH(1,0)-N	4.89E-06	0.500	-0.009	-	-	1.000	-	2415.83	-3.313
15	ARMA(0,1)-GJR-GARCH(1,0)-N	3.33E-06	0.500	-	-0.009	-	1.000	-	2016.47	-2.764
16	ARMA(1,1)-GJR-GARCH(1,0)-N	1.23E-04	0.305	0.767	-0.050	-	0.127	-	4267.69	-5.855
17	ARMA(0,0)-GJR-GARCH(0,1)-N	1.21E-06	-	-	-	0.997	-	-	3668.48	-5.036
18	ARMA(1,0)-GJR-GARCH(0,1)-N	6.46E-07	-	0.695	-	0.997	-	-	4147.61	-5.693
19	ARMA(0,1)-GJR-GARCH(0,1)-N	6.51E-07	-	-	0.580	0.998	-	-	4002.63	-5.494
20	ARMA(1,1)-GJR-GARCH(0,1)-N	1.25E-06	-	0.703	-0.016	0.994	-	-	4145.28	-5.689
21	ARMA(0,0)-GJR-GARCH(1,1)-N	9.52E-07	0.144	-	-	0.834	0.045	-	3824.91	-5.248
22	ARMA(1,0)-GJR-GARCH(1,1)-N	9.13E-07	0.072	0.763	-	0.903	0.050	-	4306.23	-5.908
23	ARMA(0,1)-GJR-GARCH(1,1)-N	3.77E-06	0.115	-	0.589	0.878	-0.000	-	4137.46	-5.676
24	ARMA(1,1)-GJR-GARCH(1,1)-N	2.35E-06	0.075	0.815	-0.123	0.894	0.051	-	4319.25	-5.925
25	ARMA(0,0)-E-GARCH(1,0)-N	-8.15E+00	-0.006	-	-	-	0.905	-	3869.42	-5.311
26	ARMA(1,0)-E-GARCH(1,0)-N	-8.62E+00	-0.074	0.715	-	-	0.432	-	4239.87	-5.819
27	ARMA(0,1)-E-GARCH(1,0)-N	-8.40E+00	-0.001	-	0.518	-	0.449	-	4066.71	-5.581
28	ARMA(1,1)-E-GARCH(1,0)-N	-8.63E+00	-0.082	0.752	-0.101	-	0.442	-	4241.87	-5.820
29	ARMA(0,0)-E-GARCH(0,1)-N	1.61E-04	-	-	-	1.000	-	-	3674.53	-5.045
30	ARMA(1,0)-E-GARCH(0,1)-N	1.94E-04	-	0.697	-	1.000	-	-	4155.06	-5.703
31	ARMA(0,1)-E-GARCH(0,1)-N	1.79E-04	-	-	0.583	1.000	-	-	4008.53	-5.502
32	ARMA(1,1)-E-GARCH(0,1)-N	-1.56E-04	-	0.705	-0.018	1.000	-	-	4155.02	-5.702
33	ARMA(0,0)-E-GARCH(1,1)-N	-1.57E-01	-0.013	-	-	0.980	0.443	-	3848.71	-5.281
34	ARMA(1,0)-E-GARCH(1,1)-N	-6.13E-01	-0.054	0.775	-	0.928	0.320	-	4327.06	-5.937
35	ARMA(0,1)-E-GARCH(1,1)-N	-6.24E-01	-0.015	-	0.575	0.925	0.321	-	4139.77	-5.680
36	ARMA(1,1)-E-GARCH(1,1)-N	-7.19E-01	-0.066	0.819	-0.126	0.916	0.332	-	4330.61	-5.940

Table A.1: *Overview of the first 36 of the fitted ARMA-GARCH models for the spot price data*

