CHALMERS | GÖTEBORG UNIVERSITY

MASTER'S THESIS

Modeling dependences between time series in the electricity market using copulas

STEFAN THAM

Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg, Sweden 2014

Thesis for the Degree of Master of Science

Modeling dependences between time series in the electricity market using copulas

Stefan Tham

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE − 412 96 Göteborg, Sweden
Göteborg, January 2014

Abstract

In this thesis we investigate the relationship of the Swedish electricity prices with outdoor temperatures and the water levels of the reservoirs from the hydro power plants. We will try to model the dependence between these time series with copulas. Methods of estimating copula models from the most popular copula families are described and the methods performances are compared in a simulation study. To help us see if our estimated copula models have a good fit, goodness-of-fit tests are performed. We find that several of our procedures on the data improve our models and that there are copulas that fit well on some of our data sets, especially the data that pairs the electricity prices and water levels.

Acknowledgments

I would like to thank my supervisor Patrik Albin for all his support and encouragement. He has been at most helpful to me during this time, with his consultations and advices. Also thank you to SMHI for providing me data from their weather stations and Nord Pool Spot for contributing with their data as well.

Contents

1 Introduction and Background **11**
1.1 The electricity market 12
1.2 Influences of the electricity prices 12

2 Theory **15**
2.1 Time Series with trend and seasonal components 15
2.2 Introduction to copulas 16
2.3 Dependence 18
2.4 Archimedean copulas 18
2.5 Examples of bivariate copula families 19
2.6 Mixed copulas 24
2.7 Extension to trivariate copulas 24
2.7.1 The copula families in three dimensions 24

3 Method **27**
3.1 Estimation of trend and seasonal components in time series 27
3.2 Estimating copula models 29
3.2.1 Pseudo observations 29
3.2.2 Empirical copula 29
3.2.3 Pseudo maximum likelihood estimation 29
3.2.4 Minimum distance estimation 31
3.2.5 Simulation study 32
3.3 Goodness-of-fit test for copulas 34
3.4 Data description 35

4 Results **39**
4.1 Copula modeling with electricity prices and water levels. 39
4.2 Copula modeling with electricity prices and temperatures 41
4.3 Mixed copulas 44
4.4 Trivariate data 44
4.5 Adaption to positive dependence 45

5 Conclusion **47**

A References **49**

1 Introduction and Background

Electricity prices have a direct relationship with natural physical conditions such as precipitation and outdoor temperatures. The water supply in the hydro plants reservoirs decides how much electricity can be produced and the outdoor temperatures rules how much electricity we need to heat our households.

In this thesis we will model the dependence between the time series of Swedish electricity prices, water levels in reservoirs and outdoor temperature by using copulas. Traditionally the pairwise dependence between variables are modeled with classical bivariate distributions such as the normal, lognormal and gamma. The problem with this approach is that the individual univariate distributions of the variables must belong to the same parametric family. Copulas do not have to deal with this type of problem as the marginals of a copula can be of different types and also be different from the joint distribution. We shall see that the copula models can capture important dependence features such as asymmetry and heavy-tail behavior.

The thesis is structured as followed: Chapter 1 describes information about the Swedish electricity market which will be needed to understand the data we will use for modeling. Chapter 2 introduces the theory of the mathematical models that this thesis employs and Chapter 3 deals with the estimation methods of these models and how to evaluate them. Chapter 4 presents the results of the modeling which is discussed in Chapter 5.

1.1 The electricity market

The Swedish electricity market has developed significantly and is very different from what it used to be 15 years ago. In 1996 a legislation was introduced, with a new system market that made it possible to produce and trade electricity in competition. The electricity market consists of several independent players. At one end we have the electricity *producers* that generates and feeds the network with electricity from hydro power, nuclear power, wind power and thermal power. On the other end are the electricity *consumers* which are everything from industries to households. The *network owners* are responsible of transmitting the electricity from the producers to the consumers. This is done via the national grid and regional and local networks. The *power trading company* are the ones that sell electricity to consumers. They buy electricity from either the producers directly, or from an organized market place such as Nord Pool. Nord Pool provides standard agreements between the actors on the market. They have a spot market for physical trading of financial contracts on the hours for the following 24-hour period. Sweden, Norway, Finland and Denmark form a joint, open electricity market without border tariffs for spot trading and the players can trade freely within all four countries [11].

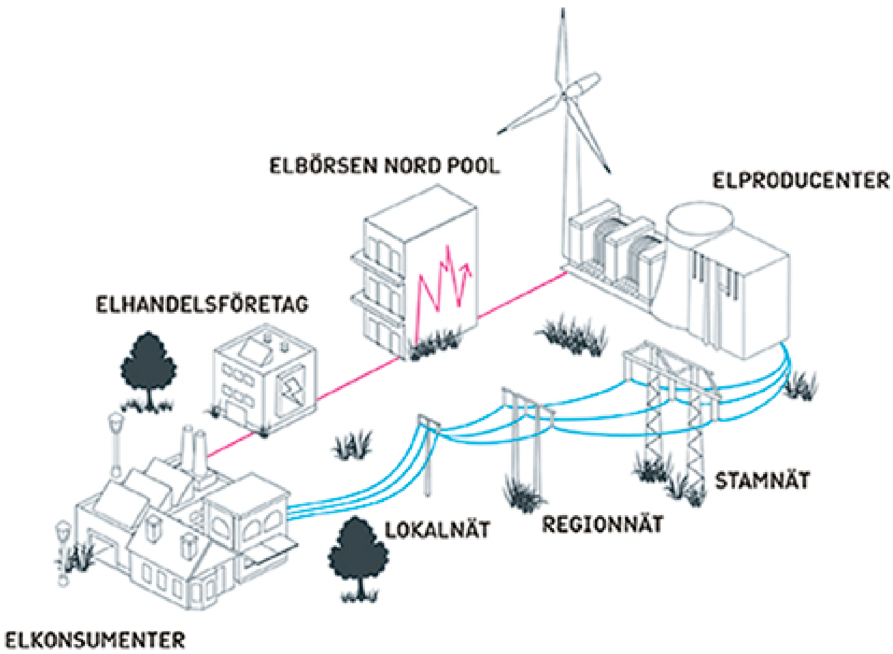

Figure 1: An overview of the electricity market. Source: Svenska Kraftnät [10]

1.2 Influences of the electricity prices

The electricity prices are reflected by its *supply* and *demand* and what makes the electricity trading market unique is the inability to store the asset. Electricity is therefore more or less directly produced in relation to the demand.

In Sweden, the main sources of energy are from hydro power and nuclear power where about half of the electricity produced are from hydro power and about half is from nuclear power. Water from rain and snow is stored in the hydro plants reservoirs and the water levels are thus highly dependent on the precipitation. When the availability of water is high the electricity price tend to be low and vice versa. The overall production capacity is affected by the water levels and also the number of nuclear power plants that are online, as they sometimes need downtime for maintenance.

The lead consumers of electricity are households that need heating and industries that are running. The demand varies over time as most industries only run during the day and are inactive at night and weekends. As for the households the people also consume according to the daily time where they consume less when they are out for work. Additionally the yearly season affects our need for electricity as the outdoor temperature decides how much is required to heat our homes. Low temperatures during the winter raises the demand which in turn raises the electricity prices. The most commonly recognized seasonal periods are thus the 24-hour period, the 7-day period and the 52-weeks period. Figures 2 demonstrates how the electricity prices varies throughout the day and figure 3 compares the prices and outdoor temperatures in the year 2012.

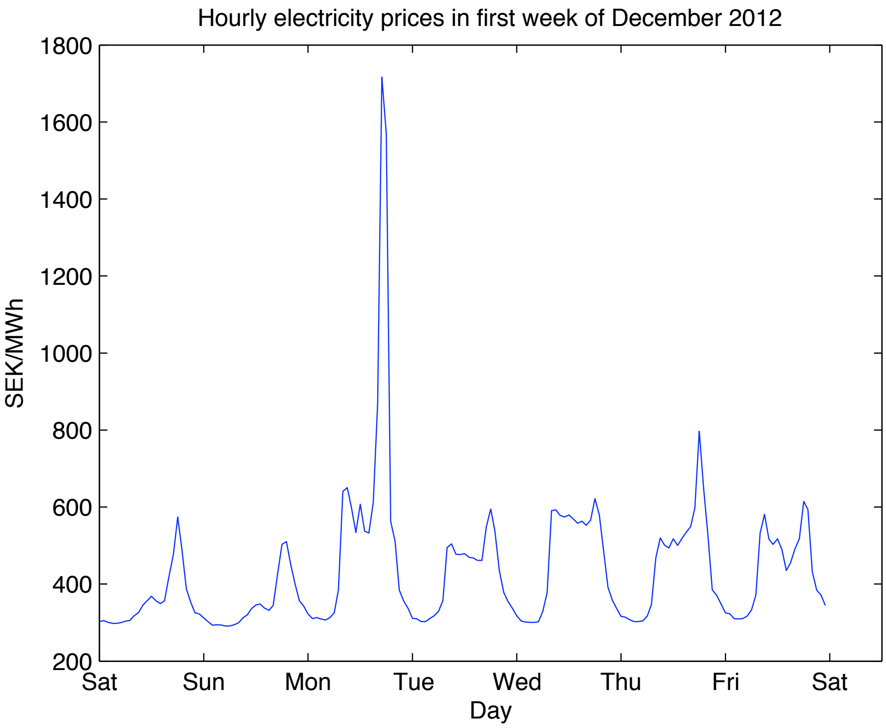

Figure 2: A plot of hourly electricity prices in first week of December 2012. The graph demonstrates how the prices relate to the hour of the day.

Electricity is also priced with respect to geographical location. Since most of the electricity is produced in the north of Sweden, the electricity needs to be transferred for long distances through the power grids to the souther regions where the population is denser and have a higher demand of electricity. This transmission leads to different prices between the regions. As a result, Sweden has been divided into four price regions (SE1-SE4).

Another factor that affects the electricity prices are political decisions. With the global warming at large, EU introduced in 2005 a new system for the electricity producers with the goal of minimizing the emission of greenhouse gases such as carbon dioxide. The system of trading with emission rights changed the behavior of the electricity prices drastically and lead to a significant increase of which is demonstrated in Figure 4.

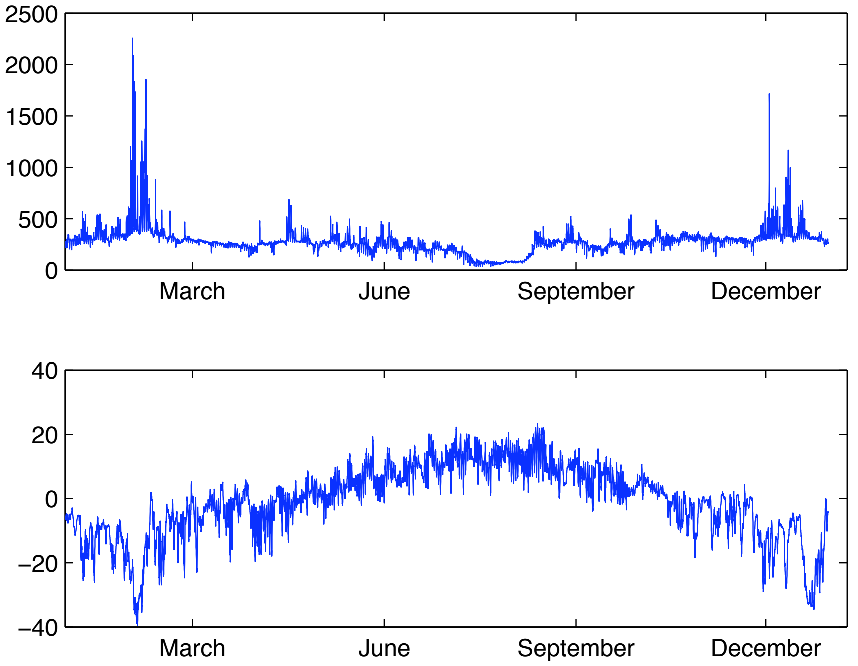

Figure 3: Electricity prices in 2012 (top) and temperature [Gällivare, Sweden] (bottom). As we can see, the prices are significantly higher during the winter season.

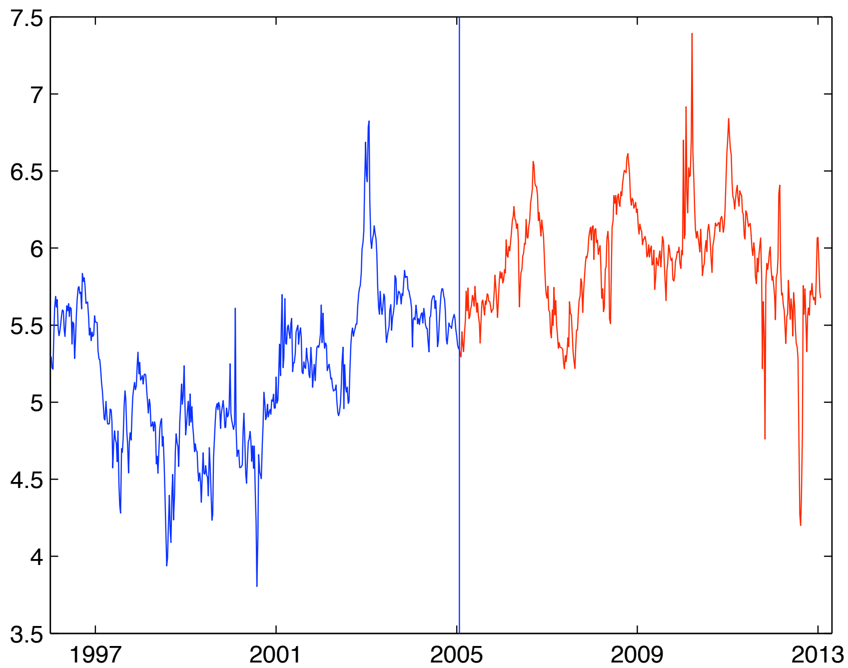

Figure 4: The log-scaled electricity prices from 1996 to 2012. The vertical line marks year 2005 when the emission legislation was introduced. We see a clear distinction in price behavior after 2005.

