GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Master Thesis

Multivariate modelling of the dependency structure between article sales of a sportswear manufacturer

Author

Petros Christanas

from Nuremberg

Matriculation Number

11604278

Applied Statistics M.Sc.

Chair of Statistics and Econometrics

Supervisors

Prof. Dr. Thomas Kneib

Dipl.-Vw. Quant. Fabian H. C. Raters

Submitted June 25, 2020

Processing time of 20 weeks

Confidentiality Clause

Statutory Declaration

Acknowledgments

Contents

1 Introduction

Nowadays, online shopping is gradually becoming people's favourite purchasing standard. As designer and fashion brands adjust to this new way of shopping, they are not only promoting their products via third party providers, but also have their own e-Commerce websites. Likewise, *adidas* has grown its eCom channel tremendously over the past few years and has gained a large pool of casual and regular customers.

The aim of some use cases of the adidas Advanced Analytics Hub is to generate sales[1] forecasts for individual articles, usually on a weekly or monthly level. This is not always trivial since industrial big data are quite noisy, e.g. different types of campaigns and promotions influence the demand quantity dramatically over time. Another latent effect is sales cannibalization between newer and older articles or articles of similar traits. To identify this effect in a quantifiable way, the purpose of this thesis is to capture a dependence structure between article demand quantities over time, applied to transactional eCom data.

In the remaining of Chapter 1, a brief overview of the sports brand adidas is given in Section 1.1 and in Section 1.2 we will have a first look into our data sources and a data dictionary. Chapter 2 introduces some notions on relevant statistical methodology. In Chapter 3, we will have a closer look into the copula framework which comprises the major ingredient of the modelling part of this thesis. In Chapter 4 we perform some exploratory data analysis to deep dive into the behaviour of article sales and to stepwise investigate some hierarchical properties of our data. Chapter 5 analyses the results of modelling copulas to the data and we will examine some diagnostics using different approaches. A final conclusion will be summarised in Chapter 6, where we point out the main findings of this thesis.

1.1 adidas

After World War II, the *"Dassler Brothers Shoe Factory"* (German: *"Gebrüder Dassler Schuhfabrik"*), which was led by *Adolf Dassler* (aka *Adi Dassler*) and his brother Rudolph, was dissolved. The brothers split up and formed their own firms. As a result, the sports shoe factory *"Adi Dassler adidas Sportschuhfabrik"* was founded on August 18th 1949

[1] By sales we actually mean sale quantities in units throughout this thesis

by Adolf Dassler in Herzogenaurach, a small town in Germany [adidas-group.com].

(a) adidas Performance (b) adidas Originals

Figure 1.1: Two of the adidas-group logos: Performance (left) & Originals (right) [adidas.com media-center]

Today, just over 70 years later, the sportswear designer and manufacturer is known as the *"adidas AG"* (short: *adidas*) and is one of the world's biggest sports and fashion brands. The global headquarters of are located in the birthplace Herzogenaurach and the company is employing over 59,000 people worldwide, with *Kasper Rørsted* leading the brand as CEO since October 1st 2016. In 2019, adidas produced over 1.1 billion sports and sports lifestyle products and is nowadays sponsoring a vast range of athletes, artists and organizations across the globe (e.g. the FIFA World Cup™).

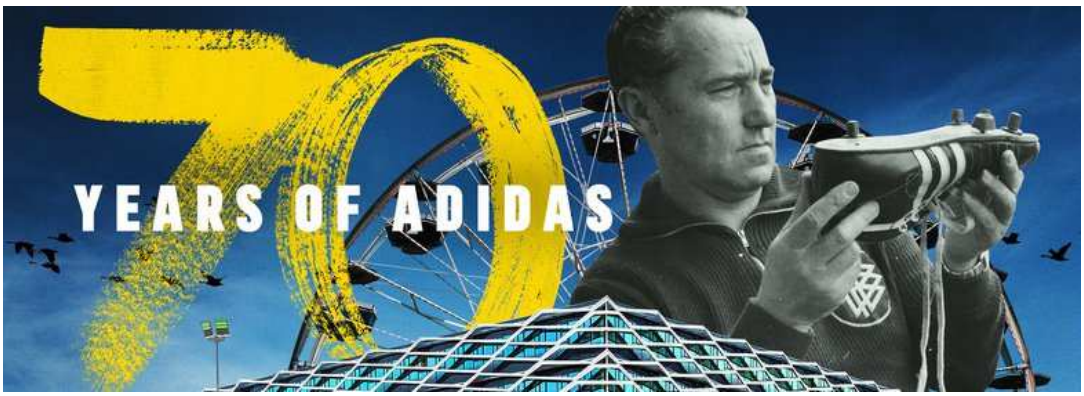


Figure 1.2: adidas celebrates its 70th anniversary and the opening of the ARENA building [adidas 70 years, 2019]

With the company's rapid growth, the data produced and acquired from social media, e-Commerce (eCom), transactions, demographics and other channels require suitable instruments and stuff to generate business relevant insights. As the need for data-driven decision making keeps increasing, *"Data and Analytics (DNA)"* became a sub-department of IT. Within DNA, the *"Data Science & AI Solutions"* team consisting of data scientists,

analysts, engineers and other roles is responsible for carrying out analytical tasks, for the most part in the context of various DNA use cases.

1.2 Data Sources

Throughout each season, transactional data are collected from online purchases of the sports brand's e-Commerce website. Specifically, weekly sales data for western European countries consisting of 109 observed weeks in total are provided. A short description is depicted in Table 1.1.

Column	Description	Values
week_id	Calender week of a specific year (YYYYWW)	Factors: *201648, ..., 201852*
article_number	Unique article identification number (article ID)	Factors: *10669, 10, ...*
min_date_of_week	Minimum date of the respective week; always a Monday (YYYY-MM-DD)	Dates: *2016-11-28, ..., 2018-12-24*
art_min_price	Minimal recorded price of the article	Non-negative (integer) value
month_id	Calender month of a specific year (YYYYMM)	Factors: *201612, ..., 201812*
season	Season of year (format: SSYY) (Spring-Summer [SS]: December - May) Fall-Winter [FW]: June - November)	Factors: *SS17, FW17,* *SS18, FW18, SS19*
bf_w	Weekly "Black Friday" promotion intensity of the article	Between *0* and *1*
ff_w	Weekly "Friends & Family" promotion intensity of the article	Between *0* and *1*
ot_w	Weekly article promotion intensity of "Other" type	Between *0* and *1*
gross_demand_quantity	Weekly amount of added articles to shopping cart	Non-negative (integer) value
base_price_locf	Retail price of the article without any discounts	Non-negative (integer) value
total_markdown_pct	Total markdown percentage of the article	Non-negative
day_of_month	Day of the month	Integers: *1 - 31*
month_of_year	Month of the year	Factors: *January, ..., December*
year	Year	Integers: *2016, 2017, 2018*
week_of_year	Week of the year	Integers: *1 - 52*

Table 1.1: Transactional raw data description from online purchases of western European countries

Due to legal regulations of the company, some columns had to undergo anonymization in order for the data to be released. To ensure data protection and confidentiality, numeric variables (with exception of time-indicating columns) were transformed. As a consequence for the analysis part, most integer values were converted to float numbers. This fact should be kept in mind by the reader, since the above table serves as a reminder and reference point for the data documentation.