	Model	ω	α	ϕ	θ	β	γ	ν	LLH	AIC
37	ARMA(0,0)-GARCH(1,0)-t	1.08E-04	0.867	-	-	-	-	12.081	3928.39	-5.392
38	ARMA(1,0)-GARCH(1,0)-t	1.21E-04	0.421	0.775	-	-	-	4.035	4364.21	-5.989
39	ARMA(0,1)-GARCH(1,0)-t	1.57E-04	0.346	-	0.567	-	-	5.585	4147.47	-5.692
40	ARMA(1,1)-GARCH(1,0)-t	1.22E-04	0.430	0.809	-0.098	-	-	3.994	4366.64	-5.991
41	ARMA(0,0)-GARCH(0,1)-t	8.62E-07	-	-	-	0.998	-	4.772	3747.01	-5.143
42	ARMA(1,0)-GARCH(0,1)-t	5.63E-07	-	0.764	-	0.999	-	2.493	4327.05	-5.938
43	ARMA(0,1)-GARCH(0,1)-t	2.77E-07	-	-	0.596	1.000	-	2.757	4103.04	-5.631
44	ARMA(1,1)-GARCH(0,1)-t	9.83E-07	-	0.773	-0.021	0.998	-	2.477	4323.23	-5.932
45	ARMA(0,0)-GARCH(1,1)-t	1.01E-04	0.750	-	-	0.065	-	11.860	3927.69	-5.390
46	ARMA(1,0)-GARCH(1,1)-t	2.25E-06	0.111	0.766	-	0.887	-	4.634	4382.37	-6.013
47	ARMA(0,1)-GARCH(1,1)-t	3.30E-06	0.139	-	0.592	0.859	-	6.827	4180.92	-5.736
48	ARMA(1,1)-GARCH(1,1)-t	1.65E-06	0.102	0.815	-0.114	0.898	-	4.713	4383.65	-6.013
49	ARMA(0,0)-GJR-GARCH(1,0)-t	9.35E-06	0.977	-	-	-	0.046	6.226	3559.62	-4.884
50	ARMA(1,0)-GJR-GARCH(1,0)-t	1.26E-04	0.291	0.776	-	-	0.273	3.921	4366.40	-5.991
51	ARMA(0,1)-GJR-GARCH(1,0)-t	1.56E-04	0.450	-	0.568	-	-0.183	5.531	4148.96	-5.692
52	ARMA(1,1)-GJR-GARCH(1,0)-t	9.01E-06	0.973	0.901	-0.177	-	0.055	4.262	3889.72	-5.335
53	ARMA(0,0)-GJR-GARCH(0,1)-t	8.62E-07	-	-	-	0.998	-	4.772	3747.01	-5.143
54	ARMA(1,0)-GJR-GARCH(0,1)-t	5.63E-07	-	0.764	-	0.999	-	2.493	4327.05	-5.938
55	ARMA(0,1)-GJR-GARCH(0,1)-t	2.77E-07	-	-	0.596	1.000	-	2.757	4103.04	-5.631
56	ARMA(1,1)-GJR-GARCH(0,1)-t	9.83E-07	-	0.773	-0.021	0.998	-	2.477	4323.23	-5.932
57	ARMA(0,0)-GJR-GARCH(1,1)-t	1.01E-04	0.829	-	-	0.034	0.017	12.626	3928.36	-5.389
58	ARMA(1,0)-GJR-GARCH(1,1)-t	2.63E-06	0.095	0.778	-	0.876	0.050	4.720	4384.10	-6.014
59	ARMA(0,1)-GJR-GARCH(1,1)-t	3.34E-06	0.122	-	0.590	0.893	-0.043	6.733	4183.59	-5.738
60	ARMA(1,1)-GJR-GARCH(1,1)-t	1.28E-06	0.075	0.830	-0.126	0.900	0.050	4.810	4382.55	-6.010
61	ARMA(0,0)-E-GARCH(1,0)-t	-8.18E+00	0.046	-	-	-	0.953	10.880	3887.61	-5.335
62	ARMA(1,0)-E-GARCH(1,0)-t	-8.59E+00	-0.061	0.765	-	-	0.490	3.596	4351.24	-5.970
63	ARMA(0,1)-E-GARCH(1,0)-t	-8.44E+00	0.064	-	0.561	-	0.439	5.285	4136.23	-5.675
64	ARMA(1,1)-E-GARCH(1,0)-t	-8.59E+00	-0.057	0.789	-0.078	-	0.502	3.617	4352.47	-5.970
65	ARMA(0,0)-E-GARCH(0,1)-t	2.95E-04	-	-	-	1.000	-	3.842	3751.46	-5.149
66	ARMA(1,0)-E-GARCH(0,1)-t	-6.23E-01	-	0.756	-	0.926	-	3.090	4316.22	-5.923
67	ARMA(0,1)-E-GARCH(0,1)-t	-6.24E-01	-	-	0.587	0.925	-	4.588	4104.56	-5.633
68	ARMA(1,1)-E-GARCH(0,1)-t	-1.67E-02	-	0.784	-0.027	0.998	-	2.245	4316.65	-5.923
69	ARMA(0,0)-E-GARCH(1,1)-t	-6.04E-01	0.016	-	-	0.927	0.605	9.742	3899.73	-5.350
70	ARMA(1,0)-E-GARCH(1,1)-t	-8.71E-01	-0.082	0.791	-	0.901	0.382	4.666	4398.86	-6.034
71	ARMA(0,1)-E-GARCH(1,1)-t	-7.37E-01	0.022	-	0.582	0.914	0.374	6.313	4191.80	-5.750
72	ARMA(1,1)-E-GARCH(1,1)-t	-7.86E-01	-0.091	0.840	-0.137	0.911	0.371	4.729	4402.29	-6.037

Table A.2: *Overview of the last 36 of the fitted ARMA-GARCH models for the spot price data*