2 Theory

This chapter presents the theory that is necessary for modeling the dependence between time series with copulas. It starts with a short introduction to time series with seasonality, followed by the main topic, copulas. The fundamental definitions and theorems of copulas are explained and thereafter the different types and classes of copulas are presented.

2.1 Time Series with trend and seasonal components

A *time series* is a set of observations x_t that are observed at a specific time t. In this thesis we will consider a discrete times series, where the set of times that are recorded is a discrete set T_0. To investigate a time series, a suitable mathematical model needs to be selected in order to consider the unpredictability of the data. The time series x_t can then be assumed to be a *realization* of a *stochastic process* $\{X_t, t \in T\}$. A stochastic process is a family of random variables $\{X_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. The functions $\{X(w), w \in \Omega\}$ on T are then called the realizations of the stochastic process $\{X_t, t \in T\}$.

When modeling with time series, it is important to be aware of possible seasonal affects on the data and make proper adjustments in the case of its presence. We have previously mentioned how electricity prices are affected by the date and time. A common model for times series that contemplate these factors, is the additive model [2],

$$X_t = m_t + s_t + Y_t. \tag{1}$$

The time series X_t in Equation 1 is decomposed into three components. The *trend component* m_t is a slowly changing function. The *seasonal component* s_t is a periodic function with a known period d. A periodic function s_t with period d is a function with the property that $s_t = s_{t+d}$ and $\sum_{j=1}^{d} s_j = 0$. The final component Y_t is a *random component* that is *weakly stationary*. A time series $\{Y_t, t \in \mathbb{Z}\}$ is said to be stationary in the weak sense if

(i) $E|Y_t|^2 < \infty, \quad$ for all $t \in \mathbb{Z}$;

(ii) $E(Y_t) = M, \quad$ for all $t \in \mathbb{Z}$;

(iii) $Cov(Y_r, Y_s) = Cov(Y_{r+t}, Y_{s+t}), \quad$ for all $r, s, t \in \mathbb{Z}$.

By removing the trend and seasonal components, further analysis and model fitting are easier accomplished on the stationary series Y_t. A method of estimating and removing the deterministic components m_t and s_t is explained in Section 3.1.

2.2 Introduction to copulas

Copulas are parametrically specified joint distributions generated from given marginals. They have only quite recently become popular to statisticians and is now widely used in many fields such as finance, actuarial sciences and biostatistics for its wide range of dependence structures and flexibility in modeling. The definition is as follows:

Definition 1. *A 2-dimensional copula is a function* $C : \mathbf{I^2} \to \mathbf{I}$*, where* $\mathbf{I} = [0, 1]$*, with the following properties:*

1. $C(u, 0) = 0 = C(0, v)$ *for all* $u, v \in \mathbf{I}$*;*
2. $C(u, 1) = u$ *and* $C(1, v) = v$ *for all* $u, v \in \mathbf{I}$*;*
3. $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ *for all* $u_1, u_2, v_1, v_2 \in \mathbf{I}$ *such that* $u_1 \leq u_2$ *and* $v_1 \leq v_2$*.*

In other words, C is a 2-dimensional cumulative distribution function with marginals that are uniformly distributed over $[0, 1]$.

The definition is expandable to the multivariate case and the definition of a n-dimensional copula is analogous to the 2-dimensional. However for simplicity and readability, the theory and examples in thesis will be for the bivariate case unless otherwise stated. The density $c(u, v)$ of copula $C(u, v)$ is written as

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}.$$

The graph of a copula is a continuous surface within the unit cube $\mathbf{I^3}$ with vertices $(0, 0, 0), (1, 0, 0), (0, 1, 0)$ and $(1, 1, 1)$. Another way to present the graph of a copula is with a contour diagram which plots its level curves.

An essential property of copulas explained by Sklar, that sets the foundation of copula modeling is described in the following theorem.

Theorem 1 (Sklar's theorem)**.** *Let* H *be a joint distribution function with marginal functions* F *and* G*. There then exists a copula* C*, such that*

$$H(x, y) = C(F(x), G(y)). \tag{2}$$

Conversely, if C *is a copula and* F *and* G *are distribution functions, then the function* $H(x, y) = C(F(x), G(y))$ *is a distribution function with margins* F *and* G*. Furthermore if* F *and* G *are continuous, then* C *is unique.*

Sklar's theorem explains the relationship between bivariate distributions and their univariate marginals. When given two arbitrary marginals, we can find a bivariate distribution that "couples" them, simply by plugging a couple of univariate margins into a function which satisfies the copula definition. This eliminates restrictions the traditional method of constructing multivariate distributions has where the margins need to be of the same type. With Sklar's theorem the marginal distributions do not need to be in any way similar to each other. It also offers the flexibility when modeling in that the estimation problem can be decomposed in two steps; first by estimating the marginals and then the copula.

Sklars theorem is expressed in terms of random variables in the following manner;

Theorem 2. *Let X and Y be continuous random variables with distribution functions $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$, respectively. Then there exists a unique copula C such that*

$$P(X \leq x, Y \leq y) = C(F(x), G(y))$$

where C(u,v) is a distribution of the pair $(U, V) = (F(x), G(y))$ whose margins are uniform on the unit interval $[0, 1]$.

The copula for two independent random variables has the form $C(u, v) = uv$ and is called the product copula. Figure 5 shows the graph and contour diagram of the product copula.

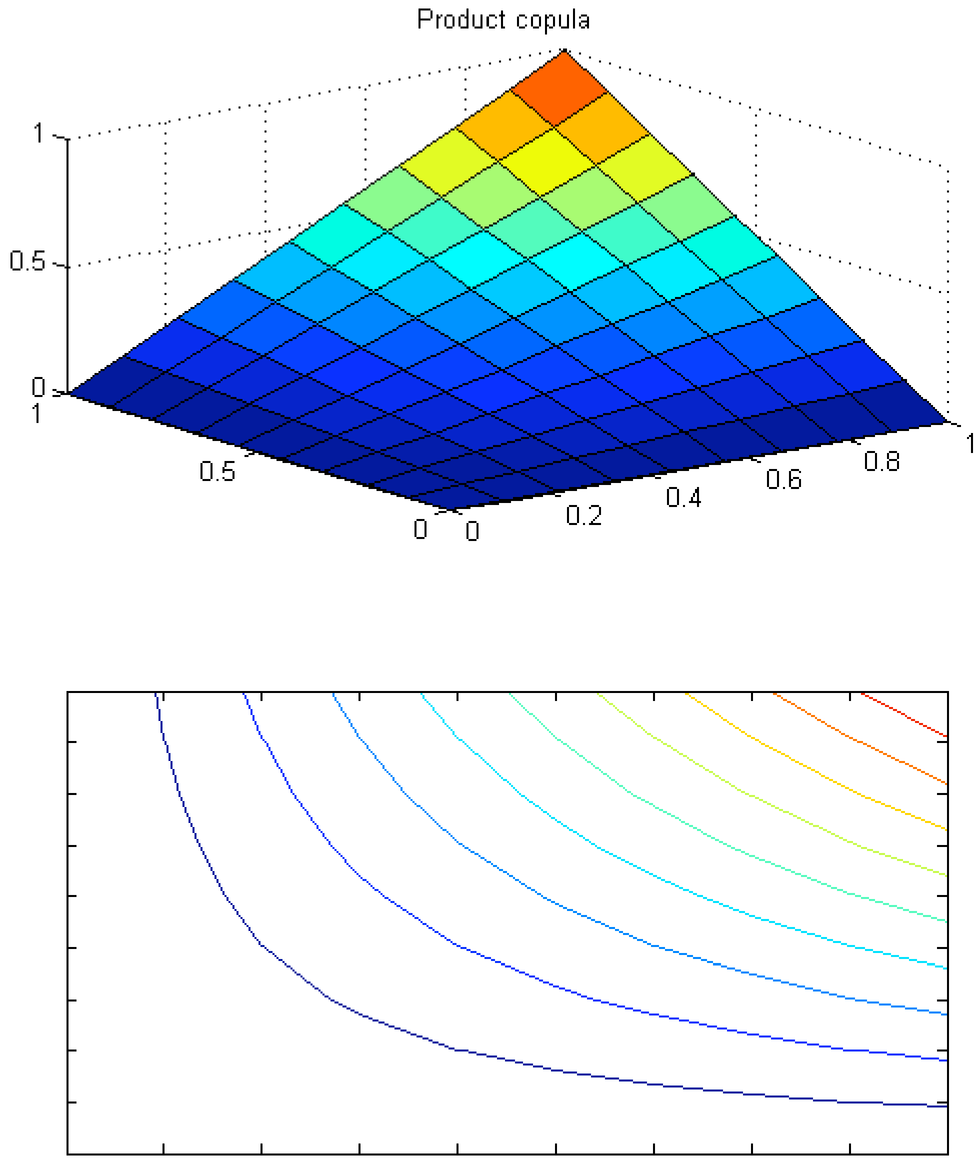

Figure 5: Graph and contour diagram of the product copula

An important property that will be of use for our inference is that copulas are invariant under strictly increasing transformations.

Theorem 3. *Let X and Y be continuous random variables with copula C. Then if g and h are strictly increasing functions on the range of X and Y, the random variables $g(X)$ and $h(Y)$ have the exact same copula C. Thus C is invariant under strictly increasing transformations of X and Y.*

2.3 Dependence

To measure the dependence between random variables, we shall look at a rank based measure called *Kendall's tau* which is defined by *concordance*. A pair of random variables X and Y are concordant if "large" values of one of the random variables is associated with "large" values of the other, and "small" values of one them is associated with "small" values of the other. More formally, let (x_i, y_i) and (x_j, y_j) be two observations from a random vector (X, Y). (x_i, y_i) and (x_j, y_j) are concordant if $x_i < x_j$ and $y_i < y_j$ or if $x_i > x_j$ and $y_i > y_j$. On the other hand (x_i, y_i) and (x_j, y_j) are said to be *discordant* if $x_i < x_j$ and $y_i > y_j$ or if $x_i > x_j$ and $y_i < y_j$.

Now to define the sample version of Kendall's tau, let $\{(x_1, y_1), ..., (x_n, y_n)\}$ be a sample from a vector (X, Y) of continuous random variables and let c be the number of concordant pairs and d the number of discordant pairs. The Kendall's tau for the sample, τ_n is

$$\tau_n = \frac{c-d}{c+d}.$$

Kendall's tau range from -1 to 1, where a positive value notions that large or small values of random variables tend to occur together while a negative value expresses that large values of one variable occurs with small values of the other variable.

Kendall's tau for random variables X and Y with copula C denoted as τ is defined by

$$\tau(C) = 4 \iint C(u, v) c(u, v) du dv - 1,$$

where $c(u, v) = \partial^2 C(u, v)/\partial u \partial v$, assuming that this derivative exists.

2.4 Archimedean copulas

Archimedean copulas is a very important class of copulas with many nice properties that makes them suitable for modeling. Their popularity comes from that

- they contain a large variety of families that together captures a wide range of dependences;
- the distribution functions have explicit formulas;
- they are easy to estimate.

The construction of Archimedean copulas is explained by the following. Let φ be a continuous, convex and strictly decreasing function from $\boldsymbol{I}$ to $[0, \infty]$ such that $\varphi(1) = 0$. Moreover let $\varphi^{[-1]}$ be the *pseudo inverse* which is defined as

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & \text{if } 0 \le t \le \varphi(0). \\ 0, & \text{if } \varphi(0) \le t \le \infty. \end{cases}$$

There exists copulas $C(u, v)$ that can be written in the form of

$$C(u, v) = \varphi^{[-1]}[\varphi(u) + \varphi(v)]. \tag{3}$$

Copulas that can be constructed in this form are the Archimedean copulas. If a copula C can be defined as in Equation 3, then the function φ is called the *generator* of C.

2.5 Examples of bivariate copula families

When modeling with copulas, C is unknown but often assumed to belong to a parametric family

$$\mathcal{C}_0 = \{C_\theta : \theta \in \Theta\}.$$

Θ is an open subset of $\mathbb{R}^p$ for some integer $p \geq 1$ and C_θ is a copula for every $\theta \in \Theta$. A copula family is thus characterized by this vector of parameters θ. We will now present the bivariate copula families that we will use for our modeling. They have been chosen because they are easy to work with, but can still capture different types of structures. All of them with the exception of the Student's t copula have only one parameter where as the Student's t have two. For more examples of families, Nelsen [12] provides an extensive compendium of the most popular ones.