Another peculiarity of this setup is to be considered, too. We will often refer to the variable *gross demand quantity* as *sales*, even though it is obviously not exactly the same. In the e-Commerce environment, there are several stages before the purchase is complete, e.g.

addition to cart, removal from cart, proceeding to checkout & even the return of bought articles. Targeting the articles added to cart, i.e. the (gross) demand quantity, provides the optimal data extraction for analytical purposes and is the closest to adequately model the dependence structure between net sales of articles.[2]

Besides the transactional data, attributes of the articles are provided and described in Table 1.2. Some attributes of special importance will be explained in more detail later on in Chapter 4.

Column	Description	Values (all Factors)
article_number	Unique article identification number (article ID)	*10669, 10, ...*
gender	Gender type of the article (Men, Women, Unisex)	*M, W, U*
age_group	Age group of the article (Adult, Infant, Junior, Kids)	*A, I, J, K*
key_category_descr	Key category of the article	*KC_1, ..., KC_15*
key_category_cluster_descr	Key category cluster of the article	*KCC_1, ..., KCC_9*
product_division_descr	Product division of the article	*Apparel, Footwear, Hardware*
product_group_descr	Product group of the article	*Bags, Balls, Footwear Accessories, Shoes, ...*
color	Consolidated color group of the article	*Beige, Black, Brown, Orange, Pink, Red, ...*
sports_category_descr	Sports category of the article	encoded: *SC_1, ..., SC_22*
sales_line_descr	Sales line of the article	encoded: *SL_1, ..., SL_379*
business_unit_descr	The article's Business Unit membership	encoded: *BU_1, ..., BU_18*
business_segment_descr	The article's Business Segment membership	encoded: *BS_1, ..., BS_49*
sub_brand_descr	Sub-brand of the article	encoded: *sub-brand_1, ..., sub-brand_4*
item_type	Item type of the article	encoded: *IT_1, ..., IT_171*
brand_element	Brand element of the article	encoded: *BE_1, ..., BE_131*
product_franchise_descr	Product franchise of the article	encoded: *franchise_1, ..., franchise_72*
product_line_descr	Product line of the article	encoded: *PL_1, ..., PL_105*
franchise_bin	Franchise indicator of the article	*Franchise, Non-Franchise*
category	Category of the article	encoded: *category_1, category_2*

Table 1.2: Article attribute data

Overall, these are the primary data sources and we will be working with data collected over two years, namely the years 2017 and 2018, while some transactions of late 2016 are attached marginally. In summary, after joining the transactional observations to the article attributes by the article ID, this translates to a dataset of over 587,000 instances and over 30 variables.

[2]Gross demand quantity will be our target as we follow the adidas norm

2 Statistical Theory & Methods

This chapter introduces various statistical methods used during the conduction of this thesis. Basic notations regarding mathematical foundations of statistics (such as linear algebra, probability theory, etc) are skipped. The chapter starts off with some fundamental notions on generalized linear models and arrives at a brief introduction to additive models

2.1 Generalized Linear Models

Generalized Linear Models (GLMs) are an extension of the classical *Linear Regression Model (LM)*

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \ldots, n$$

which in matrix notation can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where the response variable y_i can take values from several probability distributions (e.g. Poisson, Binomial, Gamma, ...), which are members of the exponential family [Fahrmeir et al., 2003]. The linear predictor

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i = \boldsymbol{x}_i'\boldsymbol{\beta} \tag{2.1}$$

is passed through a *response function h* (a one-to-one, twice differentiable transformation), such that

$$E(y_i) = h(\eta_i) \tag{2.2}$$

i.e. h ensures that the expected value of the response variable belongs to the appropriate value range. The inverse of the response function, i.e.

$$g = h^{-1}, \tag{2.3}$$

is called the *link function* and transforms the mean of the response's distribution to an unbounded continuous scale.

2.2 Mixed Effects Models

Linear Mixed Models (LMMs) are powerful tools when dealing with clustered data or data with a longitudinal structure (repeated measurements of individuals). As in the classical

LM, there are population-specific effects, namely the parameter vector of *fixed effects* $\boldsymbol{\beta}$, as well as the cluster- or individual-specific effects of such models called *random effects* [Fahrmeir et al., 2003]. Mathematically speaking, the linear predictor $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ is extended to

$$\eta_{ij} = \boldsymbol{x'}_{ij}\boldsymbol{\beta} + \boldsymbol{u}'_{ij}\boldsymbol{\gamma}_i \tag{2.4}$$

for individuals $i = 1, \ldots, m$ measured in a longitudinal setting at observed times $t_{i1} < \ldots < t_{ij} < \ldots < t_{i_{n_i}}$ or for subjects $j = 1, \ldots, n_i$ in cluster $i = 1, \ldots, m$.
In any case,

- $\boldsymbol{\beta}$ is the vector of fixed effects,
- $\boldsymbol{\gamma}_i$ is the vector of random effects,
- $\boldsymbol{x}'_{ij}$ is the vector of covariates and
- $\boldsymbol{u}'_{ij}$ is a subvector of $\boldsymbol{x}'_{ij}$.

$\boldsymbol{x}'_{ij} = (1, x_{ij1}, \ldots, x_{ijk})$ and $\boldsymbol{u}'_{ij} = (1, u_{ij1}, \ldots, u_{ijk})$ are therefore the design vectors and ε_{ij} are the error terms of the *measurement model*specified as

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{u}'_{ij}\boldsymbol{\gamma}_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right) \tag{2.5}$$

or in matrix notation

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{U}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i \tag{2.6}$$

with $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2\boldsymbol{I}_{n_i})$ and $\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \mathbf{Q})$ for individuals or clusters $i = 1, \ldots, m$ and positive definite matrix $\mathbf{Q}$. Note that $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_m$ are assumed to be mutually independent.