	Model	ω	α	ϕ	θ	β	γ	ν	LLH	AIC
1	ARMA(0,0)-GARCH(1,0)-N	1.25E-04	1.000	-	-	-	-	-	2559.24	-3.513
2	ARMA(1,0)-GARCH(1,0)-N	2.32E-05	1.000	0.051	-	-	-	-	1330.05	-1.823
3	ARMA(0,1)-GARCH(1,0)-N	5.60E-04	0.571	-	0.488	-	-	-	3081.71	-4.229
4	ARMA(1,1)-GARCH(1,0)-N	4.03E-04	0.736	0.641	0.029	-	-	-	3236.29	-4.440
5	ARMA(0,0)-GARCH(0,1)-N	4.75E-06	-	-	-	0.997	-	-	2670.77	-3.666
6	ARMA(1,0)-GARCH(0,1)-N	2.57E-06	-	0.653	-	0.997	-	-	3074.81	-4.220
7	ARMA(0,1)-GARCH(0,1)-N	2.51E-06	-	-	0.489	0.998	-	-	2925.76	-4.015
8	ARMA(1,1)-GARCH(0,1)-N	6.10E-08	-	0.764	-0.196	1.000	-	-	3087.82	-4.236
9	ARMA(0,0)-GARCH(1,1)-N	4.72E-06	0.100	-	-	0.900	-	-	2770.96	-3.802
10	ARMA(1,0)-GARCH(1,1)-N	1.54E-05	0.141	0.708	-	0.859	-	-	3170.59	-4.350
11	ARMA(0,1)-GARCH(1,1)-N	1.68E-05	0.162	-	0.541	0.838	-	-	3044.40	-4.176
12	ARMA(1,1)-GARCH(1,1)-N	2.20E-05	0.100	0.783	-0.147	0.890	-	-	3171.32	-4.349
13	ARMA(0,0)-GJR-GARCH(1,0)-N	5.55E-04	0.553	-	-	-	0.248	-	2935.87	-4.029
14	ARMA(1,0)-GJR-GARCH(1,0)-N	3.88E-04	0.374	0.664	-	-	0.965	-	3257.46	-4.469
15	ARMA(0,1)-GJR-GARCH(1,0)-N	1.21E-05	0.990	-	1.000	-	0.020	-	225.97	-0.305
16	ARMA(1,1)-GJR-GARCH(1,0)-N	3.87E-04	0.381	0.656	0.018	-	0.928	-	3257.60	-4.468
17	ARMA(0,0)-GJR-GARCH(0,1)-N	4.75E-06	-	-	-	0.997	-	-	2670.77	-3.666
18	ARMA(1,0)-GJR-GARCH(0,1)-N	2.57E-06	-	0.653	-	0.997	-	-	3074.81	-4.220
19	ARMA(0,1)-GJR-GARCH(0,1)-N	2.51E-06	-	-	0.489	0.998	-	-	2925.76	-4.015
20	ARMA(1,1)-GJR-GARCH(0,1)-N	6.10E-08	-	0.764	-0.196	1.000	-	-	3087.82	-4.236
21	ARMA(0,0)-GJR-GARCH(1,1)-N	3.67E-06	0.075	-	-	0.900	0.050	-	2756.66	-3.781
22	ARMA(1,0)-GJR-GARCH(1,1)-N	3.32E-04	0.371	0.684	-	0.082	0.840	-	3259.16	-4.470
23	ARMA(0,1)-GJR-GARCH(1,1)-N	9.87E-05	0.105	-	0.545	0.795	0.027	-	3055.92	-4.191
24	ARMA(1,1)-GJR-GARCH(1,1)-N	4.86E-06	0.075	0.795	-0.144	0.900	0.050	-	3142.13	-4.308
25	ARMA(0,0)-E-GARCH(1,0)-N	-6.76E+00	-0.115	-	-	-	0.785	-	2878.15	-3.949
26	ARMA(1,0)-E-GARCH(1,0)-N	-7.19E+00	-0.111	0.663	-	-	0.601	-	3207.23	-4.400
27	ARMA(0,1)-E-GARCH(1,0)-N	-6.97E+00	-0.099	-	0.451	-	0.571	-	3036.87	-4.166
28	ARMA(1,1)-E-GARCH(1,0)-N	-7.19E+00	-0.121	0.681	-0.050	-	0.597	-	3207.84	-4.400
29	ARMA(0,0)-E-GARCH(0,1)-N	-2.81E-04	-	-	-	1.000	-	-	2671.00	-3.666
30	ARMA(1,0)-E-GARCH(0,1)-N	7.70E-05	-	0.659	-	1.000	-	-	3076.20	-4.221
31	ARMA(0,1)-E-GARCH(0,1)-N	6.62E-05	-	-	0.500	1.000	-	-	2926.54	-4.016
32	ARMA(1,1)-E-GARCH(0,1)-N	4.37E-06	-	0.764	-0.196	1.000	-	-	3087.89	-4.236
33	ARMA(0,0)-E-GARCH(1,1)-N	-1.69E+00	-0.020	-	-	0.748	0.671	-	2897.57	-3.975
34	ARMA(1,0)-E-GARCH(1,1)-N	-3.08E+00	-0.146	0.713	-	0.569	0.653	-	3247.22	-4.454
35	ARMA(0,1)-E-GARCH(1,1)-N	-6.71E-01	-0.036	-	0.573	0.901	0.421	-	3089.28	-4.237
36	ARMA(1,1)-E-GARCH(1,1)-N	-1.79E+00	-0.106	0.763	-0.088	0.746	0.535	-	3237.54	-4.439

Table A.3: *Overview of the first 36 of the fitted ARMA-GARCH models for the load data*