- The *Gaussian* family, also known as the *Normal* family, is given by

$$\begin{aligned} C_\theta(u, v) &= N_\theta(\Phi^{-1}(u), \Phi^{-1}(v)) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(u)} \exp\left[\frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right] dsdt, \end{aligned}$$

  where Φ denotes the standard normal distribution function and N_θ is the standard normal bivariate distribution with correlation parameter θ which is restricted in $(0, 1)$. The Gaussian copula is symmetric and generates the standard joint normal distribution function whenever the margins are standard normal. Figure 6 shows a visual example of the Gaussian copula distribution.

- The *Student's t* copula family has two parameters and is defined by

$$\begin{aligned} C_{\theta,\nu}(u, v) &= T_{\theta,\nu}(T_\nu^{-1}(u), T_\nu^{-1}(v)) \\ &= \int_{-\infty}^{T_\nu^{-1}(u)} \int_{-\infty}^{T_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \left(1 + \frac{s^2 + t^2 - 2\rho st}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} dsdt, \end{aligned}$$

  where T_ν is the univariate Student's t distribution function with ν degrees of freedom and $T_{\theta,\nu}$ is the bivariate Student's t distribution with correlation parameter θ and ν degrees of freedom. The range of θ is in $(0, 1)$. The parameter ν controls the heaviness of the tails. When the number of degrees of freedom diverges to infinity, the copula converges to the Gaussian copula. An advantage with the Student's t copula over the Gaussian is that it captures more observations in the tails, where the extreme dependent values are observed.

- The *Frank* family is an Archimedean family given by

$$C_\theta(u, v) = -\frac{1}{\theta} \log\left(1 + \frac{\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1}\right).$$

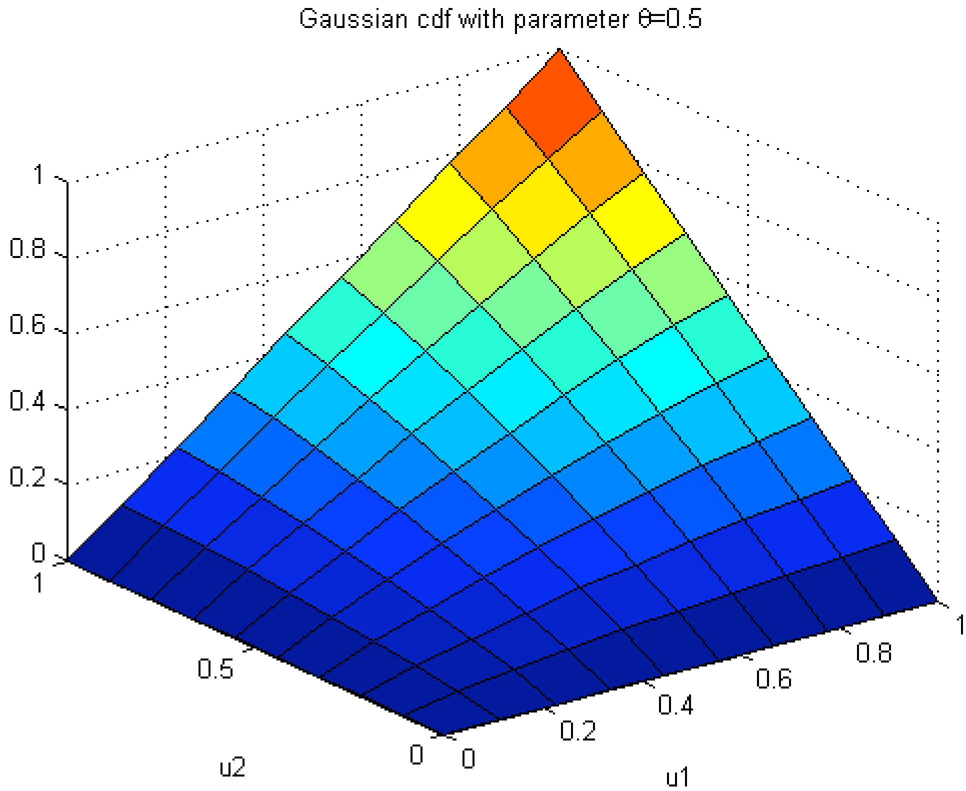

Figure 6: A plot of the Gaussian copula distribution with $\theta = 0.5$

The dependence parameter may take on any real value $(-\infty, \infty)$. When $\theta = 0$, the copula corresponds to the product copula. The main advantage with the Frank copula is that it can take on negative dependence between the marginals unlike some others. It is also symmetric, however the dependence in the tails are weak compared to the Gaussian and Student's t [15]. The generator of the Frank copula is

$$\varphi_\theta(t) = -\log \frac{e^{-\theta t} - 1}{e^{-\theta - 1}}.$$

- The *Clayton* family also belongs to the Archimedean class and is defined as

  $$C_\theta(u, v) = \left[\max\{0, u^{-\theta} + v^{-\theta} - 1\}\right]^{-1/\theta}.$$

  θ has ranges in $[-1, \infty)\backslash\{0\}$ and the marginals become independent as θ approaches 0. The Clayton copula has the generator $\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta} - 1)$.

- The third Archimedean family that we will study is the *Gumbel* copula.

  $$C_\theta(u, v) = \exp\left[-\left((-\log(u))^\theta + (-\log(v))^\theta\right)^{1/\theta}\right].$$

  Here $\theta \in [1, \infty)$. The Gumbel does not allow negative dependence. It has a strong right tail dependence and weak left tail dependence and is a good choice if outcomes are strongly correlated at high values and less correlated at low values. The Gumbel generator $\varphi(t) = (-\log t)^\theta$.

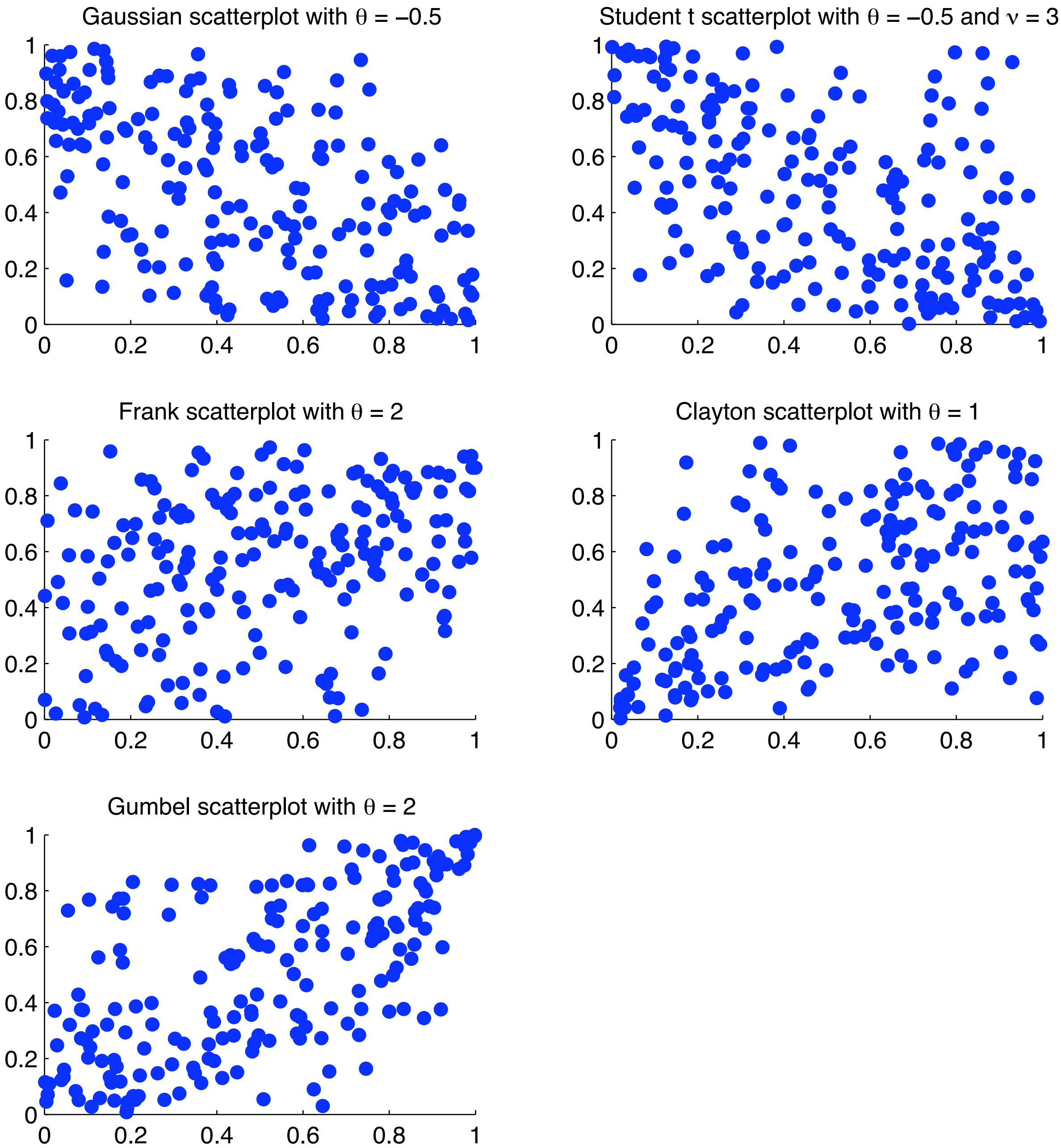

Figure 7: Examples of scatterplots with sample size 200 from the five copula families.

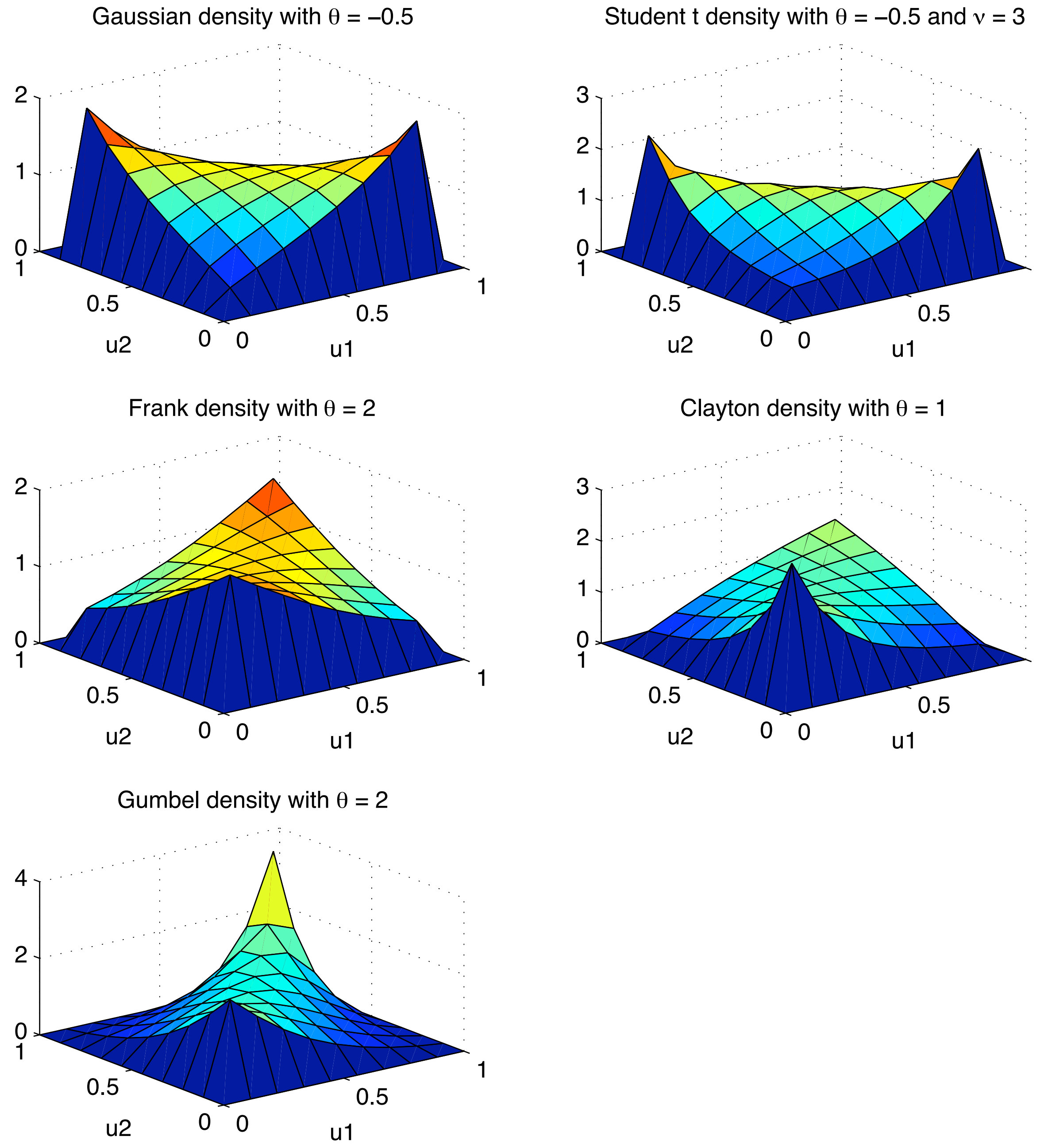

Figure 8: Examples of the densities of the five copula families.