Similar to GLMs, *Generalized Linear Mixed Models (GLMMs)* relate the linear mixed predictor (Equation 2.4) to the conditional mean $\mu_{ij} = E(y_{ij}|\boldsymbol{\gamma}_i)$ via a suitable response function h, such that $\mu_{ij} = h(\eta_{ij})$ and thus the conditional density of y_{ij} belongs to the exponential family.

2.3 Additive Models

Additive Models expand models with just a linear predictor

$$\eta_i^{lin} = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$$

(such as the LM) to

$$y_i = \eta_i^{add} + \varepsilon_i, \tag{2.7}$$

where

$$\eta_i^{add} = f_1(z_{i1}) + \ldots + f_q(z_{iq}) + \eta_i^{lin}, \quad i = 1, \ldots, n. \tag{2.8}$$

The functions $f_1(z_1), \ldots, f_q(z_q)$ are non-linear univariate *smooth effects* of the *continuous* covariates $z_1, \ldots, z_q$ and are defined as

$$f_j(z_j) = \sum_{l=1}^{d_j} \gamma_{jl} B_l(z_j) \tag{2.9}$$

with $B_l(z_j)$ being *basis functions* for $j = 1, \ldots, q$ and d_j the number of basis functions for covariate z_j. The regression coefficients of the basis functions $B_l(z_j)$ are labeled as γ_{jl}. There is a wide variety of basis functions which can be used to flexibly model the data in a non-parametric manner. For more content on basis functions the reader can refer to Wood [2017] and Fahrmeir et al. [2003]. The basis functions evaluated at the observed covariate values are summarized in the design matrices $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_q$ and the additive model 2.7 can be written in matrix notation as

$$\boldsymbol{y} = \boldsymbol{Z}_1\boldsymbol{\gamma}_1 + \ldots + \boldsymbol{Z}_q\boldsymbol{\gamma}_q + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{2.10}$$

Accordingly, the vector of function values evaluated at the observed covariate values $z_{1j}, \ldots, z_{nj}$ is denoted by $\boldsymbol{f}_j = (f_j(z_{1j}), \ldots, f_j(z_{nj}))'$ and therefore $\boldsymbol{f}_j = \boldsymbol{Z}_j\boldsymbol{\gamma}_j$. To ensure identifiability of the additive model, the smooth functions $f_j(z_j)$ are centered around zero, such that

$$\sum_{i=1}^{n} f_1(z_{i1}) = \ldots = \sum_{i=1}^{n} f_q(z_{iq}) = 0.$$

A convenient trait of additive models is that they also support the incorporation of random effects. Random coefficient terms can straightforwardly be added to the model. Analogously to Section 2.2, data are considered to be measured in a longitudinal setting with individuals $i = 1, \ldots, m$ observed at times $t_{i1} < \ldots < t_{ij} < \ldots < t_{i_{n_i}}$ or clustered data with subjects $j = 1, \ldots, n_i$ in clusters $i = 1, \ldots, m$. Without loss of generality (w.l.o.g.),[3] we can simply add to Equation 2.10 the terms $\boldsymbol{Z}_0\boldsymbol{\gamma}_0$ and $\boldsymbol{Z}_1\boldsymbol{\gamma}_1$ representing the design matrices and coefficients of the random intercepts and random slopes respectively. Explicitly, the coefficients are formulated as $\gamma_0 = (\boldsymbol{\gamma}_{01}, \ldots, \boldsymbol{\gamma}_{0i}, \ldots, \boldsymbol{\gamma}_{0m})'$ and

[3] for the indexes

$\gamma_1 = (\gamma_{11}, \ldots, \gamma_{1i}, \ldots, \gamma_{1m})'$, whereas the design matrices are expressed as

$$\boldsymbol{Z}_0 = \begin{pmatrix} \mathbf{1}_1 & & & & \mathbf{0} \\ & \ddots & & & \\ & & \mathbf{1}_i & & \\ & & & \ddots & \\ & & & & \mathbf{1}_m \end{pmatrix} \qquad \boldsymbol{Z}_1 = \begin{pmatrix} \boldsymbol{x}_1 & & & & \mathbf{0} \\ & \ddots & & & \\ & & \boldsymbol{x}_i & & \\ & & & \ddots & \\ & & & & \boldsymbol{x}_m \end{pmatrix}.$$

More details and technicalities regarding mixed effects in additive models can be found in the respective literature.

Extensions of additive models to non-normal responses are consequently called *Generalized Additive Models (GAMs)*, which were first introducd by Hastie and Tibshirani [1986]. If additionally random effects are included, they are called *Generalized Additive Mixed Models (GAMMs)*.

Thus far, models with main effects and conceivably random effects have been introduced. Accordingly, these types of effects can likewise be combined with covariate interactions and/or spatial effects. Such models can be described in a unified framework and are titled as (possibly *Generalized*) *Structured Additive Regression Models (STARs)*,

$$y = f_1(\nu_1) + \ldots + f_q(\nu_q) + \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon.$$

The covariates $\nu_1, \ldots, \nu_q$ can be one- or multidimensional and the functions can be of different structure determining the type of effect.

3 Copulas & Dependence Structures

Multivariate distributions consist of the marginal distributions and the dependence structure between those marginals. These components can be specified separately in a single framework with the help of copula functions. This chapter introduces the concept of modelling such dependence structures with copulas, which is the main focus of this thesis. The core elements on this subject were picked up from McNeil et al. [2015] and Ruppert and Matteson [2015].

3.1 Introduction to Copulas

A d-dimensional function $C : [0,1]^d \to [0,1]$ is called a *copula*, if it is a Cumulative Distribution Function (CDF) with uniform margins, i.e.

$$P\left(U_1 \leq u_1, \ldots, U_d \leq u_d\right) = C\left(u_1, \ldots, u_d\right)$$

where $U_i,\ i = 1, \ldots, d$ are uniformly distributed Random Variables (RVs) in $[0,1]$.
Since C is a CDF, following properties emerge:

- $C(\boldsymbol{u}) = C\left(u_1, \ldots, u_d\right)$ is increasing in each component $u_i, i = 1, \ldots, d$.
- The i^{th} marginal distribution is obtained by setting $u_j = 1$ for $j \neq i$ and it has to be uniformly distributed
$$C\left(1, \ldots, 1, u_i, 1, \ldots, 1\right) = u_i$$
- For $a_i \leq b_i$, the probability $P\left(U_1 \in [a_1, b_1], \ldots, U_d \in [a_d, b_d]\right)$ mus be non-negative, so we obtain the *rectangle inequality*
$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C\left(u_{1,i_1}, \ldots, u_{d,i_d}\right) \geq 0, \tag{3.1}$$
where $u_{j,1} = a_j$ and $u_{j,2} = b_j$.