	Model	ω	α	ϕ	θ	β	γ	ν	LLH	AIC
37	ARMA(0,0)-GARCH(1,0)-t	4.22E-04	0.870	-	-	-	-	6.661	3008.47	-4.128
38	ARMA(1,0)-GARCH(1,0)-t	4.10E-04	0.560	0.727	-	-	-	3.767	3428.04	-4.703
39	ARMA(0,1)-GARCH(1,0)-t	5.38E-04	0.611	-	0.485	-	-	4.301	3199.85	-4.390
40	ARMA(1,1)-GARCH(1,0)-t	4.14E-04	0.549	0.759	-0.077	-	-	3.681	3429.82	-4.704
41	ARMA(0,0)-GARCH(0,1)-t	3.22E-06	-	-	-	0.998	-	3.515	2828.05	-3.881
42	ARMA(1,0)-GARCH(0,1)-t	5.73E-06	-	0.732	-	0.996	-	2.513	3344.77	-4.589
43	ARMA(0,1)-GARCH(0,1)-t	2.49E-06	-	-	0.484	0.998	-	3.037	3130.09	-4.294
44	ARMA(1,1)-GARCH(0,1)-t	4.18E-06	-	0.772	-0.142	0.997	-	2.601	3350.99	-4.596
45	ARMA(0,0)-GARCH(1,1)-t	1.50E-04	0.100	-	-	0.757	-	5.324	2914.02	-3.997
46	ARMA(1,0)-GARCH(1,1)-t	2.91E-04	0.222	0.750	-	0.304	-	3.952	3420.88	-4.692
47	ARMA(0,1)-GARCH(1,1)-t	1.36E-05	0.080	-	0.518	0.909	-	4.667	3153.02	-4.324
48	ARMA(1,1)-GARCH(1,1)-t	1.50E-05	0.078	0.792	-0.154	0.901	-	3.741	3367.17	-4.617
49	ARMA(0,0)-GJR-GARCH(1,0)-t	4.08E-04	0.892	-	-	-	0.050	6.145	3008.18	-4.127
50	ARMA(1,0)-GJR-GARCH(1,0)-t	4.46E-04	0.435	0.730	-	-	0.344	3.492	3430.54	-4.705
51	ARMA(0,1)-GJR-GARCH(1,0)-t	4.86E-04	0.586	-	0.483	-	0.087	4.646	3200.66	-4.390
52	ARMA(1,1)-GJR-GARCH(1,0)-t	3.27E-04	0.561	0.720	0.008	-	0.255	4.085	3425.82	-4.698
53	ARMA(0,0)-GJR-GARCH(0,1)-t	3.22E-06	-	-	-	0.998	-	3.515	2828.05	-3.881
54	ARMA(1,0)-GJR-GARCH(0,1)-t	5.73E-06	-	0.732	-	0.996	-	2.513	3344.77	-4.589
55	ARMA(0,1)-GJR-GARCH(0,1)-t	2.49E-06	-	-	0.484	0.998	-	3.037	3130.09	-4.294
56	ARMA(1,1)-GJR-GARCH(0,1)-t	4.18E-06	-	0.772	-0.142	0.997	-	2.601	3350.99	-4.596
57	ARMA(0,0)-GJR-GARCH(1,1)-t	4.03E-04	0.827	-	-	0.020	0.069	6.733	3008.93	-4.126
58	ARMA(1,0)-GJR-GARCH(1,1)-t	1.39E-04	0.287	0.757	-	0.613	0.034	3.231	3410.48	-4.676
59	ARMA(0,1)-GJR-GARCH(1,1)-t	4.50E-05	0.241	-	0.531	0.764	-0.065	5.035	3188.45	-4.371
60	ARMA(1,1)-GJR-GARCH(1,1)-t	3.02E-04	0.348	0.785	-0.101	0.175	0.265	3.936	3438.06	-4.713
61	ARMA(0,0)-E-GARCH(1,0)-t	-6.87E+00	0.021	-	-	-	0.940	6.080	2959.51	-4.060
62	ARMA(1,0)-E-GARCH(1,0)-t	-7.26E+00	-0.013	0.728	-	-	0.670	3.362	3413.48	-4.682
63	ARMA(0,1)-E-GARCH(1,0)-t	-7.07E+00	0.026	-	0.480	-	0.609	4.209	3177.69	-4.358
64	ARMA(1,1)-E-GARCH(1,0)-t	-7.27E+00	-0.009	0.751	-0.060	-	0.652	3.378	3414.41	-4.682
65	ARMA(0,0)-E-GARCH(0,1)-t	7.41E-05	-	-	-	1.000	-	3.449	2828.57	-3.881
66	ARMA(1,0)-E-GARCH(0,1)-t	-6.27E-01	-	0.729	-	0.909	-	2.742	3342.81	-4.586
67	ARMA(0,1)-E-GARCH(0,1)-t	-6.25E-01	-	-	0.483	0.909	-	3.429	3128.99	-4.293
68	ARMA(1,1)-E-GARCH(0,1)-t	3.75E-04	-	0.773	-0.144	1.000	-	2.628	3354.10	-4.600
69	ARMA(0,0)-E-GARCH(1,1)-t	-1.76E+00	0.036	-	-	0.745	0.802	6.242	2977.31	-4.083
70	ARMA(1,0)-E-GARCH(1,1)-t	-7.76E-01	-0.028	0.755	-	0.894	0.403	3.473	3405.16	-4.669
71	ARMA(0,1)-E-GARCH(1,1)-t	-7.23E-01	0.014	-	0.526	0.899	0.435	4.558	3202.00	-4.390
72	ARMA(1,1)-E-GARCH(1,1)-t	-8.96E-01	-0.031	0.805	-0.130	0.877	0.401	3.475	3413.18	-4.679

Table A.4: *Overview of the last 36 of the fitted ARMA-GARCH models for the load data*

	Model	ω	α	ϕ	θ	β	γ	ν	LLH	AIC
1	ARMA(0,0)-GARCH(1,0)-N	9.28E-02	0.388	-	-	-	-	-	-615.55	0.848
2	ARMA(1,0)-GARCH(1,0)-N	9.03E-02	0.239	0.493	-	-	-	-	-477.42	0.660
3	ARMA(0,1)-GARCH(1,0)-N	9.40E-02	0.238	-	0.408	-	-	-	-507.87	0.702
4	ARMA(1,1)-GARCH(1,0)-N	9.01E-02	0.242	0.460	0.047	-	-	-	-477.07	0.661
5	ARMA(0,0)-GARCH(0,1)-N	1.53E-04	-	-	-	0.999	-	-	-696.88	0.960
6	ARMA(1,0)-GARCH(0,1)-N	2.80E-03	-	0.481	-	0.976	-	-	-506.12	0.699
7	ARMA(0,1)-GARCH(0,1)-N	3.95E-04	-	-	0.424	0.997	-	-	-537.38	0.742
8	ARMA(1,1)-GARCH(0,1)-N	4.96E-04	-	0.466	0.019	0.996	-	-	-506.16	0.701
9	ARMA(0,0)-GARCH(1,1)-N	1.40E-02	0.242	-	-	0.675	-	-	-591.13	0.816
10	ARMA(1,0)-GARCH(1,1)-N	1.60E-03	0.067	0.486	-	0.920	-	-	-412.27	0.572
11	ARMA(0,1)-GARCH(1,1)-N	1.55E-03	0.067	-	0.428	0.921	-	-	-442.84	0.614
12	ARMA(1,1)-GARCH(1,1)-N	1.60E-03	0.067	0.467	0.024	0.920	-	-	-412.18	0.573
13	ARMA(0,0)-GJR-GARCH(1,0)-N	9.28E-02	0.412	-	-	-	-0.048	-	-615.37	0.849
14	ARMA(1,0)-GJR-GARCH(1,0)-N	8.99E-02	0.296	0.491	-	-	-0.099	-	-476.51	0.660
15	ARMA(0,1)-GJR-GARCH(1,0)-N	9.36E-02	0.308	-	0.409	-	-0.127	-	-506.26	0.701
16	ARMA(1,1)-GJR-GARCH(1,0)-N	8.96E-02	0.299	0.457	0.049	-	-0.101	-	-476.12	0.661
17	ARMA(0,0)-GJR-GARCH(0,1)-N	1.53E-04	-	-	-	0.999	-	-	-696.88	0.960
18	ARMA(1,0)-GJR-GARCH(0,1)-N	2.80E-03	-	0.481	-	0.976	-	-	-506.12	0.699
19	ARMA(0,1)-GJR-GARCH(0,1)-N	3.95E-04	-	-	0.424	0.997	-	-	-537.38	0.742
20	ARMA(1,1)-GJR-GARCH(0,1)-N	4.96E-04	-	0.466	0.019	0.996	-	-	-506.16	0.701
21	ARMA(0,0)-GJR-GARCH(1,1)-N	1.18E-02	0.258	-	-	0.707	-0.064	-	-589.80	0.816
22	ARMA(1,0)-GJR-GARCH(1,1)-N	1.70E-03	0.095	0.482	-	0.918	-0.047	-	-409.86	0.570
23	ARMA(0,1)-GJR-GARCH(1,1)-N	1.61E-03	0.092	-	0.428	0.920	-0.044	-	-439.76	0.611
24	ARMA(1,1)-GJR-GARCH(1,1)-N	1.69E-03	0.095	0.460	0.029	0.918	-0.047	-	-409.72	0.571
25	ARMA(0,0)-E-GARCH(1,0)-N	-1.97E+00	0.022	-	-	-	0.544	-	-632.96	0.874
26	ARMA(1,0)-E-GARCH(1,0)-N	-2.17E+00	0.033	0.501	-	-	0.351	-	-482.02	0.668
27	ARMA(0,1)-E-GARCH(1,0)-N	-2.13E+00	0.047	-	0.411	-	0.332	-	-514.54	0.712
28	ARMA(1,1)-E-GARCH(1,0)-N	-2.17E+00	0.035	0.471	0.044	-	0.355	-	-481.71	0.669
29	ARMA(0,0)-E-GARCH(0,1)-N	-4.85E-02	-	-	-	0.974	-	-	-696.83	0.960
30	ARMA(1,0)-E-GARCH(0,1)-N	-5.27E-02	-	0.481	-	0.976	-	-	-506.12	0.699
31	ARMA(0,1)-E-GARCH(0,1)-N	-4.97E-02	-	-	0.423	0.977	-	-	-537.31	0.742
32	ARMA(1,1)-E-GARCH(0,1)-N	-5.27E-02	-	0.466	0.019	0.976	-	-	-506.06	0.701
33	ARMA(0,0)-E-GARCH(1,1)-N	-3.45E-01	0.017	-	-	0.832	0.476	-	-592.72	0.820
34	ARMA(1,0)-E-GARCH(1,1)-N	-3.56E-02	0.007	0.485	-	0.984	0.133	-	-411.50	0.572
35	ARMA(0,1)-E-GARCH(1,1)-N	-3.31E-02	0.008	-	0.427	0.985	0.132	-	-442.46	0.615
36	ARMA(1,1)-E-GARCH(1,1)-N	-3.56E-02	0.008	0.457	0.036	0.984	0.134	-	-411.29	0.573