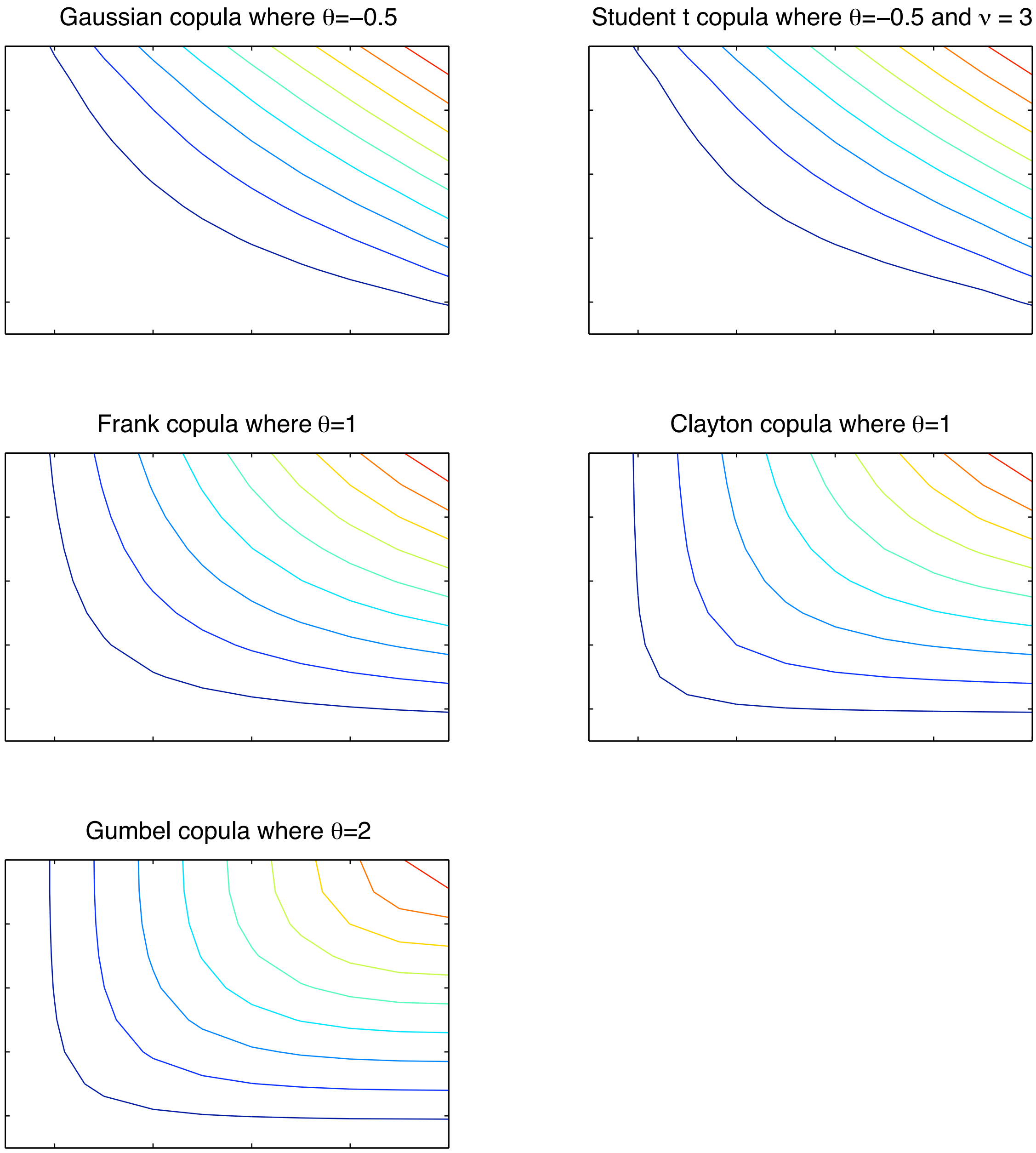

Figure 9: Examples of the contour plots of the five copula distributions.

2.6 Mixed copulas

A *mixed copula* combines different copulas with the intention of creating more general dependence structures. The mixed copula has the form of

$$C_{\text{mix}}(u, v, \beta, \theta) = \beta_1 C_1(u, v, \theta_1) + ... + \beta_s C_s(u, v, \theta_s),$$

where $\beta = (\beta_1, ..., \beta_s)$ are the *weight parameters* with the conditions $\beta_1 + ... + \beta_s = 1$ and $0 \leq \beta_k \leq 1, k = 1, ..., s$. The sequence $\{C_k(u, v, \theta_k\}_{k=1}^{s}$ are known copulas with parameters $\{\theta_k\}_{k=1}^{s}$. $\theta = (\theta_1, ..., \theta_s)$ are called the *associate parameters* in the mixture copula and that controls the degree of dependence. The weight parameters or the *shape parameters* $\beta = (\beta_1, ..., \beta_s)$ decides the shape of the dependence and how much load the respective copulas have in the mixed copula. The higher the value of θ_j, the more the j-th copula is appropiate for the data. A mixture of copulas is also a copula. For this study we will consider a mixed copula that includes two copulas,

$$C_{\text{mix}}(u, v, \beta, \theta_1, \theta_2) = \beta C_1(u, v, \theta_1) + (1 - \beta) C_2(u, v, \theta_2).$$

The copulas C_1 and C_2 will be from the five families previously described.

2.7 Extension to trivariate copulas

In this section we will show how copulas work in the three dimensional case. Most of the definitions, theorems and inference methods from the earlier sections are analogues to the trivariate case. For example Sklar's theorem in the trivariate case is expressed in that a 3-dimensional continuous distribution function F with marginal distributions F_1, F_2 and F_3, we can find a 3-dimensional copula C satisfying

$$F(x_1, x_2, x_3) = C(F_1(x_1), F_2(x_2), F_3(x_3)).$$

2.7.1 The copula families in three dimensions

Recall how bivariate Archimedean copulas are constructed as showed by Equation 3 in Section 2.4.

$$C(u, v) = \varphi^{[-1]}[\varphi(u) + \varphi(v)].$$

We can include a third marginal distribution as trivariate Archimedean copulas with generator φ are represented analogously by

$$C(u, v, w) = \varphi^{[-1]}(\varphi(u) + \varphi(v) + \varphi(w)).$$

The requirement is that φ^{-1} has to be completely monotonic on $[0, \infty]$ [3]. The corresponding trivariate Archimedean copulas that we are working with have one dependence parameter as well.

We are now ready to present our copula families in three dimensions:

- Let $\boldsymbol{\theta}$ be a symmetric, positive definite 3-dimensional matrix with diagonal elements 1, and $\Phi_{\boldsymbol{\theta}}$ the standardized trivariate normal distribution with correlation matrix $\boldsymbol{\theta}$. The trivariate Gaussian copula is defined as:

  $$C_{\boldsymbol{\theta}}(u, v, w) = \Phi_{\boldsymbol{\theta}}(\Phi^{-1}(u), \Phi^{-1}(v), \Phi^{-1}(w)),$$

  where Φ^{-1} is the inverse of the standard normal univariate normal distribution function.

- Let $\boldsymbol{\theta}$ be a symmetric, positive definite 3-dimensional matrix with diagonal elements 1 and $T_{\boldsymbol{\theta},\nu}$ be the standardized trivariate Student's t distribution with correlation matrix $\boldsymbol{\theta}$ and ν degrees of freedom. The trivariate Student's t copula is defined as:

$$C_{\theta,\nu}(u,v,w) = T_{\boldsymbol{\theta},\nu}(T_\nu^{-1}(u), T_\nu^{-1}(v), T_\nu^{-1}(w)),$$

  where T_ν^{-1} is the inverse of the univariate Student's t distribution with ν degrees of freedom.

- The generator for the Frank copula is

$$\varphi(u) = \log\left(\frac{e^{-\theta u}-1}{e^{-\theta}-1}\right),$$

  with

$$\varphi^{-1}(t) = -\frac{1}{\theta}\log(1+e^t(e^{-\theta}-1)),$$

  which is completely monotonic for $\theta > 0$. This gives the three dimensional Frank copula

$$C(u,v,w) = -\frac{1}{\theta}\log\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)(e^{-\theta w}-1)}{(e^{-\theta}-1)^2}\right), \quad \theta > 0.$$

  As opposed to the bivariate case, the dependence parameter for the Frank copula can only be positive.

- The generator for the Clayton copula is given by $\varphi(u) = u^{-\theta} - 1$, with inverse $\varphi^{-1}(t) = (t+1)^{-\frac{1}{\theta}}$. It is completely monotonic for $\theta > 0$. The copula is

$$C(u,v,w) = \left[u^{-\theta}+v^{-\theta}+w^{-\theta}-2\right]^{-\frac{1}{\theta}}, \quad \theta > 0.$$

- The Gumbel copula has generator $\varphi(u) = (-\log u)^{-\theta}$ and the inverse $\varphi^{-1}(t) = exp(-t^{fra1\theta}$ completely monotonic if $\theta > 1$. This gives the copula

$$C(u,v,w) = \exp\left\{-\left[(-\log u)^\theta + (-\log v)^\theta + (-\log w)^\theta\right]^{\frac{1}{\theta}}\right\}, \quad \theta > 1.$$

3 Method

In this chapter we will present the methods of copula modeling. First we explain how to deal with the trend and and seasonal components in time series. Thereafter we describe methods of estimating copula models and how to evaluate them with a goodness-of-fit test. The chapter is concluded with a description of the data to be analyzed in our modeling.

3.1 Estimation of trend and seasonal components in time series

Recall the additive model of a time series:

$$X_t = m_t + s_t + Y_t.$$

Brockwell and Davies [2] suggests a method of estimating the trend and seasonal component with a moving average technique which goes as follows: Suppose that we have observations of a time series $\{x_1, x_2, ..., x_n\}$ and assume that there exists a trend and seasonal component. In the first step the trend component is estimated with a symmetric moving average of the series x_t. For an even period d and an integer q such that $2q = d$, m_t is estimated by

$$\hat{m}_t = (0.5x_{t-q} + x_{t-q+1} + ... + x_{t+q-1} + 0.5x_{t+q})/d, \quad q < t \leq n - q.$$

If d is odd such that $2q + 1$ for some integer q, we instead estimate m_t by

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t+j} \quad q + 1 \leq t \leq n - q.$$

Since x_t is not observed for $t \leq 0$ and $t > n$, we define $x_t := x_1$ for $t < 1$ and $x_t := x_n$ for $t > n$. The next step is to the estimate seasonal component s_t, which we will denote by $\hat{s}_t$. For $k = 1, ..., d$ define the weight function w_k as the average of the deviations

$$\{x(k + jd) - \hat{m}(k + jd) : q < k + jd \leq n - q\}.$$

We estimate the seasonal component by

$$\hat{s}_k = w_k - \frac{1}{d} \sum_{i=1}^{d} w_i, \; k = 1, ..., d$$

and $\hat{s}_k = \hat{s}_k = \hat{s}_{k-d}, k > d$. Now we deseasonalize our data to get the series $d_t = X_t - \hat{s}_t$.

We should now reestimate the trend, $\tilde{m}_t$ from the non-seasonal data $\{d_t\}$. The reestimation of the trend is done in order to have a parametric form for the trend that can be extrapolated for the purpose of prediction and simulation. The estimated random component is given by

$$\hat{Y}_t = x_t - \tilde{m}_t - \hat{s}_t, \quad t = 1, ..., n.$$

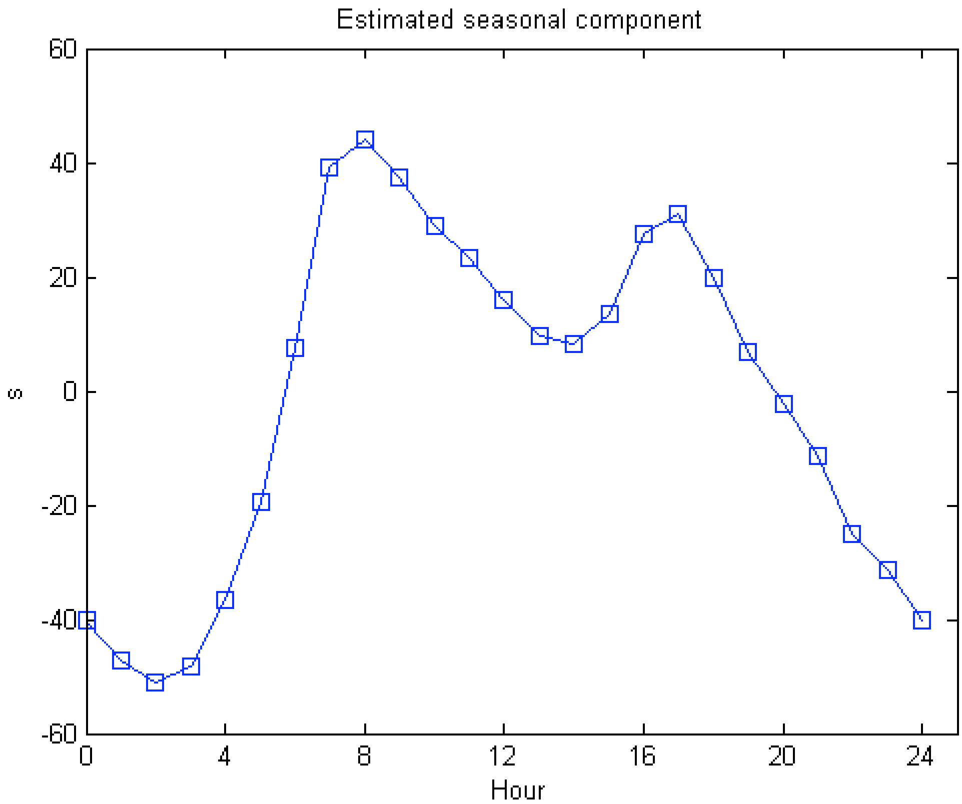

Figure 10: Estimated seasonal component of hourly electricity prices in 2012 with period 24.