The reverse is also true, i.e. any function C that satisfies the above properties is a copula. Furthermore, $C\left(1, u_1, \ldots, u_{d-1}\right)$ is also a $(d-1)$-dimensional copula and thus all k-dimensional marginals with $2 < k < d$ are copulas.

Generalized Inverse

For a CDF, the *generalized inverse* is defined by

$$F^{\leftarrow}(y) := \inf\{x : F(x) \geq y\}$$

(similar to the definition of a *quantile function*).

□

Probability Transformation

If a RV Y has a continuous CDF F, then

$$F(Y) \sim U[0,1]. \tag{3.2}$$

□

The reverse of the *probability transformation* is the *quantile transformation*.

Quantile Transformation

If $U \sim U[0,1]$ and F be a CDF, then

$$P\left(F^{\leftarrow}(U) \leq x\right) = F(x) \tag{3.3}$$

□

The above two transformations allow us to move back and forth between $\mathbb{R}^d$ and $[0,1]^d$ and are the primary building blocks when it comes to copulas. Against this backdrop, *Sklar's theorem* is introduced, which is considered the foundation of all copula related applications.

Sklar's Theorem [Sklar, 1959]

Let F be a d-dimensional CDF with marginal distributions $F_i,\ i = 1,\ldots,d$. Then there exists a copula C such that

$$F(x_1,\ldots,x_d) = C(F_1(x_1),\ldots,F_d(x_d)) \tag{3.4}$$

for all $x_i \in \mathbb{R},\ i = 1,\ldots,d$.

The copula C is unique, if $\forall i = 1,\ldots,d$, F_i is continuous. Otherwise C is uniquely determined only on $Ran(F_1) \times \ldots \times Ran(F_d)$, where $Ran(F_i)$ is the range of F_i.

Conversely, if C is a d-dimensional copula and $F_1,\ldots,F_d$ are univariate CDF's, then F as defined in Equation 3.4 is a d-dimensional CDF.

□

If the copula has a Probability Density Function (PDF), then the *copula density* is defined as

$$c(\boldsymbol{u}) = \frac{\partial^d C\left(u_1,\ldots,u_d\right)}{\partial u_1 \cdots \partial u_d} \tag{3.5}$$

for a differentiable copula function C and the realization of a random vector $\boldsymbol{u} = (u_1, \ldots, u_d)$.

By virtue of Equation 3.4 in Sklar's theorem and given that

$$C(\boldsymbol{u}) = F\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right), \tag{3.6}$$

i.e. inversible CDFs $F_i,\ i = 1, \ldots, d,$ we can rewrite the copula density to

$$c(u_1, \ldots, u_d) = \frac{f\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right)}{\prod_{i=1}^{d} f_i\left(F_i^{\leftarrow}(u_i)\right)} \tag{3.7}$$

for densities f of F and $f_1, \ldots, f_d$ of the corresponding marginals.

Invariance Principal

Suppose the RVs $X_1, \ldots, X_d$ have continuous marginals and copula C. For strictly increasing functions $T_i : \mathbb{R} \to \mathbb{R}, i = 1, \ldots, d$, the RVs $T_1(X_1), \ldots, T_d(X_d)$ also have copula C.

□

Fréchet-Hoeffding Bounds

Let $C(\boldsymbol{u}) = C(u_1, \ldots, u_d)$ be any d-dimensional copula.

Then, for

$$W(\boldsymbol{u}) = \max\left\{\sum_{i=1}^{d} u_i - d + 1,\ 0\right\} \tag{3.8}$$

as well as

$$M(\boldsymbol{u}) = \min_{1 \le i \le d}\{u_i\}, \tag{3.9}$$

it holds that

$$W(\boldsymbol{u}) \le C(\boldsymbol{u}) \le M(\boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^d. \tag{3.10}$$

W is called the *lower Fréchet-Hoeffding bound* and M the *upper Fréchet-Hoeffding bound*.

Note that W is a copula if and only if $d = 2$, whereas M is a copula for all $d \ge 2$ (more on this later in Section 3.2.1).

□

3.2 Copula Classes

In this section we will take a look at three very popular *copula classes*, namely *fundamental, elliptical and archimedean copulas*. For each class, a few (parametric) *copula families*, which are widely used, will be presented.

3.2.1 Fundamental Copulas

Fundamental copulas are a basic class of copulas, which emerge directly from the copula framework and do not depend on any parametric components.

Independence Copula

It is well known that the joint CDF of a finite set of RVs $X_i, i = 1, \ldots, n$, is equal to the product of the marginals if and only if the RVs X_i are mutually independent, i.e.

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i))$$

$\forall x_1, \ldots, x_n$.

Equally, the exact same concept applies when we talk about the *independence copula*

$$\Pi(\boldsymbol{u}) = \prod_{i=1}^{d} u_i. \tag{3.11}$$

As a result of Sklar's theorem the RVs u_i are independent if and only if their copula is the independence copula, i.e.

$$C(\boldsymbol{u}) = \Pi(\boldsymbol{u})$$

and thus the copula density would be

$$c(\boldsymbol{u}) = 1, \quad \boldsymbol{u} \in [0,1]^d.$$

□

From Equation 3.10, it is obvious that the Fréchet-Hoeffding bounds correspond to the extreme cases of perfect dependence between the RVs $X_i, i = 1, \ldots, d$.

Comonotonicity Copula

Consider the RVs $X_1, \ldots, X_d$ and strictly increasing transformations $T_1, \ldots, T_d$ and $X_i = T(X_i)$ for $i = 2, \ldots, d$. Making use of the *invariance principal*, it can be shown that these RVs have as copula the upper Fréchet-Hoeffding bound

$$M(\boldsymbol{u}) = \min\{u_1, \ldots, u_d\}.$$

Since there is perfect positive dependence between those RVs, M is called the *comonotonicity copula*. The number of dimensions d can be any finite number greater than or equal to 2 for M to be a copula, as the minimum remains well defined.

□

Countermonotonicity Copula

Similar to the comonotonic case, it can be shown that if two RVs X_1 and X_2 are perfectly negatively dependent, their copula is the lower Fréchet-Hoeffding bound

$$W(\boldsymbol{u}) = \max\left\{\sum_{i=1}^{d} u_i - d + 1,\ 0\right\}.$$

Therefore, W is known as the *countermonotonicity copula*. Because of the fact that countermonotonicity is not valid for a dimension greater than 2, we end up with the restriction $d = 2$ for W to be indeed a copula.