Table A.5: *Overview of the first 36 of the fitted ARMA-GARCH models for the solar data*

	Model	ω	α	ϕ	θ	β	γ	ν	LLH	AIC
37	ARMA(0,0)-GARCH(1,0)-t	9.20E-02	0.401	-	-	-	-	45.283	-615.58	0.850
38	ARMA(1,0)-GARCH(1,0)-t	8.97E-02	0.261	0.526	-	-	-	8.716	-466.34	0.646
39	ARMA(0,1)-GARCH(1,0)-t	9.32E-02	0.251	-	0.419	-	-	13.366	-502.80	0.696
40	ARMA(1,1)-GARCH(1,0)-t	8.96E-02	0.261	0.502	0.035	-	-	8.775	-466.12	0.647
41	ARMA(0,0)-GARCH(0,1)-t	3.73E-03	-	-	-	0.975	-	17.378	-693.56	0.957
42	ARMA(1,0)-GARCH(0,1)-t	1.20E-04	-	0.514	-	0.999	-	7.545	-490.43	0.679
43	ARMA(0,1)-GARCH(0,1)-t	3.08E-04	-	-	0.437	0.997	-	10.254	-528.49	0.731
44	ARMA(1,1)-GARCH(0,1)-t	1.20E-04	-	0.506	0.011	0.999	-	7.562	-490.40	0.680
45	ARMA(0,0)-GARCH(1,1)-t	1.34E-02	0.241	-	-	0.681	-	97.390	-591.13	0.817
46	ARMA(1,0)-GARCH(1,1)-t	1.32E-03	0.072	0.505	-	0.919	-	13.060	-404.93	0.563
47	ARMA(0,1)-GARCH(1,1)-t	1.31E-03	0.070	-	0.434	0.920	-	20.296	-439.39	0.610
48	ARMA(1,1)-GARCH(1,1)-t	1.31E-03	0.072	0.488	0.023	0.919	-	13.093	-404.84	0.564
49	ARMA(0,0)-GJR-GARCH(1,0)-t	9.22E-02	0.423	-	-	-	-0.056	79.049	-615.21	0.851
50	ARMA(1,0)-GJR-GARCH(1,0)-t	8.92E-02	0.314	0.522	-	-	-0.093	8.849	-465.79	0.647
51	ARMA(0,1)-GJR-GARCH(1,0)-t	9.27E-02	0.335	-	0.418	-	-0.149	13.100	-501.03	0.695
52	ARMA(1,1)-GJR-GARCH(1,0)-t	8.91E-02	0.315	0.497	0.036	-	-0.093	8.918	-465.57	0.648
53	ARMA(0,0)-GJR-GARCH(0,1)-t	3.73E-03	-	-	-	0.975	-	17.378	-693.56	0.957
54	ARMA(1,0)-GJR-GARCH(0,1)-t	1.20E-04	-	0.514	-	0.999	-	7.545	-490.43	0.679
55	ARMA(0,1)-GJR-GARCH(0,1)-t	3.08E-04	-	-	0.437	0.997	-	10.254	-528.49	0.731
56	ARMA(1,1)-GJR-GARCH(0,1)-t	1.20E-04	-	0.506	0.011	0.999	-	7.562	-490.40	0.680
57	ARMA(0,0)-GJR-GARCH(1,1)-t	1.13E-02	0.259	-	-	0.712	-0.068	89.149	-589.65	0.817
58	ARMA(1,0)-GJR-GARCH(1,1)-t	1.37E-03	0.101	0.497	-	0.917	-0.050	13.414	-402.84	0.562
59	ARMA(0,1)-GJR-GARCH(1,1)-t	1.33E-03	0.099	-	0.432	0.920	-0.050	19.114	-435.97	0.607
60	ARMA(1,1)-GJR-GARCH(1,1)-t	1.36E-03	0.101	0.476	0.029	0.918	-0.050	13.436	-402.71	0.563
61	ARMA(0,0)-E-GARCH(1,0)-t	-1.97E+00	0.025	-	-	-	0.550	100.000	-632.88	0.875
62	ARMA(1,0)-E-GARCH(1,0)-t	-2.17E+00	0.022	0.531	-	-	0.368	8.693	-470.98	0.654
63	ARMA(0,1)-E-GARCH(1,0)-t	-2.13E+00	0.050	-	0.424	-	0.337	13.079	-509.49	0.707
64	ARMA(1,1)-E-GARCH(1,0)-t	-2.17E+00	0.022	0.514	0.025	-	0.369	8.734	-470.85	0.655
65	ARMA(0,0)-E-GARCH(0,1)-t	-4.71E-02	-	-	-	0.975	-	17.381	-693.56	0.957
66	ARMA(1,0)-E-GARCH(0,1)-t	-4.02E-01	-	0.514	-	0.812	-	7.638	-490.44	0.679
67	ARMA(0,1)-E-GARCH(0,1)-t	-4.71E-02	-	-	0.437	0.978	-	10.501	-528.44	0.731
68	ARMA(1,1)-E-GARCH(0,1)-t	-3.47E-01	-	0.505	0.012	0.838	-	7.665	-490.42	0.681
69	ARMA(0,0)-E-GARCH(1,1)-t	-3.34E-01	0.019	-	-	0.838	0.475	99.379	-592.64	0.821
70	ARMA(1,0)-E-GARCH(1,1)-t	-3.13E-02	0.009	0.499	-	0.987	0.144	13.002	-404.08	0.563
71	ARMA(0,1)-E-GARCH(1,1)-t	-2.92E-02	0.011	-	0.431	0.987	0.139	19.133	-438.70	0.611
72	ARMA(1,1)-E-GARCH(1,1)-t	-3.12E-02	0.009	0.475	0.034	0.987	0.144	13.031	-403.90	0.564