3.2 Estimating copula models

Suppose that we have an independent and identically distributed (IID) random sample $(X_1, Y_1), ..., (X_n, Y_n)$. The sample is assumed to be drawn from a copula distribution $C(F(x), G(y))$, with associated marginals F and G, and belongs to a parametric family $\{C_\theta : \theta \in \Theta\}$ where θ is a q-dimensional parameter.

Following Sklar's theorem, the model estimation problem can be decomposed in that the marginals are estimated independently of the copula. Selecting the copula is done by first specifying a parametric family and then estimating its dependence parameter. We will present two methods to estimate the dependence parameter θ; the pseudo maximum likelihood estimate and the minimum distance estimate. These methods will be investigated in a simulation study to see their effectiveness and accuracy.

3.2.1 Pseudo observations

When modeling with copulas, the marginals F and G are often unknown. The natural replacement are their respective empirical distribution functions. A rescaled version of the empirical distribution will be used and is defined as

$$\hat{F}_n(x) = \frac{1}{n+1}\sum_{i=1}^{n}\mathbf{1}(X_i \leq x), \quad \hat{G}_n(y) = \frac{1}{n+1}\sum_{i=1}^{n}\mathbf{1}(Y_i \leq y).$$

The scaling factor of $1/(n+1)$ instead of the traditional $1/n$ is to avoid problems with the log-likelihood function, that will later be introduced, to blow up at the boundary of $[0,1]^2$. The transformation of the sample is a collection $(U_1, V_1), ..., (U_n, V_n)$ where

$$(U_k, V_k) = (\hat{F}_n(X_k), \hat{G}_n(Y_k)), \quad k = 1, ..., n$$

and will be called *pseudo-observations*. This works as a transformation of the data to the "copula scale", which is the unit cube. The pseudo-observations can be interpreted as a sample from the underlying copula and the estimation methods and goodness of fit tests will be based on this. However they are not mutually independent and their components are only approximately uniform $(0, 1)$. These features must be taken into account when dealing with any inference based on these transformations.

3.2.2 Empirical copula

A nonparametric and objective representation of the underlying copula is the *empirical copula*. For the bivariate case, the empirical copula constructed with pseudo-observations is defined as followed,

$$C_n(u, v) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(U_i \leq u, V_i \leq v), \quad (u, v) \in [0,1]^2.$$

This is a consistent estimation of the true underlying copula C and a weak convergence of the empirical copula has been shown by Fermanian [5].

3.2.3 Pseudo maximum likelihood estimation

The maximum likelihood estimate is a well known method that is popular for its efficiency and nice properties. Given an independent and identically distributed sample

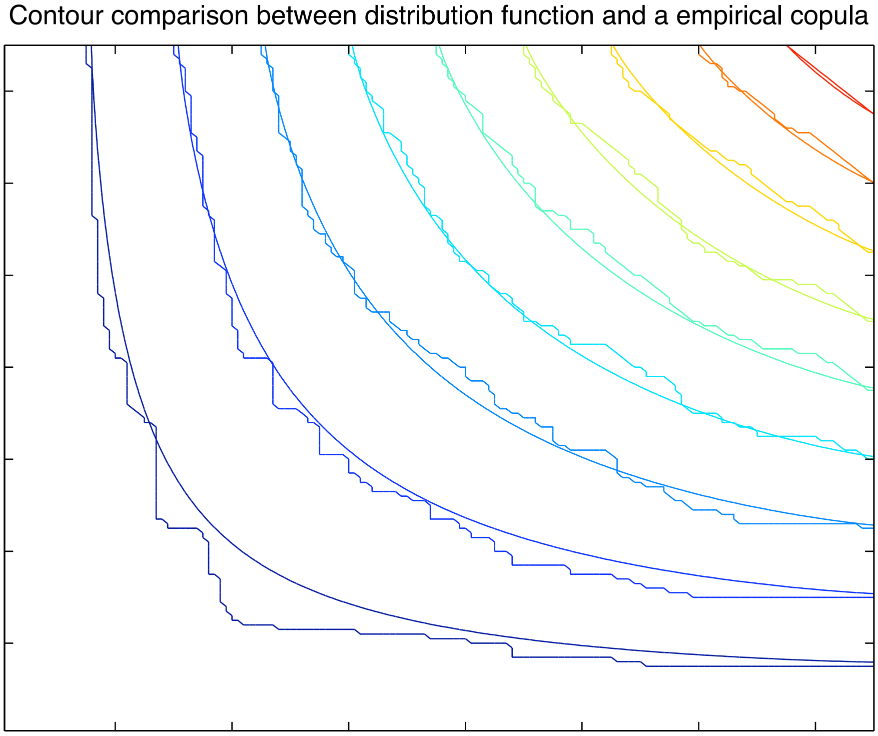

Figure 11: Contour plot comparison between a Gaussian copula with parameter $\theta = 0.25$ and a empirical copula sample generated from the same copula. The solid lines is the contour plot from a Gaussian copula and the dashed line represents the empirical copula.

$\{(X_i, Y_i)\}_{i=1}^n$ from a distribution $C_\theta(F(x), G(y))$, the classic maximum likelihood estimator of the dependence parameter θ is the value that maximizes the log-likelihood function:

$$\ell(\theta) = \sum_{i=1}^n \log[c_\theta(F(X_i), G(Y_i)].$$

This requires that C is absolutely continuous with density c_θ. In our case the marginal distribution functions F and G are unknown. This is solved by making use of the pseudo observations. With this we instead get the pseudo likelihood function

$$\ell(\theta) = \sum_{i=1}^n \log[c_\theta(U_i, V_i)].$$

The pseudo maximum likelihood estimate (PML), denoted by $\hat{\theta}_{\mathrm{PML}}$ is the value that solves the equation

$$\frac{\partial}{\partial\theta}\ell(\theta) = \sum_{i=1}^n \frac{\dot{c}_\theta(F(X_k), G(Y_k))}{c_\theta(F(X_k), G(Y_k))} = 0, \tag{4}$$

where $\dot{c}_\theta = (\partial c_\theta/\partial\theta_1, \partial c_\theta/\partial\theta_2)$. Genest et al [6] shows that the root of Equation 4 is unique. Moreover $\hat{\theta}_{\mathrm{PML}}$ also has all the other properties as the classic maximum likelihood estimate, such as consistency and asymptotic normality. The downside with this method is that the likelihood function can be difficult to compute and involves numerical work and also requires the existence of a density.

The pseudo maximum likelihood extends to the three dimensional case in the sense of estimating a multidimensional dependence parameter $\theta = (\theta_1, ..., \theta_q)$ from a trivariate copula $C(F_1(x_1), F_2(x_2), F_3(x_3))$. A random sample $\{(X_{1i}, X_{2i}, X_{3i})\}_{i=1}^n$ from a distribution C_θ would have the estimate $\hat{\theta}_{\text{PML}}$ of θ as the solution to the system

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \log[c_\theta(U_{1i}, U_{2i}, U_{3i})] = 0 \quad (1 \leq j \leq q).$$

3.2.4 Minimum distance estimation

The minimum distance method finds the copula that is measured closest to the empirical copula. Assume that the distribution function (X, Y) is associated with copula D and we want to fit the data to a family of copulas $\{C_\theta; \theta \in \Theta\}$. Define the minimum distance functional T on the space of the copula by

$$T(D) \equiv \arg\min_{\theta} d(D, C_\theta).$$

Here d is a distance statistic that measures how close two distributions C and D are with each other. The measures that we will consider are the Kolmogorov-Smirnov statistic (KS),

$$d(C, D) = \sup_{u \in (0,1)^2} |C(u) - D(u)|.$$

There is also the Cramér-von Mises statistic (CvM),

$$d(C, D) = \int_{(0,1)^2} (C(u) - D(u))^2 \, du.$$

When given a sample $\{X_i, Y_i\}_{k=1}^n$ that is assumed to belong to a specified copula family, the minimum distance estimator (MD), denoted by $\hat{\theta}_n^{MD}$, is defined as

$$\hat{\theta}_n^{\text{MD}} = T(C_n) = \arg\min_{\theta} d(C_n, C_\theta),$$

where C_n is the empirical copula based on $\{X_i, Y_i\}_{k=1}^n$. In other words, the fitted copula of a parametric family is the one that is measured closest to the empirical copula.

T has some nice properties that are investigated by Tsukahara [16]; under the basic assumptions,

(i) For every copula $D, t \mapsto d(D, C_t)$ is continuous;

(ii) $d(C_t, C_\theta) = 0$ if and only if $t = \theta$.

Assuming that Θ is compact and that (i), and (i) hold, we get that:

1. $T(D)$ exists for every copula D and $T(C_\theta) = \theta$ uniquely.

2. T is continuous at C_θ: for any $\epsilon > 0$, there exists a $\delta > 0$ such that $d(D, C_\theta) < \delta \implies |T(D) - \theta| < \epsilon$.

Furthermore the MD estimator $\hat{\theta}_n^{\text{MD}}$ converges locally uniformly to the estimand $T(D)$ whenever the true D is close to C_θ. Finally the MD estimator also has the property of asymptotic normality. The estimator is appropriate to apply when a slight deviation from a given parametric family is anticipated.

3.2.5 Simulation study

Before we analyze our data, we want to be certain that our estimation methods estimate the copula dependence parameter well. We thus perform a simulation study. In the study we consider the bivariate case where we generate 500 samples from the copulas with sizes $n = 200$. The copulas that are tested are: Frank, Clayton, Gumble and Gaussian. The dependence parameter is estimated on each sample with our three different methods; Pseudo maximum likelihood and the Minimum distance estimator (MD) with the Kolmogorv-Smirnov and Cramér-von Mises distances. We do the simulation on three different true parameter values of the copulas. Thereafter the estimates are evaluated by computing the Monte Carlo based bias and mean squared error (MSE). The results of the simulation study are presented in Table 1.

From the study we see that the estimated bias and MSE for all the methods are quite small and we can thus feel comfortable using them on our data. We also notice that the PML performs significantly better than both the MD estimators as it causes less bias and MSE. What is not presented in the table is that the code for PML computes much faster than the MD. We can thus conclude that PML the most efficient and optimal estimation method and will the one that will be mostly used for the upcoming analysis.

The results between KS and CvM statistics were quite even, although CvM performed slightly better overall. Another observation is that as the true dependence parameter increases, the estimation becomes more difficult and the bias and MSE increases with it.

Table 1: Simulation results measured in estimated bias and MSE of the estimation methods for different copula families.

Family	Real parameter value	Estimate method	$\widehat{bias}$	MSE
Gaussian	$\theta = -0.75$	PML	0.0045	0.0010
		MD(KS)	0.0305	0.0028
		MD(CvM)	0.0308	0.0026
	$\theta = 0$	PML	0.0039	0.0050
		MD(KS)	0.0397	0.0080
		MD(CvM)	0.0428	0.0077
	$\theta = 0.25$	PML	−0.0026	0.0042
		MD(KS)	0.0285	0.0066
		MD(CvM)	0.0314	0.0062
Clayton	$\theta = 1$	PML	0.0408	0.0315
		MD(KS)	0.1104	0.0654
		MD(CvM)	0.1039	0.0552
	$\theta = 4$	PML	−0.0349	0.2373
		MD(KS)	0.4034	0.6957
		MD(CvM)	0.4456	0.6408
	$\theta = 8$	PML	−0.1776	0.6768
		MD(KS)	0.8093	1.8439
		MD(CvM)	1.0744	2.0701
Frank	$\theta = 1$	PML	0.0219	0.1831
		MD(KS)	0.1844	0.2698
		MD(CvM)	0.1873	0.2435
	$\theta = 4$	PML	0.0168	0.2941
		MD(KS)	0.2504	0.3815
		MD(CvM)	0.2737	0.3625
	$\theta = 8$	PML	0.0304	0.5466
		MD(KS)	0.4993	0.9364
		MD(CvM)	0.6134	0.9905
Gumbel	$\theta = 1.5$	PML	0.0231	0.0099
		MD(KS)	0.0535	0.0157
		MD(CvM)	0.0604	0.0151
	$\theta = 4$	PML	−0.0049	0.1069
		MD(KS)	0.2473	0.2762
		MD(CvM)	0.3601	0.3000
	$\theta = 8$	PML	−0.1543	0.4663
		MD(KS)	0.7504	1.5585
		MD(CvM)	1.2792	2.2453