□

3.2.2 Elliptical Copulas

Copulas which can be derived from known multivariate distributions like for example the *Multivariate Normal (or Gaussian) Distribution* or the *Multivariate Student's t-Distribution* are called *implicit copulas*. *Elliptical copulas* are implicit copulas which arise via Sklar's theorem from elliptical distributions like the mentioned examples.

Gaussian Copula

W.l.o.g., for a random vector $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{P})$ and *correlation matrix* $\boldsymbol{P}$, the *Gaussian copula (family)* is given by

$$C_{\boldsymbol{P}}^{Ga}(\boldsymbol{u}) = \Phi_{\boldsymbol{P}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right), \tag{3.12}$$

where Φ is the CDF of $\mathcal{N}(0, \sigma^2)$ and $\Phi_{\boldsymbol{P}}$ is the CDF of $\mathcal{N}_d(\boldsymbol{0}, \boldsymbol{P})$.
There are special cases to this copula family, namely for $d = 2$ and correlation ρ, the *bivariate Gaussian copula* C_ρ^{Ga} is equivalent to

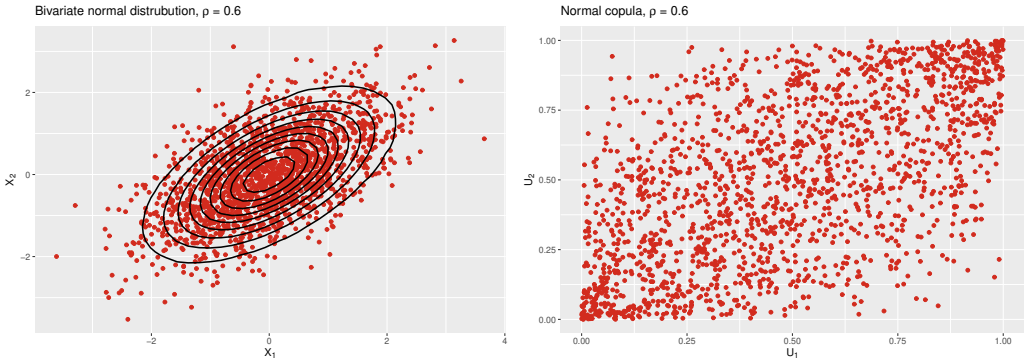
- the independence copula Π if $\rho = 0$,
- the comonotonicity copula M if $\rho = 1$ and
- the countermonotonicity copula W if $\rho = -1$

The density of the Gaussian copula is given by

$$c_{\boldsymbol{P}}^{\mathrm{Ga}}(\boldsymbol{u}) = \frac{1}{\sqrt{\det \boldsymbol{P}}} \exp\left(-\frac{1}{2}\boldsymbol{x}'\left(\boldsymbol{P}^{-1} - \boldsymbol{I}_d\right)\boldsymbol{x}\right), \tag{3.13}$$

where $\boldsymbol{x} = \left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)$.

□



(a) Gaussian distribution with contour lines

(b) Gaussian copula

Figure 3.1: Bivariate Gaussian distribution and Gaussian copula for Pearson's $\rho = 0.6$ and simulated sample of size $n = 1800$, both with standard normal marginals

t-Copula

Consider w.l.o.g. $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{0}, \boldsymbol{P})$ (multivariate Student's t-distribution) with ν Degrees of Freedom (d.o.f.) and $\boldsymbol{P}$ a correlation matrix, then the *t-copula (family)* is given by

$$C_{\nu,\boldsymbol{P}}^{t}(\boldsymbol{u}) = t_{\nu,\boldsymbol{P}}\left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\right), \tag{3.14}$$

where t_ν is the CDF of the univariate Student's t-distribution and $t_{\nu,\boldsymbol{P}}$ is the CDF of the multivariate Student's t-distribution (both with ν d.o.f.).

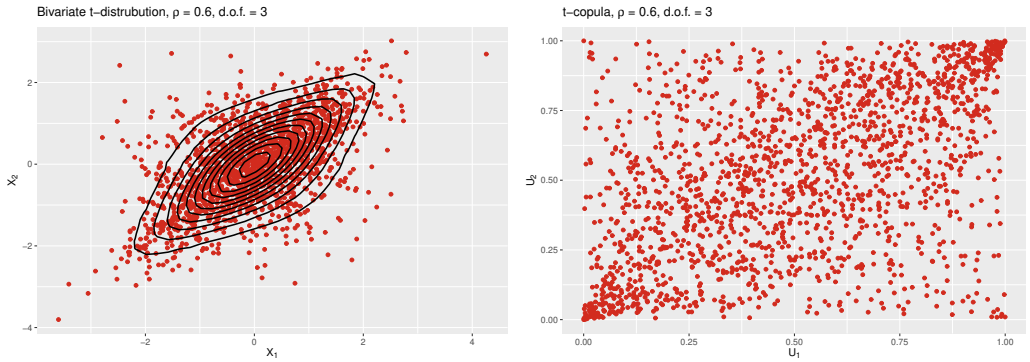
For the *bivariate t-copula* ($d = 2$), the special cases are the same as for the Gaussian copula except that $d = 0$ does not yield the independence copula (unless $\nu \to \infty$ in which case $C_{\nu,\rho}^{t} = C_{\rho}^{Ga}$).

The density of $C_{\nu,\boldsymbol{P}}^{t}$ is given by

$$c_{\nu,\boldsymbol{P}}^{t}(\boldsymbol{u}) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\sqrt{\det \boldsymbol{P}}} \left(\frac{\Gamma(\nu/2)}{\Gamma((\nu+1)/2)}\right)^d \frac{\left(1 + \boldsymbol{x}'\boldsymbol{P}^{-1}\boldsymbol{x}/\nu\right)^{-(\nu+d)/2}}{\prod_{j=1}^{d}\left(1 + x_j^2/\nu\right)^{-(\nu+1)/2}}, \tag{3.15}$$

where $\boldsymbol{x} = \left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\right)$.

□



(a) t-distribution with contour lines

(b) t-copula

Figure 3.2: Bivariate t-distribution and t-copula with 3 degrees of freedom for Pearson's $\rho = 0.6$ and simulated sample of size $n = 1800$, both with standard normal marginals

3.2.3 Archimedean Copulas

Unlike implicit copulas, *explicit copulas* can be specified directly by taking into account certain constructional principles. The most important aspects of a such explicit copulas, in particular *archimedean copulas*, are showcased in this subsection. Archimedian copulas are of the general form

$$C(\boldsymbol{u}) = \phi^{-1}\left(\phi\left(u_1\right) + \cdots + \phi\left(u_d\right)\right), \tag{3.16}$$

where the function $\phi : [0,1] \to [0,\infty)$ is the *(archimedean) generator* and satisfies the following properties:

- ϕ is strictly decreasing in the entire domain $[0,1]$
- We set $\phi(1) = 0$
- If $\phi(0) = \lim_{u \to 0^-} \phi(u) = \infty$, then ϕ is called *strict*.