Table A.6: *Overview of the last 36 of the fitted ARMA-GARCH models for the solar data*

	Model	ω	α	ϕ	θ	β	γ	ν	LLH	AIC
1	ARMA(0,0)-GARCH(1,0)-N	3.30E-01	0.350	-	-	-	-	-	-1517.38	2.087
2	ARMA(1,0)-GARCH(1,0)-N	3.33E-01	0.013	0.580	-	-	-	-	-1275.61	1.756
3	ARMA(0,1)-GARCH(1,0)-N	3.46E-01	0.033	-	0.529	-	-	-	-1316.53	1.813
4	ARMA(1,1)-GARCH(1,0)-N	3.31E-01	0.008	0.455	0.192	-	-	-	-1266.84	1.746
5	ARMA(0,0)-GARCH(0,1)-N	4.73E-04	-	-	-	0.999	-	-	-1575.21	2.166
6	ARMA(1,0)-GARCH(0,1)-N	3.17E-04	-	0.581	-	0.999	-	-	-1275.37	1.756
7	ARMA(0,1)-GARCH(0,1)-N	3.33E-04	-	-	0.534	0.999	-	-	-1316.81	1.813
8	ARMA(1,1)-GARCH(0,1)-N	3.13E-04	-	0.455	0.192	0.999	-	-	-1266.51	1.745
9	ARMA(0,0)-GARCH(1,1)-N	3.30E-01	0.350	-	-	0.000	-	-	-1517.38	2.088
10	ARMA(1,0)-GARCH(1,1)-N	1.42E-03	0.007	0.582	-	0.988	-	-	-1273.35	1.755
11	ARMA(0,1)-GARCH(1,1)-N	1.08E-03	0.007	-	0.534	0.990	-	-	-1314.88	1.812
12	ARMA(1,1)-GARCH(1,1)-N	1.38E-03	0.007	0.457	0.191	0.989	-	-	-1264.61	1.744
13	ARMA(0,0)-GJR-GARCH(1,0)-N	3.31E-01	0.288	-	-	-	0.118	-	-1516.19	2.087
14	ARMA(1,0)-GJR-GARCH(1,0)-N	3.10E-01	0.000	0.580	-	-	0.155	-	-1267.93	1.747
15	ARMA(0,1)-GJR-GARCH(1,0)-N	3.29E-01	0.000	-	0.524	-	0.147	-	-1309.87	1.805
16	ARMA(1,1)-GJR-GARCH(1,0)-N	3.05E-01	0.000	0.460	0.194	-	0.163	-	-1258.61	1.736
17	ARMA(0,0)-GJR-GARCH(0,1)-N	4.73E-04	-	-	-	0.999	-	-	-1575.21	2.166
18	ARMA(1,0)-GJR-GARCH(0,1)-N	3.17E-04	-	0.581	-	0.999	-	-	-1275.37	1.756
19	ARMA(0,1)-GJR-GARCH(0,1)-N	3.33E-04	-	-	0.534	0.999	-	-	-1316.81	1.813
20	ARMA(1,1)-GJR-GARCH(0,1)-N	3.13E-04	-	0.455	0.192	0.999	-	-	-1266.51	1.745
21	ARMA(0,0)-GJR-GARCH(1,1)-N	3.31E-01	0.288	-	-	0.000	0.118	-	-1516.19	2.088
22	ARMA(1,0)-GJR-GARCH(1,1)-N	1.20E-03	0.005	0.582	-	0.989	0.004	-	-1273.27	1.756
23	ARMA(0,1)-GJR-GARCH(1,1)-N	8.00E-04	0.006	-	0.534	0.991	0.002	-	-1314.85	1.813
24	ARMA(1,1)-GJR-GARCH(1,1)-N	1.02E-03	0.005	0.457	0.191	0.990	0.004	-	-1264.51	1.745
25	ARMA(0,0)-E-GARCH(1,0)-N	-7.64E-01	-0.091	-	-	-	0.558	-	-1519.12	2.091
26	ARMA(1,0)-E-GARCH(1,0)-N	-1.12E+00	-0.311	0.608	-	-	-0.199	-	-1246.25	1.717
27	ARMA(0,1)-E-GARCH(1,0)-N	-1.04E+00	-0.196	-	0.538	-	-0.051	-	-1303.05	1.795
28	ARMA(1,1)-E-GARCH(1,0)-N	-1.14E+00	-0.331	0.491	0.184	-	-0.233	-	-1234.68	1.703
29	ARMA(0,0)-E-GARCH(0,1)-N	-5.69E-05	-	-	-	1.000	-	-	-1574.94	2.166
30	ARMA(1,0)-E-GARCH(0,1)-N	-1.07E-01	-	0.581	-	0.901	-	-	-1275.76	1.757
31	ARMA(0,1)-E-GARCH(0,1)-N	-5.51E-05	-	-	0.534	1.000	-	-	-1316.53	1.813
32	ARMA(1,1)-E-GARCH(0,1)-N	-1.08E-01	-	0.455	0.192	0.901	-	-	-1266.90	1.746
33	ARMA(0,0)-E-GARCH(1,1)-N	-6.58E-01	-0.080	-	-	0.145	0.578	-	-1517.96	2.091
34	ARMA(1,0)-E-GARCH(1,1)-N	-1.04E+00	-0.311	0.610	-	0.075	-0.196	-	-1246.07	1.719
35	ARMA(0,1)-E-GARCH(1,1)-N	-1.10E+00	-0.199	-	0.538	-0.049	-0.060	-	-1303.00	1.797
36	ARMA(1,1)-E-GARCH(1,1)-N	-9.97E-01	-0.333	0.496	0.187	0.125	-0.230	-	-1234.01	1.703