3.3 Goodness-of-fit test for copulas

Continuing from the problem of estimating the dependence parameter θ of a copula model, there is the issue of testing the validity of the null hypothesis:

$$H_0 : C \in \mathcal{C}_0,$$

for some specific parametric copula family $\mathcal{C}_0 = \{C_\theta : \theta \in \Theta\}$. In other words, we want to test that the dependence structure of the copula C is well represented by a specific parametric family $\mathcal{C}_0$. Since the underlying copula C is invariant by continuous and strictly increasing transformations of its components (Theorem 3), Genest and Remillard [9] proposed a semi-parametric bootstrap test based on a maximally invariant statistics with respect to a ranking transformation. The pseudo observations that were described earlier will constitute the statistic on the test where large values of statistic will lead to rejection of the estimated model. The statistic in mind is the Cramér-von Mises statistic which involves the empirical copula

$$\sum_{t=1}^{n} \left(C_{\theta_n}(U_t, V_t) - C_n(U_t, V_t)\right)^2 .$$

Let $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ be independent copies of a random vector (X, Y) and $(U_1, V_1), (U_2, V_2), ..., (U_n, V_n)$ be their respective pseudo observations. The procedure consists of the following steps:

1. Compute the empirical copula C_n and estimate θ with $\theta_n = T_n[(U_1, V_1), ..., (U_n, V_n)]$ where T_n is the functional for one of the estimation methods from Section 3.2.
2. Compute
$$S_n = \sum_{t=1}^{n} \{C_{\theta_n}(U_t, V_t) - C_n(U_t, V_t)\}^2.$$
3. For a large integer N, repeat the following bootstrap steps for every $k \in \{1, ..., N\}$
   (a) Generate a random sample $(X^*_{1,k}, Y^*_{1,k}), ..., (X^*_{n,k}, Y^*_{n,k})$ from distribution C_{θ_n} and compute their pseudo vectors $(U^*_{1,k}, V^*_{1,k}), ..., (U^*_{n,k}, V^*_{n,k})$.
   (b) Estimate the pseudo bootstrap samples empirical copula
   $$C^*_{n,k} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(U^*_{1,k} \leq u, V^*_{1,k} \leq v).$$
   and the dependence parameter θ by $\theta^*_{n,k} = T_n[(U^*_{1,k}, V^*_{1,k}), ..., (U^*_{n,k}, V^*_{n,k})]$.
   (c) Let
   $$S^*_{n,k} = \sum_{i=1}^{n} (C^*_{n,k}(U, V) - C_{\theta^*_{n,k}}(U, V))^2.$$

The p-value is then given by $\sum_{k=1}^{N} \mathbf{1}(S^*_{n,k} > S_n)/N$. If the distance between the observed distribution and fitted distribution is sufficiently short, the p-value will be high, suggesting that the null hypothesis can not be rejected.

The test is consistent, meaning that if $C \notin C_0$, then the null hypothesis is rejected with probability 1 as $n \to \infty$ [8].

3.4 Data description

Nord Pool Spot [14] provided data of the spot electricity prices as well as the water levels in the Swedish hydro power plants from 1996 to 2012. The electricity prices are measured hourly and weekly in SEK/ MWh, while the water levels are measured once a week in GWh. SMHI [13] provided data of outside temperatures from different weather stations throughout Sweden. The data consists of hourly measures from 1996-2012.

A note on notation:

The time series of electricity prices, water levels and outdoor temperature will be denoted E, W and T respectively. If the electricity prices have been deaseasonlized with a period of d, then it will have notation $E_{(d)}$. Data sets that pairs two time series are for instance denoted as $[E_{(d)}, W]_{(ot)}$, where the observation times (ot) can be either hourly (h), or weekly (w).

With the data provided, the sets of interest to investigate are the following:

Electricity prices and Water levels, [E,W]

Weekly observations of the Stockholm electricity prices and the water levels in the Swedish hydro reservoirs with the time periods 1996-2012 and 2005-2012. The electricity prices are tested with and without estimated trend and seasonal components with a period of 52 weeks. There is a total of 4 data sets.

1. $[E, W]_{(w)}(1996 - 2012)$.
2. $[E, W]_{(w)}(2005 - 2012)$.
3. $[E_{(52)}, W]_{(w)}(1996 - 2012)$.
4. $[E_{(52)}, W]_{(w)}(2005 - 2012)$.

Data sets 1 and 3 are of size 887, and 417 for sets 2 and 4.

Electricity prices and Outdoor temperature [E,T]

We will investigate data sets of whole 2012 but also for specific months, December and July. With and without seasonality with daily and weekly period i.e. 24 and 168 hours. Here the temperatures are from Gallivare (Northern Sweden, close to one of the biggest hydro power plants). We also investigate the weekly measures of electricity prices and temperature from 1996-2012 and 2005-2012. A total of 13 datasets.

1. $[E, T]_{(h)}(2012)$.
2. $[E_{(24)}, T]_{(h)}(2012)$.
3. $[E_{(168)}, T]_{(h)}(2012)$.
4. $[E, T]_{(h)}(\text{December}, 2012)$.
5. $[E_{(24)}, T]_{(h)}(\text{December}, 2012)$.
6. $[E_{(168)}, T]_{(h)}(\text{December}, 2012)$.
7. $[E, T]_{(h)}(\text{July}, 2012)$.
8. $[E_{(24)}, T]_{(h)}(\text{July}, 2012)$.
9. $[E_{(168)}, T]_{(h)}(\text{July}, 2012)$.

10. $[E, T]_{(w)}(1996 - 2012)$.

11. $[E, T]_{(w)}(2005 - 2012)$.

12. $[E_{(52)}, T]_{(w)}(1996 - 2012)$.

13. $[E_{(52)}, T]_{(w)}(2005 - 2012)$.

Data sets 1-3 are of size 8783, sets 4-9 are of size 745. Data sets 10 and 12 are of size 887 and 11 and 13 have size 417.

Electricity prices, Water levels and Outdoor temperature [E,W,T]

We will also try to fit a trivariate model by investigating the data of the weekly observations of electricity prices, water levels and temperature together.

1. $[E, W, T]_{(w)}(1996 - 2012)$.

2. $[E, W, T]_{(w)}(2005 - 2012)$.

3. $[E_{(52)}, W, T]_{(w)}(1996 - 2012)$.

4. $[E_{(52)}, W, T]_{(w)}(2005 - 2012)$.

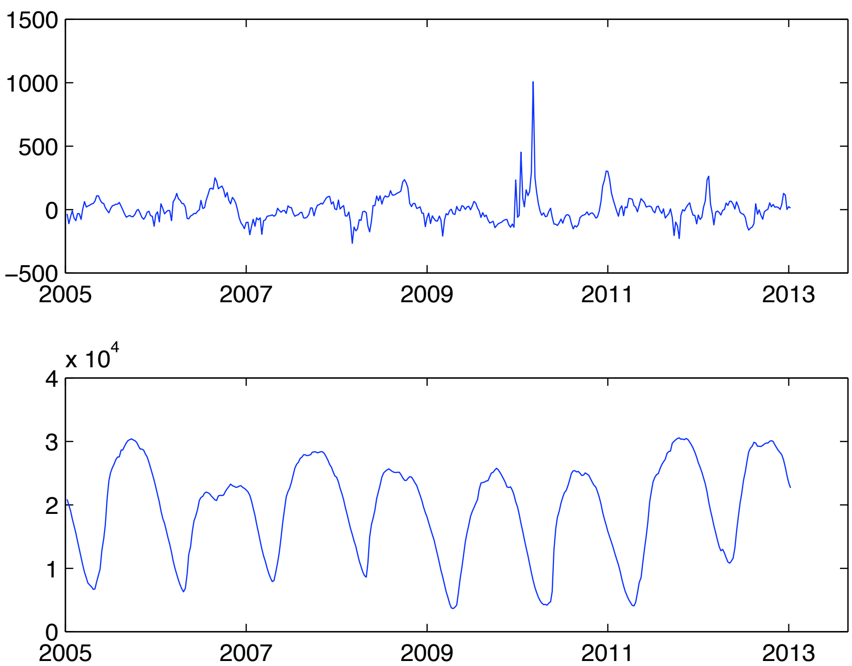

Figure 12: Plot of the $[E_{(52)}, W]_{(w)}(2005-2012)$ data set. Deseasonlized electricity prices (top) and water levels (bottom)

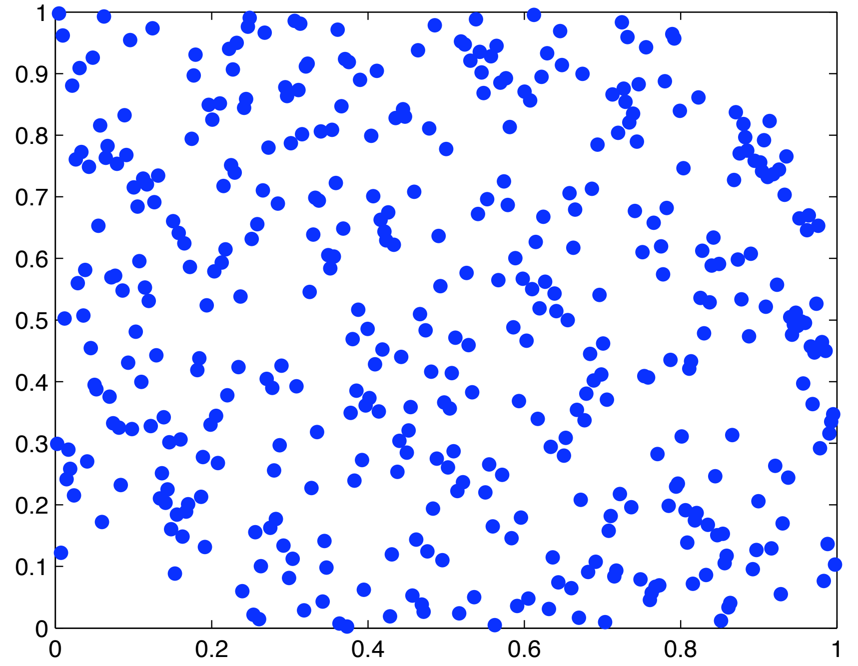

Figure 13: The pairs of pseudo observations of $[E_{(52)}, W]_{(w)}(2005 - 2012)$ gives us a hint of their dependence.

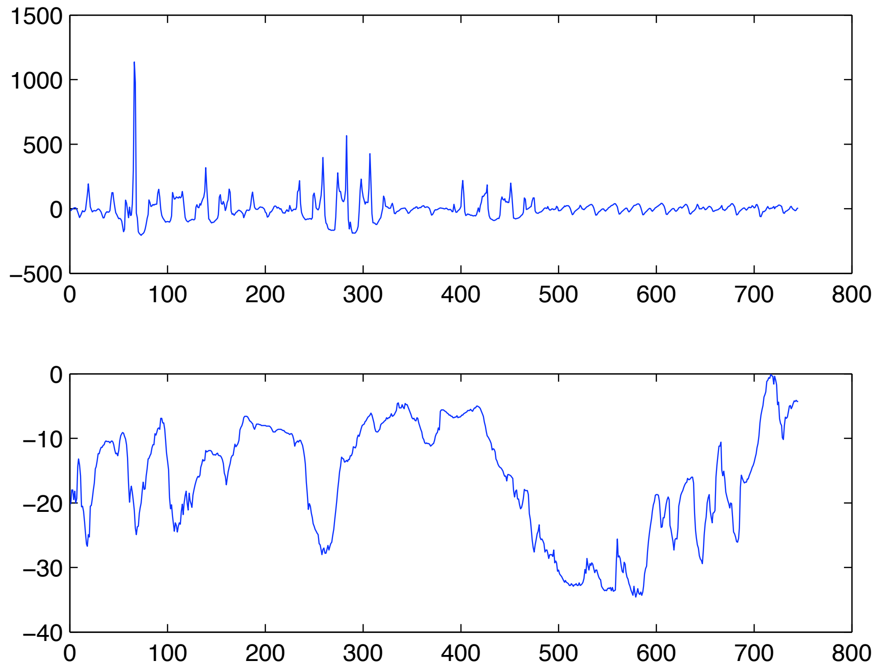

Figure 14: Plot of the $[E_{(24)}, T]_{(h)}(\text{December}, 2012)$, Deseasonalized electricity prices (top) and temperature (bottom)

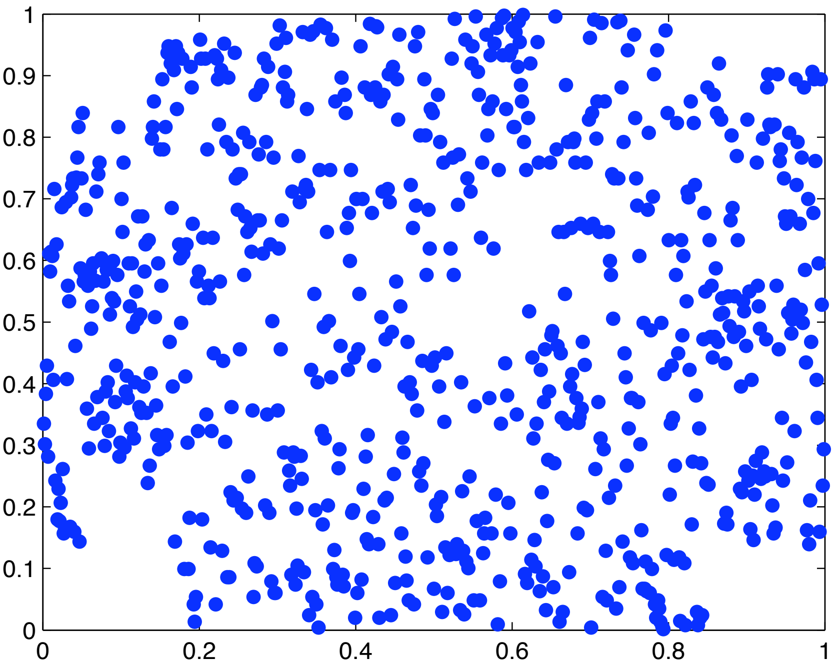

Figure 15: The pseudo observations of $[E_{(24)}, T]_{(h)}(\text{December}, 2012)$ showing the dependence between electricity prices and temperature.