Based on Equation 3.16 and according to the form of the generator, we can construct several copula families. Three of the most popular ones are the *Gumbel*, the *Clayton* and the *Frank copula*, which will be discussed.[4] The advantage of such copulas lies in the fact that they interpolate between certain fundamental dependence structures.

[4]We will look into these copulas for the bivariate case ($d = 2$) only.

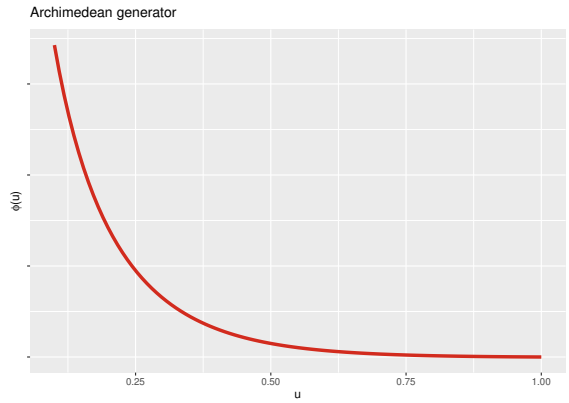


Figure 3.3: Shape of a generator function

Clayton Copula

If the generator takes on the form

$$\phi_{Cl}(u) = \frac{1}{\theta}\left(u^{-\theta} - 1\right) \tag{3.17}$$

then we obtain the *Clayton copula* given by

$$C_\theta^{Cl}(u_1, u_2) = \left(\max\left\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\right\}\right)^{-\frac{1}{\theta}}, \tag{3.18}$$

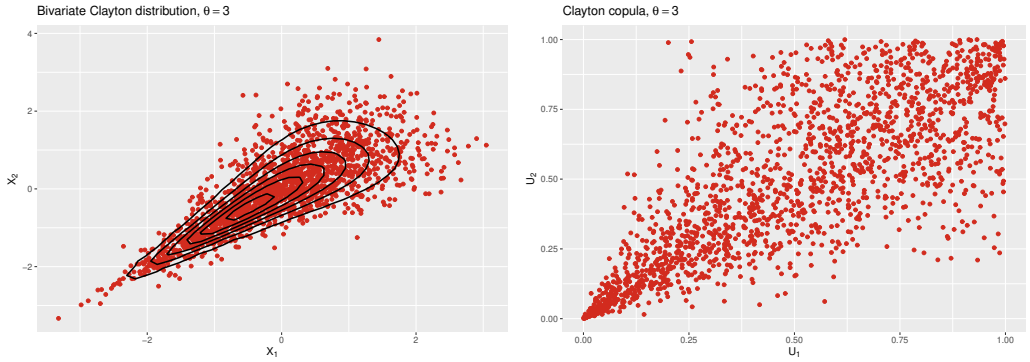
where $\theta \in [-1, \infty)\backslash\{0\}$.

For $\theta > 0$ the generator of the Clayton copula is strict and we arrive at

$$C_\theta^{Cl}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}. \tag{3.19}$$

Note that for $\theta = -1$, we obtain the lower Fréchet-Hoeffding bound W, whereas for the limits $\theta \to 0$ and $\theta \to \infty$ we arrive at the independence copula Π and the comonotonicity copula M respectively.

□



(a) Clayton distribution with contour lines

(b) Clayton copula

Figure 3.4: Bivariate Clayton distribution and Clayton copula for Kendall's $\tau = 0.6$ and simulated sample of size $n = 1800$, both with standard normal marginals

Gumbel Copula

If the generator takes on the form

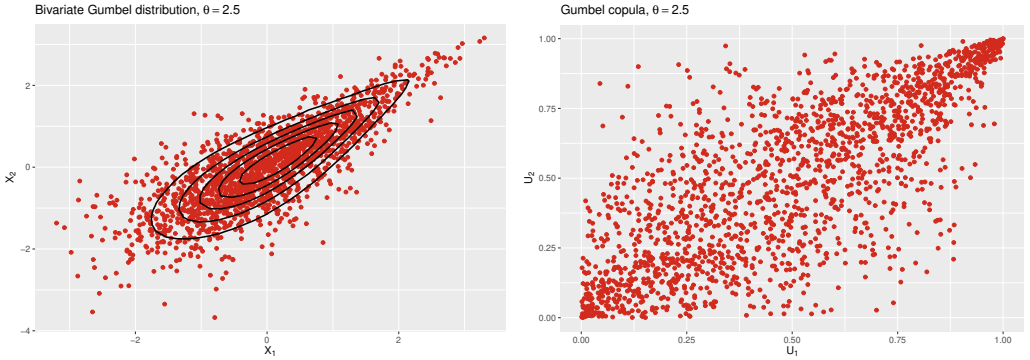
$$\phi_{Gu}(u) = (-\ln u)^{\theta}, \quad \theta \in [1, \infty), \tag{3.20}$$

then we arrive at the *Gumbel copula* given by

$$C_{\theta}^{Gu}(u_1, u_2) = \exp\left[-\left((-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right)^{\frac{1}{\theta}}\right]. \tag{3.21}$$

Note that for $\theta = 1$, we obtain the independence copula Π, while for $\theta \to \infty$ the Gumbel copula converges to the comonotonicity copula M. Strictness holds for the entire parameter range of θ.

□



(a) Gumbel distribution with contour lines

(b) Gumbel copula

Figure 3.5: Bivariate Gumbel distribution and Gumbel copula for Kendall's $\tau = 0.6$ and simulated sample of size $n = 1800$, both with standard normal marginals