Table A.7: *Overview of the first 36 of the fitted ARMA-GARCH models for the wind data*

	Model	ω	α	ϕ	θ	β	γ	ν	LLH	AIC
37	ARMA(0,0)-GARCH(1,0)-t	3.30E-01	0.355	-	-	-	-	100.000	-1518.88	2.091
38	ARMA(1,0)-GARCH(1,0)-t	3.34E-01	0.013	0.581	-	-	-	99.992	-1275.89	1.758
39	ARMA(0,1)-GARCH(1,0)-t	3.46E-01	0.033	-	0.530	-	-	99.999	-1317.21	1.815
40	ARMA(1,1)-GARCH(1,0)-t	3.31E-01	0.008	0.457	0.191	-	-	99.996	-1267.19	1.748
41	ARMA(0,0)-GARCH(0,1)-t	4.75E-04	-	-	-	0.999	-	99.999	-1575.95	2.169
42	ARMA(1,0)-GARCH(0,1)-t	1.07E-01	-	0.582	-	0.683	-	99.997	-1276.04	1.758
43	ARMA(0,1)-GARCH(0,1)-t	7.28E-02	-	-	0.535	0.797	-	99.998	-1317.92	1.816
44	ARMA(1,1)-GARCH(0,1)-t	3.13E-04	-	0.457	0.191	0.999	-	97.736	-1266.90	1.747
45	ARMA(0,0)-GARCH(1,1)-t	3.30E-01	0.355	-	-	0.000	-	100.000	-1518.88	2.092
46	ARMA(1,0)-GARCH(1,1)-t	1.44E-03	0.007	0.583	-	0.988	-	99.864	-1273.72	1.756
47	ARMA(0,1)-GARCH(1,1)-t	1.09E-03	0.007	-	0.535	0.990	-	100.000	-1315.61	1.814
48	ARMA(1,1)-GARCH(1,1)-t	1.40E-03	0.007	0.459	0.190	0.988	-	93.291	-1265.07	1.746
49	ARMA(0,0)-GJR-GARCH(1,0)-t	3.31E-01	0.297	-	-	-	0.112	100.000	-1517.86	2.090
50	ARMA(1,0)-GJR-GARCH(1,0)-t	3.10E-01	0.000	0.582	-	-	0.158	84.209	-1268.31	1.749
51	ARMA(0,1)-GJR-GARCH(1,0)-t	3.30E-01	0.000	-	0.526	-	0.148	99.999	-1310.66	1.807
52	ARMA(1,1)-GJR-GARCH(1,0)-t	3.05E-01	0.000	0.462	0.194	-	0.167	98.331	-1258.91	1.738
53	ARMA(0,0)-GJR-GARCH(0,1)-t	4.75E-04	-	-	-	0.999	-	99.999	-1575.95	2.169
54	ARMA(1,0)-GJR-GARCH(0,1)-t	1.07E-01	-	0.582	-	0.683	-	99.997	-1276.04	1.758
55	ARMA(0,1)-GJR-GARCH(0,1)-t	7.28E-02	-	-	0.535	0.797	-	99.998	-1317.92	1.816
56	ARMA(1,1)-GJR-GARCH(0,1)-t	3.13E-04	-	0.457	0.191	0.999	-	97.736	-1266.90	1.747
57	ARMA(0,0)-GJR-GARCH(1,1)-t	3.31E-01	0.297	-	-	0.000	0.112	100.000	-1517.86	2.092
58	ARMA(1,0)-GJR-GARCH(1,1)-t	1.24E-03	0.005	0.583	-	0.989	0.004	99.742	-1273.65	1.758
59	ARMA(0,1)-GJR-GARCH(1,1)-t	8.18E-04	0.006	-	0.535	0.991	0.002	99.993	-1315.58	1.815
60	ARMA(1,1)-GJR-GARCH(1,1)-t	1.03E-03	0.005	0.459	0.190	0.990	0.004	99.809	-1264.94	1.747
61	ARMA(0,0)-E-GARCH(1,0)-t	-7.59E-01	-0.087	-	-	-	0.565	100.000	-1520.91	2.095
62	ARMA(1,0)-E-GARCH(1,0)-t	-1.12E+00	-0.313	0.610	-	-	-0.199	100.000	-1246.96	1.720
63	ARMA(0,1)-E-GARCH(1,0)-t	-1.04E+00	-0.197	-	0.540	-	-0.050	100.000	-1304.03	1.798
64	ARMA(1,1)-E-GARCH(1,0)-t	-1.14E+00	-0.334	0.493	0.184	-	-0.234	100.000	-1235.14	1.705
65	ARMA(0,0)-E-GARCH(0,1)-t	-5.43E-05	-	-	-	1.000	-	100.000	-1575.70	2.169
66	ARMA(1,0)-E-GARCH(0,1)-t	-4.99E-05	-	0.582	-	1.000	-	99.981	-1275.45	1.757
67	ARMA(0,1)-E-GARCH(0,1)-t	-1.90E-01	-	-	0.535	0.815	-	99.995	-1317.92	1.816
68	ARMA(1,1)-E-GARCH(0,1)-t	-5.06E-05	-	0.457	0.191	1.000	-	99.998	-1266.64	1.747
69	ARMA(0,0)-E-GARCH(1,1)-t	-6.51E-01	-0.076	-	-	0.148	0.585	100.000	-1519.73	2.094
70	ARMA(1,0)-E-GARCH(1,1)-t	-1.02E+00	-0.313	0.612	-	0.084	-0.195	100.000	-1246.74	1.721
71	ARMA(0,1)-E-GARCH(1,1)-t	-1.10E+00	-0.200	-	0.539	-0.055	-0.060	100.000	-1303.96	1.799
72	ARMA(1,1)-E-GARCH(1,1)-t	-9.93E-01	-0.336	0.498	0.188	0.127	-0.231	100.000	-1234.47	1.705