4 Results

We will now present the results from our analysis of the data sets in a collection of tables and figures. For each data set, the dependence parameter was estimated for the copula families Frank, Clayton, Gumbel, Gaussian and Student's t. The method of estimation used was the pseudo-maximum likelihood. Thereafter the Cramér-von Mises statistic was calculated and the goodness-of-fit test was performed to evaluate the best model. The CvM statistics in the tables are displayed in their square roots, for a more comprehensible view. The goodness-of-fit tests were executed with 1000 bootstrap samples with the size of the original data set.

4.1 Copula modeling with electricity prices and water levels.

Table 2: Estimate of dependence parameter, $\hat{\theta}$ for $[E, W]$ data sets. The estimates clearly suggests a negative dependence. Moreover the really high estimated degrees of freedom for the Student t copula indicates that the Gaussian copula is sufficient enough.

Data set	Frank	Clayton	Gumbel	Gaussian	Student's t
1. $[E, W]_{(w)}(1996 - 2012)$	-0.877	0.000	1.000	-0.189	$-0.191, \hat{\nu} = 3.876 \cdot 10^6$
2. $[E, W]_{(w)}(2005 - 2012)$	-1.125	0.000	1.000	-0.221	$-0.226, \hat{\nu} = 1.402 \cdot 10^7$
3. $[E_{(52)}, W]_{(w)}(1996 - 2012)$	-0.612	0.000	1.000	-0.114	$-0.116, \hat{\nu} = 1.402 \cdot 10^7$
4. $[E_{(52)}, W]_{(w)}(2005 - 2012)$	-0.577	0.000	1.000	-0.122	$-0.125, \hat{\nu} = 1.577 \cdot 10^7$

Table 3: Cramér-von Mises statistic for the $[E, W]$ data sets. We notice that by deseasonlizing the prices and by concentrating on shorter time periods, the statistics decreases. Since the Frank, Gaussian and Student t copulas can capture negative dependence, they have lower statistics and make a better fit than Clayton and Gumbel.

Data set	Frank	Clayton	Gumbel	Gaussian	Student's t
1. $[E, W]_{(w)}(1996 - 2012)$	0.3992	0.4880	0.4880	0.4201	0.4234
2. $[E, W]_{(w)}(2005 - 2012)$	0.3611	0.5046	0.5046	0.3619	0.3637
3. $[E_{(52)}, W]_{(w)}(1996 - 2012)$	0.1970	0.3321	0.3321	0.1975	0.1985
4. $[E_{(52)}, W]_{(w)}(2005 - 2012)$	0.1449	0.2098	0.2098	0.1539	0.1565

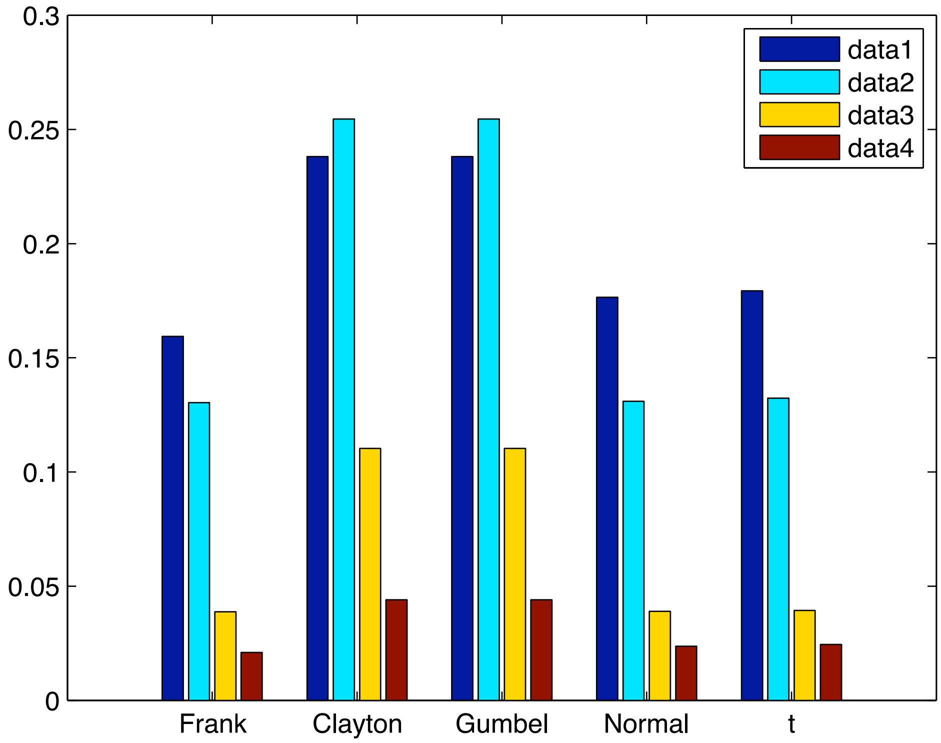

Figure 16: A visual representation the CvM statistic.

Table 4: p-value from goodness of fit test for $[E, W]$ data sets. Data set 4 with the deaseaonalized prices and shorter time period clearly have insignificant p-values, which suggests that the statistics are sufficiently low and that we can not reject the null hypothesis that the dependence can be modeled with the Frank, Gaussian and Student's t copulas.

Data set	Frank	Clayton	Gumbel	Gaussian	Student's t
1. $[E, W]_{(w)}(1996 - 2012)$	0	0	0	0	0
2. $[E, W]_{(w)}(2005 - 2012)$	0	0	0	0	0
3. $[E_{(52)}, W]_{(w)}(1996 - 2012)$	0.0080	0.0020	0.0040	0.0180	0.0160
4. $[E_{(52)}, W]_{(w)}(2005 - 2012)$	0.3270	0.0660	0.0690	0.2630	0.2270

4.2 Copula modeling with electricity prices and temperatures

Table 5: Estimates of dependence parameters, $\hat{\theta}$ of the $[E, T]$ data.

Data set	Frank	Clayton	Gumbel	Gaussian	Student's t
1. $[E, T]_{(h)}(2012)$	-4.746	0	1	-0.600	$-0.613, \hat{\nu} = 10$
2. $[E_{(24)}, T]_{(h)}(2012)$	-0.163	0.122	1	-0.009	$-0.024, \hat{\nu} = 5$
3. $[E_{(168)}, T]_{(h)}(2012)$	0.232	0.173	1.001	0.039	$0.042, \hat{\nu} = 5$
4. $[E, T]_{(h)}(\text{Dec}, 2012)$	0.108	0	1	-0.052	$-0.038, \hat{\nu} = 27$
5. $[E_{(24)}, T]_{(h)}(\text{Dec}, 2012)$	-0.161	0	1	-0.027	$-0.028, \hat{\nu} = 4.66 \cdot 10^6$
6. $[E_{(168)}, T]_{(h)}(\text{Dec}, 2012)$	-1.331	0	1	-0.196	$-0.198, \hat{\nu} = 1.60 \cdot 10^7$
7. $[E, T]_{(h)}(\text{July}, 2012)$	1.462	0.338	1.125	0.264	$0.267, \hat{\nu} = 1.41 \cdot 10^7$
8. $[E_{(24)}, T]_{(h)}(\text{July}, 2012)$	0.241	0.100	1	0.044	$0.043, \hat{\nu} = 16$
9. $[E_{(168)}, T]_{(h)}(\text{July}, 2012)$	-0.584	0	1	-0.079	$-0.080, \hat{\nu} = 3.759 \cdot 10^6$
10. $[E, T]_{(w)}(1996 - 2012)$	-0.907	0	1	-0.200	$-0.169, \hat{\nu} = 7$
11. $[E, T]_{(w)}(2005 - 2012)$	-1.182	0	1	-0.227	$-0.201, \hat{\nu} = 4$
12. $[E_{(52)}, T]_{(w)}(1996 - 2012)$	0.184	0	1	-0.036	$0.0232, \hat{\nu} = 5$
13. $[E_{(52)}, T]_{(w)}(2005 - 2012)$	-0.015	0	1	-0.053	$-0.016, \hat{\nu} = 4$

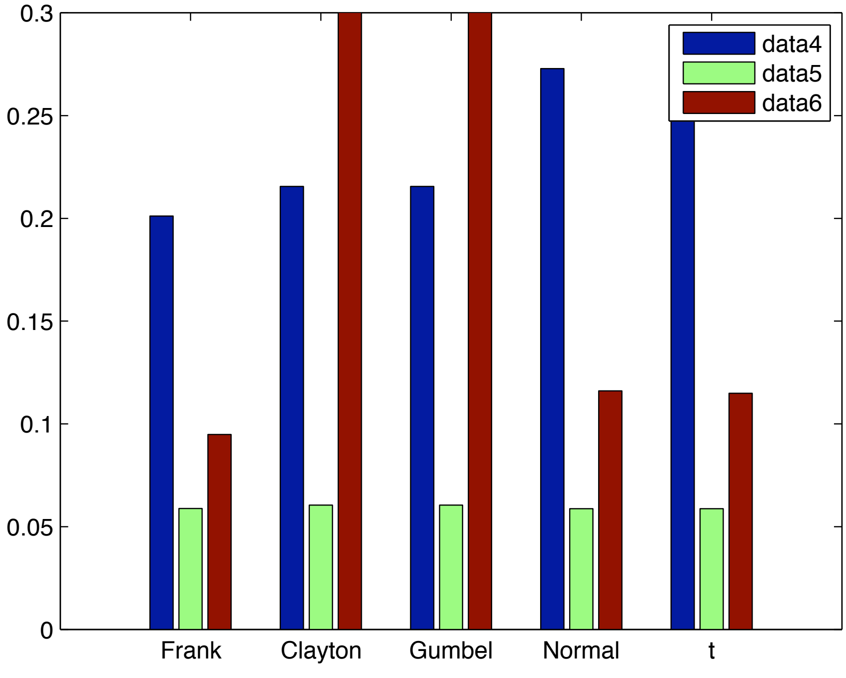

Figure 17: A visual representation of a few of the CvM statistics.

Table 6: Cramér-von Mises statistic for $[E, T]$ data. The statistics are overall pretty high and the fit of the models are not that great. The models improve when we only look at the specific months instead of the whole year and the 24 hour deseasoning is slightly better than the 168 hour deseasoning.

Data set	Frank	Clayton	Gumbel	Gaussian	Student's t
1. $[E, T]_{(h)}(2012)$	0.8526	6.3459	6.3459	0.7596	0.6686
2. $[E_{(24)}, T]_{(h)}(2012)$	0.6957	1.2252	0.7609	0.7292	0.6535
3. $[E_{(168)}, T]_{(h)}(2012)$	0.7537	0.8752	0.8420	0.7576	0.7099
4. $[E, T]_{(h)}(\text{December}, 2012)$	0.4486	0.4643	0.4643	0.5224	0.5055
5. $[E_{(24)}, T]_{(h)}(\text{December}, 2012)$	0.2425	0.2460	0.2460	0.2423	0.2423
6. $[E_{(168)}, T]_{(h)}(\text{December}, 2012)$	0.3081	0.6111	0.6111	0.3407	0.3390
7. $[E, T]_{(h)}(\text{July}, 2012)$	0.3330	0.3202	0.4445	0.3220	0.3214
8. $[E_{(24)}, T]_{(h)}(\text{July}, 2012)$	0.2661	0.2640	0.2864	0.2640	0.2617
9. $[E_{(168)}, T]_{(h)}(\text{July}, 2012)$	0.2610	0.3501	0.3501	0.2632	0.2629
10. $[E, T]_{(w)}(1996 - 2012)$	0.2848	0.4391	0.4391	0.3203	0.2615
11. $[E, T]_{(w)}(2005 - 2012)$	0.2573	0.4206	0.4206	0.2627	0.2293
12. $[E_{(52)}, T]_{(w)}(1996 - 2012)$	0.2358	0.2915	0.2661	0.3659	0.2394
13. $[E_{(52)}, T]_{(w)}(2005 - 2012)$	0.1871	0.1852	0.1806	0.2390	0.1856

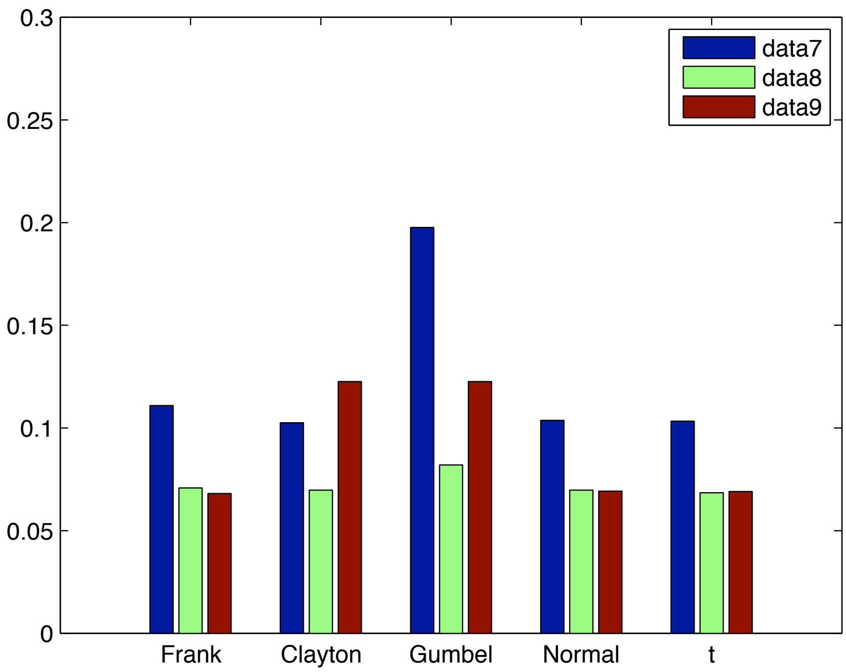

Figure 18: A visual representation of a few of the CvM statistics.