Frank Copula

If the generator takes on the form

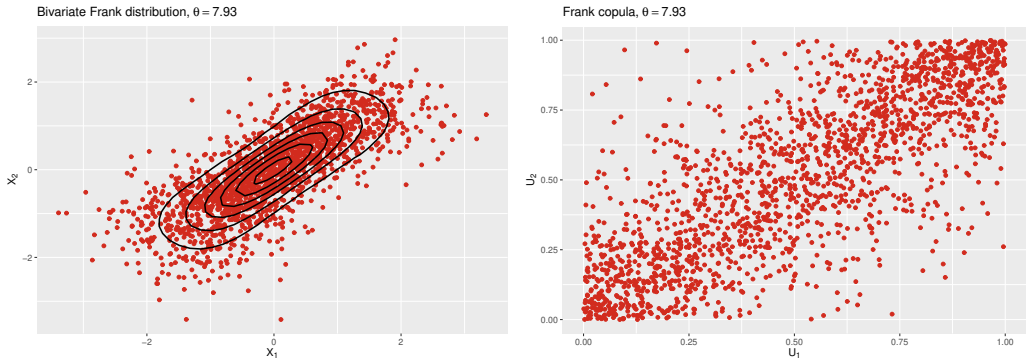
$$-\ln\left(\frac{\mathrm{e}^{-\theta u} - 1}{\mathrm{e}^{-\theta} - 1}\right), \quad \theta \in \mathbb{R}\backslash\{0\}, \tag{3.22}$$

we obtain the *Frank copula* given by

$$C_{\theta}^{Fr}(u_1, u_2) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta u_1} - 1\right)\cdot\left(e^{-\theta u_2} - 1\right)}{e^{-\theta} - 1}\right). \tag{3.23}$$

The Frank copula is strict in the parameter range of θ and interpolates between W ($\theta \to -\infty$), Π ($\theta \to 0$) and M ($\theta \to \infty$).

□



(a) Frank distribution with contour lines

(b) Frank copula

Figure 3.6: Bivariate Frank distribution and Frank copula for Kendall's $\tau = 0.6$ and simulated sample of size $n = 1800$, both with standard normal marginals

3.3 Dependence Measures

Dependence measures allow us to summarize a particular kind of dependence into a single number.[5] Recall the Fréchet-Hoeffding bounds (Equation 3.8 and Equation 3.9). They are an example of such kind of dependence measures. After all, they represent perfect negative or positive dependence. In this section, we will take a closer look into three classes of dependence measures along with appropriate association metrics.

3.3.1 Linear Correlation

Undoubtedly, the most famous association metric for two RVs X_1 and X_2 is the *Linear or Pearson's correlation coefficient*

$$\rho\left(X_1, X_2\right) = \frac{\operatorname{Cov}\left(X_1, X_2\right)}{\sqrt{\operatorname{Var}(X_1)}\sqrt{\operatorname{Var}(X_2)}} \in [-1, 1]. \tag{3.24}$$

Note that $E(X_1) < \infty$ and $E(X_2) < \infty$ have to hold, i.e. the first two moments have to exist for ρ to be defined.

The Pearson correlation coefficient is interpretable for RVs which have (approximately) a linear relationship, where $\rho = -1$ indicates perfect negative linear correlation, $\rho = 1$ indicates perfect positive linear correlation and $\rho = 0$ indicates no correlation between X_1 and X_2. However, comprehensibility of this measure comes along with some drawbacks:

- A correlation of 0 is in general not equivalent to independence. This property holds only for normally distributed RVs.[6]

[5] In the bivariate case

[6] e.g. $X_2 = X_1^2$ implies perfect dependence, yet $\rho(X_1, X_2) = 0$. Conversely though, independence always yields $\rho = 0$.

- ρ is invariant only under linear transformations, but not under transformations in general.
- Given the marginals and correlation ρ, one is able to construct a joint distribution only for the class of elliptical distributions.
- Given the marginals, only for elliptically distributed RVs any $\rho \in [-1, 1]$ is attainable.

3.3.2 Rank Correlation

To compensate some of the drawbacks of linear correlation, we can take advantage of correlation measures based on the ranks of data. *Rank correlation coefficients*, like the ones presented below, are always defined and obey to the invariance principal. This means that these coefficients only depend on the underlying copula and they can thereof be directly derived.

Spearman's Rho

Consider two RVs X_1 and X_2 with continuous CDFs F_1 and F_2, then the *Spearman's rho correlation coefficient* is simply the linear correlation between the CDFs

$$\rho_S = \rho\left(F_1\left(X_1\right), F_2\left(X_2\right)\right). \tag{3.25}$$

The reason being is that by applying the CDF to data, naturally a multiple of the ranks of the data are obtained, which essentially is equivalent to

$$\rho_S = \rho(Ran(X_1), Ran(X_2)) \tag{3.26}$$

Due to the invariance principle, we also obtain Spearman's rho directly from the unique copula via

$$\rho_{\mathrm{S}} = 12 \int_0^1 \int_0^1 C\left(u_1, u_2\right) \mathrm{d}u_1 \mathrm{d}u_2 - 3. \tag{3.27}$$

□

Kendall's Tau

Let $X_1 \sim F_1$ and $X_2 \sim F_2$ be two RV and let $(\tilde{X}_1, \tilde{X}_2)$ be an independent copy[7] of (X_1, X_2). Then *Kendall's tau* is defined by

$$\begin{aligned} \rho_\tau &= E\left[\operatorname{sign}\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right)\right)\right] \\ &= P\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right) > 0\right) - P\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right) < 0\right). \end{aligned} \tag{3.28}$$

[7] An independent copy $\tilde{X}$ of a RV X is a RV that inherits from the same distribution as X and is independent of X.

Similarly to Spearman's rho, using the invariance principal, we can directly derive Kendall's tau from the unique copula by

$$\rho_\tau (X_1, X_2) = 4 \int_0^1 \int_0^1 C (u_1, u_2) \, dC (u_1, u_2) - 1. \tag{3.29}$$

□

Both $\rho_S, \rho_\tau \in [-1, 1]$ and any value within this interval is attainable for an arbitrary copula class in contrast to the Pearson coefficient. If any of these rank correlations is -1 (or 1), we are in the countermonotonic (or comonotonic) case. If ρ_S (or ρ_τ) $= 0$, this does not necessarily imply independence between X_1 and X_2, although the opposite direction holds. Furthermore, they are not limited to be invariant just under linear transformations .

3.3.3 Tail Dependence

Coefficients of tail dependence express the strength of the dependence in the extremes of distributions, i.e. the joint tails. We distinguish between *lower* and *upper tail dependence* between $X_j \sim F_j, j = 1, 2$ and provided that the below limits exist, they are given by

$$\lambda_l = \lim_{q \to 0^+} P \left(X_2 \leq F_2^{\leftarrow}(q) | X_1 \leq F_1^{\leftarrow}(q) \right) \tag{3.30}$$

and

$$\lambda_u = \lim_{q \to 1^-} P \left(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q) \right). \tag{3.31}$$

If λ_l (or λ_u) $= 0$, then we say that X_1 and X_2 are *asymptotically independent* in the lower (or upper) tail,[8] otherwise we have lower (or upper) tail dependence.