Table A.8: *Overview of the last 36 of the fitted ARMA-GARCH models for the wind data*

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics*, 44:182–198.

Akaike, H. (1973). Information theory and an extension of the likelihood ratio principle. In Petrov, B. N., editor, *Proceedings of the Second International Symposium of Information Theory*, pages 257–281, Akademiai Kiado, Budapest.

Arzelà, C. (1895). Sulle funzioni di linee. *Mem. Accad. Sci. Ist. Bologna Cl. Sci. Fis. Mat*, 5(5):55–74.

Bedford, T. and Cooke, R. M. (2002). Vines — a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068.

Benth, F. E., Benth, J. S., and Koekebakker, S. (2008). *Stochastic Modeling of Electricity and Related Markets*, volume 11 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific.

Billingsley, P. (1995). *Probability and Measure*. Wiley, 3 edition.

Brechmann, E. C. (2010). Truncated and simplified regular vines and their applications. Diploma thesis, Technische Universität München.

Carley, H. and Taylor, M. (2003). A new proof of sklar's theorem. In Cuadras, C., Fortiana, J., and Lallena, J. R., editors, *Distributions with given marginals and Statistical Modelling*, pages 29–34. Kluwer, Dordrecht.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

Dißmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computation Statistics and Data Analysis*, 59:52–69.

Durante, F., Fernández-Sánchez, J., and Sempi, C. (2012). Sklar's theorem via regularization techniques. *Nonlinear Analysis*, 75:769–774.

Durante, F. and Sempi, C. (2016). *Principles of Copula Theory*. CRC Press.

ENTSO-E (2019). ENTSO-E transparency platform. `https://transparency.entsoe.eu`. Last accessed: 2019-08-22.

Ghalanos, A. (2019). *rugarch: Univariate GARCH models*. R package version 1.4-1.

Henry, L. and Wickham, H. (2019). *purrr: Functional Programming Tools*. R package version 0.3.2.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2018). *copula: Multivariate Dependence with Copulas*. R package version 0.999-19.1.

Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *The Canadian Journal of Statistics*, 22(1):47–64.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. In Rüschendorf, L., Schweizer, B., and Taylor, M., editors, *Distributions with fixed marginals and related topics*, volume 28 of *IMS Lecture Notes — Monograph Series*, pages 120–141. Institute of Mathematical Statistics, Hayward, CA.

Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley.

Manner, H. (2007). Estimation and model selection of copulas with an application to exchange rates. Research Memorandum 056, Maastricht University, Maastricht Research School of Economics of Technology and Organization (METEOR).

Manner, H., Fard, F. A., Pourkhanali, A., and Tafakori, L. (2019). Forecasting the joint distribution of australian electricity prices using dynamic vine copulae. *Energy Economics*, 78:143–164.

Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-square statistics for tests of fit. *The Annals of Statistics*, 3(3):599–616.

Morales-Nápoles, O., Cooke, R., and Kurowicka, D. (2010). About the number of vines and regular vines on n nodes. submitted for publication.

Müller, K. and Wickham, H. (2019). *tibble: Simple Data Frames*. R package version 2.1.3.

Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., and Erhardt, T. (2019). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.2.0.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schweizer, B. and Sklar, A. (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Mathematica*, 52:43–52.

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

Sklar, A. (1996). Random variables, distribution functions, and copulas — a personal look backward and forward. In Rüschendorf, L., Schweizer, B., and Taylor, M., editors, *Distributions with fixed marginals and related topics*, volume 28 of *IMS Lecture Notes — Monograph Series*, pages 1–14. Institute of Mathematical Statistics, Hayward, CA.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. (2019). *feather: R Bindings to the Feather 'API'*. R package version 0.3.4.

Wickham, H., François, R., Henry, L., and Müller, K. (2019). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.3.

Wickham, H., Hester, J., and Francois, R. (2018). *readr: Read Rectangular Text Data*. R package version 1.3.1.