Table 7: p-values from goodness of fit results for $[E, T]$ data. Most of the values are 0 or close to 0 so it might be better to look at the statistics instead, to find the best models.

Data set	Frank	Clayton	Gumbel	Gaussian	Student's t
1. $[E, T]_{(h)}(2012)$	0	0	0	0	0
2. $[E_{(24)}, T]_{(h)}(2012)$	0	0	0	0	0
3. $[E_{(168)}, T]_{(h)}(2012)$	0	0	0	0	0
4. $[E, T]_{(h)}(\text{December}, 2012)$	0	0	0	0	0
5. $[E_{(24)}, T]_{(h)}(\text{December}, 2012)$	0	0.014	0.017	0.001	0
6. $[E_{(168)}, T]_{(h)}(\text{December}, 2012)$	0	0	0	0	0
7. $[E, T]_{(h)}(\text{July}, 2012)$	0	0	0	0	0
8. $[E_{(24)}, T]_{(h)}(\text{July}, 2012)$	0	0	0.004	0	0
9. $[E_{(168)}, T]_{(h)}(\text{July}, 2012)$	0	0.001	0.001	0	0
10. $[E, T]_{(w)}(1996 - 2012)$	0	0.001	0	0	0.068
11. $[E, T]_{(w)}(2005 - 2012)$	0	0	0	0	0
12. $[E_{(52)}, T]_{(w)}(1996 - 2012)$	0	0.003	0.004	0	0
13. $[E_{(52)}, T]_{(w)}(2005 - 2012)$	0.016	0.171	0.154	0.002	0.036

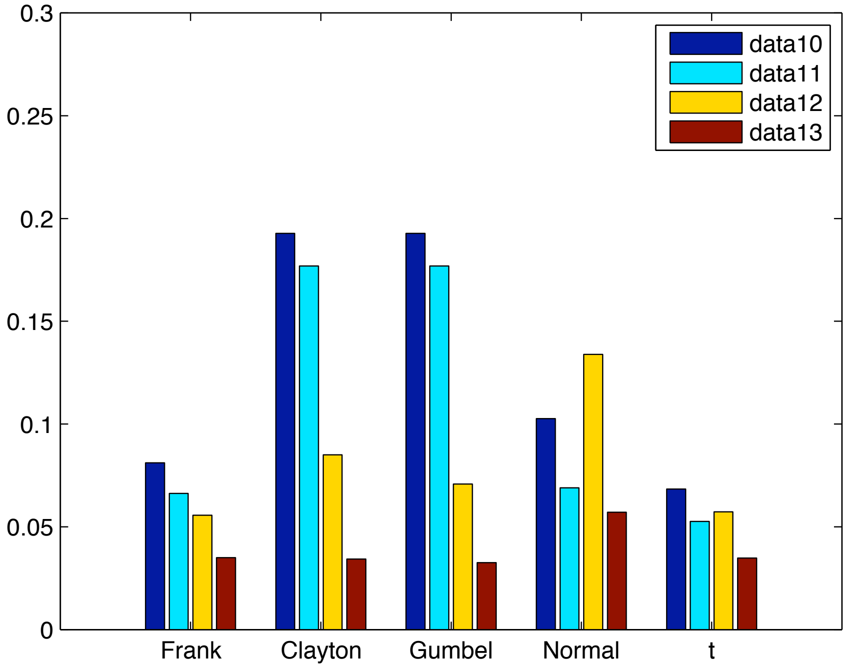

Figure 19: A visual representation of a few of the CvM statistics from.

4.3 Mixed copulas

Mixed copulas with two copulas were fitted on the data sets with the hope of finding a more suitable copula model. Recall the mixed copula

$$C_{\text{mix}}(u, v, \beta, \theta_1, \theta_2) = \beta C_1(u, v, \theta_1) + (1 - \beta)C_2(u, v, \theta_2).$$

The parameters β, θ_1 and θ_2 were estimated on mixed copulas combining the pairs of the families Frank, Clayton, Gumbel and Gaussian. The estimation procedure is harder than normally since it requires an estimation of three parameters. For most of the data sets, the estimated shape parameter β was either 0 or 1, implying that a mixed copula was not more appropriate than a single copula. Table 8 presents the data sets where a good estimated mixed copulas were found and their Cramér-von Mises statistic.

Table 8: Mixed copula results on the $[E, T]$ data. When compared to Table 6 we see that the Cramér-von Mises statistics in this table is slightly lower than, showing an improvement.

Data set	Copula 1	Copula 2	$\hat{\beta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	CvM stat.
1. $[E, T]_{(h)}(2012)$	Gaussian	Gumbel	0.861	-0.704	1.000	0.6364
2. $[E_{(24)}, T]_{(h)}(2012)$	Frank	Clayton	0.666	-1.745	1.000	0.4837
3. $[E_{(168)}, T]_{(h)}(2012)$	Frank	Clayton	0.592	-1.599	1.000	0.4889
7. $[E, T]_{(h)}(\text{July}, 2012)$	Frank	Clayton	0.773	1.070	1.000	0.3066
8. $[E_{(24)}, T]_{(h)}(\text{July}, 2012)$	Clayton	Gumbel	0.774	0.129	1.000	0.2627

4.4 Trivariate data

Estimating three dimensional copulas are a lot more difficult than in two dimensions due to the more complex structures. The problem especially arises in the estimation procedure where numerical issues occurs for the pseudo maximum likelihood estimation. When an estimate could not be found on a data set, due to convergence problems, the minimum distance method was used instead. Table 9 presents the statistics for the estimated copulas in three dimensions.

Table 9: Cramér-von Mises statistics for the trivariate data $[E, W, T]$. The statistics are surprisingly low, and are on the same level as the two dimensional copulas showed.

Data	Frank	Clayton	Gumbel	Gaussian	Student's t
$[E, W, T]_{(w)}(1996 - 2012)$	0.5564	0.5546	0.6181	0.5263	0.5255
$[E, W, T]_{(w)}(2005 - 2012)$	0.6051	0.4044	0.5381	0.3247	0.3242
$[E_{(52)}, W, T]_{(w)}(1996 - 2012)$	0.4471	0.4630	0.4592	0.4539	0.4529
$[E_{(52)}, W, T]_{(w)}(2005 - 2012)$	0.2910	0.2832	0.2919	0.3172	0.3159

4.5 Adaption to positive dependence

To solve the issue of when negative dependence is implied to the data, we transformed the electricity prices by changing the sign of the values. The transformation will hopefully benefit the Clayton and Gumbel copulas and make them more relevant, since they only account for positive dependence. We present the results of the transformation of the $[E, W]$ data in tables 10 and 11 below, and notice that the Cramér-von Mises statistics improves for the Clayton and Gumbel copulas. Compare to tables 2 and 3 to see the difference. Similar results were seen for the transformed $[E, T]$ data sets although there were no significant decrease of any data set.

Table 10: Estimate of dependence parameter, $\hat{\theta}$ for $[E, W]$ data sets, where the sign has been changed for the electricity price observations in order to account for positive values of the dependence parameter.

Data set	Frank	Clayton	Gumbel	Gaussian	Student's t
1. $[\text{-}E, W]_{(w)}(1996-2012)$	0.876	0.094	1.131	0.189	$0.191, \hat{\nu} = 3.667 \cdot 10^6$
2. $[\text{-}E, W]_{(w)}(2005-2012)$	1.125	0.173	1.129	0.221	$0.226, \hat{\nu} = 1.491 \cdot 10^7$
3. $[\text{-}E_{(52)}, W]_{(w)}(1996-2012)$	0.612	0.073	1.041	0.114	$0.116, \hat{\nu} = 1.677 \cdot 10^7$
4. $[\text{-}E_{(52)}, W]_{(w)}(2005-2012)$	0.577	0.083	1.057	0.122	$0.125, \hat{\nu} = 1.584 \cdot 10^7$

Table 11: Cramér-von Mises statistic for the $[E, W]$ data sets. We notice that the transform of the data resulted in a decrease of the Clayton and Gumbel statistics. Since the Frank, Gaussian and Student's t are symmetric, their respective statistics have not been changed.

Data set	Frank	Clayton	Gumbel	Gaussian	Student's t
1. $[\text{-}E, W]_{(w)}(1996-2012)$	0.4026	0.4450	0.3571	0.4123	0.4148
2. $[\text{-}E, W]_{(w)}(2005-2012)$	0.3751	0.4373	0.3590	0.3663	0.3663
3. $[\text{-}E_{(52)}, W]_{(w)}(1996-2012)$	0.2032	0.2642	0.2542	0.1982	0.1982
4. $[\text{-}E_{(52)}, W]_{(w)}(2005-2012)$	0.1575	0.1876	0.1682	0.1487	0.1490

5 Conclusion

Copula models have been estimated and evaluated on time series observations from the electricity market. Looking at the Cramér-von Mises statistics in Tables 3 and 6 we note that the removal of the trend and seasonal components, and by adapting the data for positive dependence, is essential to find a good copula model, as the procedures significantly lowers the statistics. Moreover we noticed that it was easier to fit a copula when we were concentrating on shorter time periods. Instead of fitting the copula over a whole year, the copula model should instead reflect the month or season in focus. The same thing can be said when modeling over several years as a longer time period has to deal with a change behavior of the electricity price from new regulations, such as the emission regulation introduced in year 2005. It is therefore suggested that different models should be used on different seasons.

Of all the data sets, there was one in particular that fit very well, the $[E_{(52)}, W]_{(w)}(2005-2012)$ data set. The data set had high, insignificant p-values for the Frank, Gaussian and Student's t copulas with p-values of $0.327, 0.263$ and 0.227 respectively, which indicates sufficiently short statistics for the symmetric copulas. We can thus not reject the null hypothesis that the dependence between weekly observations of electricity prices and water levels can be modeled with these copulas. There were no models with an exciting p-value for the electricity price and temperature data, but some of them had a noteworthy low CvM statistics, specifically the sets of the deseasonalized months. The $[E_{(52)}, T]_{(w)}(2005-2012)$ data set had high p-values and low test statistics, less than 0.2 However the estimated models were the corresponding of the product copula which is unexciting.

The modeling of the more general mixed copulas also improved the statistics for some of the data sets. A mixed copula could however not be found on most of the data sets, as they are a lot more difficult to estimate than a single copula. Instead of our pseudo maximum likelihood and minimum distance estimates, other more advanced estimation methods could be applied to mixed copulas to find better fitted models. Estimating three dimensional copulas was also hard due to the complex structure of their structures and none of our models did fit quite well. Although they don't provide a good fit, our three dimensional copula families are convenient to work with and are more plausible than other simpler models that assume independence.

It is of course possible to improve all of our methods and procedures to tune our models, but our result give a good indication on how well copulas fit the data.

A References

[1] Robin Axelsson. *Stochastic Modeling of Electricity prices using ARMA and ARMAX Models*. Master's thesis, Department of Mathematical Sciences, Chalmers University of Technology, University of Gothenburg, 2012.

[2] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer, New York, NY, second edition, 1991.

[3] Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula Methods in Finance*. Wiley, Chichester, England, 2004.

[4] Svensk Energi. `http://www.svenskenergi.se/`.

[5] Jean-David Fermanian, Dragan Radulovic, and Marten Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860, 2004.

[6] C. Genest, K. Ghoudi, and L.-P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, September 1995.

[7] Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12:347–368, 2007.

[8] Christian Genest and Bruno Rémillard. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré*, 44(6):1096–1127, 2008.

[9] Christian Genest, Bruno Rémillard, and David Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213, April 2009.

[10] Svenska Kraftnät. `http://svk.se/Energimarknaden/`.

[11] Svenska Kraftnät. The swedish electricity market and the role of svenska kraftnät. `http://www.svk.se/global/13_energy_market/pdf/elmarknaden_eng_2007_webb.pdf`, 2007.

[12] Roger B. Nelsen. *An introduction to copulas*. Springer, New York, NY, second edition, 2006.

[13] Sveriges meteorologiska och hydrologiska institut SMHI. `http://www.smhi.se/`.

[14] Nord Pool Spot. `http://www.nordpoolspot.com/`.

[15] Pravin K. Trivedi and David M. Zimmer. Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, 1(1):1–111, 2005.

[16] Hideatsu Tsukahara. Semiparametric estimation in copula models. *The Canadian Journal of Statistics*, 33(3):357–375, September 2005.

[17] Larry Wasserman. *All of statistics, A Concise Course in Statistical Inference*. Springer, 2004.