For continuous CDFs and by using Bayes' theorem, these expressions can be re-written to

$$\begin{aligned} \lambda_l &= \lim_{q \to 0^+} \frac{P \left(X_2 \leq F_2^{\leftarrow}(q), X_1 \leq F_1^{\leftarrow}(q) \right)}{P \left(X_1 \leq F_1^{*}(q) \right)} \\ &= \lim_{q \to 0^+} \frac{C(q, q)}{q} \end{aligned}$$

and similarly

$$\lambda_u = 2 - \lim_{q \to 1^-} \frac{1 - C(q, q)}{1 - q}.$$

Therefore, tail dependencies can be assessed by means of the copula itself when approaching the points $(0, 0)$ and $(1, 1)$. In addition, for all radially symmetric copulas (e.g. the bivariate Gaussian or the t-copula) we have $\lambda_l = \lambda_u = \lambda$.

Some examples are:

- Clayton: $\lambda_l = 2^{-1/\theta}$, $\lambda_u = 0$ (only lower tail dependence, see Figure 3.4)

[8] Not necessarily true for the other way around

- Gumbel: $\lambda_l = 0$, $\lambda_u = 2 - 2^{1/\theta}$ (only upper tail dependence, see Figure 3.5)
- Frank: $\lambda_l = 0$, $\lambda_u = 0$ (no tail dependence, see Figure 3.6)

Following such guidelines, the choice of a practicable copula can be facilitated. Table 3.1 displays an overview of the relationships between dependence measures and θ parameters of various copulas.

Copula \ Measure	τ	ρ_s	λ_l	λ_u
Gaussian	$\frac{2}{\pi} arcsin(\rho)$	$\frac{6}{\pi} arcsin(\rho)$	0	0
Student's t	$\frac{2}{\pi} arcsin(\rho)$	-	$2T_{\nu+1}(\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}})$	$2T_{\nu+1}(\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}})$
Clayton	$\frac{\theta}{\theta+2}$	-	$2^{-1/\theta}$	0
Gumbel	$\frac{\theta-1}{\theta}$	-	0	$2 - 2^{1/\theta}$
Frank	$1 - \frac{4}{\theta}(4 - D_1(\theta))$	$1 - \frac{12}{\theta}(D_1(\theta) - D_2(\theta))$	0	0

Table 3.1: Bivariate relationships in copula families, with T_ν being the Student's t-distribution function with ν degrees of freedom and $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$ being the Debye function [stanfordphd]

3.4 Structured Additive Conditional Copulas

Modelling of the marginal response distributions along with their dependence structure has been studied so far in a strictly parametric context, not considering any potentially available covariate information. In this section, the copula framework will be broadened by adding conditions given possible covariates for all model parameters, i.e. both for the parameters of the marginals as well as the copula parameter. All involved model parameters will receive *structured additive predictors* (see Section 2.3) to account for possible non-linear or random effects. We will summarily explore *Structured Additive Conditional Copulas* and for extensive literature, good references to view are Klein and Kneib [2016], Vatter and Nagler [2019] and Marra and Radice [2016].

To get started, we define $(Y_1, Y_2)'$ to be independent bivariate responses and $\boldsymbol{\nu}$ being the information contained in covariates. Ergo, Equation 3.4 of Sklar's theorem can be extended to the conditional case

$$F_{1,2}(Y_1, Y_2|\boldsymbol{\nu}) = C(F_1(Y_1|\boldsymbol{\nu}), F_2(Y_2|\boldsymbol{\nu})|\boldsymbol{\nu}) \tag{3.32}$$

in conjunction with all facets of Section 3.1 [Patton, 2006].
The marginal CDFs $F_d(y_{id}|\boldsymbol{\nu}_i)$ for observations $i = 1, \ldots, n$ can also be stated as

$$F_d\left(y_{id}|\vartheta_{i1}^{(d)}, \ldots, \vartheta_{iK_d}^{(d)}\right), \quad d = 1, 2, \tag{3.33}$$

i.e. the distribution F_d has a total of K_d parameters, denoted as $\vartheta_{i1}^{(d)}, \ldots, \vartheta_{iK_d}^{(d)}$. To relate all parameters of the marginals to structured additive predictors $\eta_i^{\vartheta_k^{(d)}}, k = 1, \ldots, K_d$ consisting of the covariates $\boldsymbol{\nu}_i$ (see Section 2.3), we employ strictly increasing response mappings $h_k^{(d)}$ to ensure proper domain allocation, i.e.

$$\vartheta_{ik}^{(d)} = h_k^{(d)}(\eta_i^{\vartheta_k^{(d)}}). \tag{3.34}$$

Assuming that the parameters of the copula can also depend on covariates $\boldsymbol{\nu}_i$ while Sklar's theorem applies as usual, the left-hand side of Equation 3.32 can equivalently be stated as

$$F_{1,2}(y_{i1}, y_{i2}|v_i) = F_{1,2}(y_{i1}, y_{i2}|\vartheta_{i1}^{(1)}, \ldots, \vartheta_{iK_1}^{(1)}, \vartheta_{i1}^{(2)}, \ldots, \vartheta_{iK_2}^{(2)}, \vartheta_{i1}^{(c)}, \ldots, \vartheta_{iK_c}^{(c)}),$$

where the last share of parameters $\vartheta_{i1}^{(c)}, \ldots, \vartheta_{iK_c}^{(c)}$ belong to the copula. Similar to Equation 3.34, the copula parameters are modelled as $\vartheta_{ik}^{(c)} = h_k^{(c)}(\eta_i^{\vartheta_k^{(c)}})$ with K_c being the number of parameters.

3.5 Vine Copulas

Vine copulas to be written down...

4 Data Exploration

In Section 1.2 the setup of the data to be treated was introduced. As can be seen in Table 1.2, each article can be assigned to a set of attributes. Besides some elemental attributes like *color*, *age group* or *gender*, the data exhibit a "natural" company-specific hierarchical structure. In Figure 4.1, we can see an example of such a hierarchy for the attributes *Key Category Cluster (KCC)* and *Business Segment (BS)* (See Table 1.2). The bottom level consists of the individual articles and at the top level we have the brand. It is important to mention that there are more inner levels between the brand and the articles than depicted in Figure 4.1 below. For example, Key Category (KC) would be the level below KCC. KCCs are aggregated sport/fashion categories and KCs add an additional layer to KCCs, namely the *Product Division* covering Footwear, Apparel and Accessories/Hardware. The BS supplements the KC with a consumer driven "gender" perception. Within the scope of this thesis, we will be concerned with the hierarchical structure of Figure 4.1 and in particular our KCCs of interest are *"KCC 2"*, *"KCC 6"* and *"KCC 8"*.

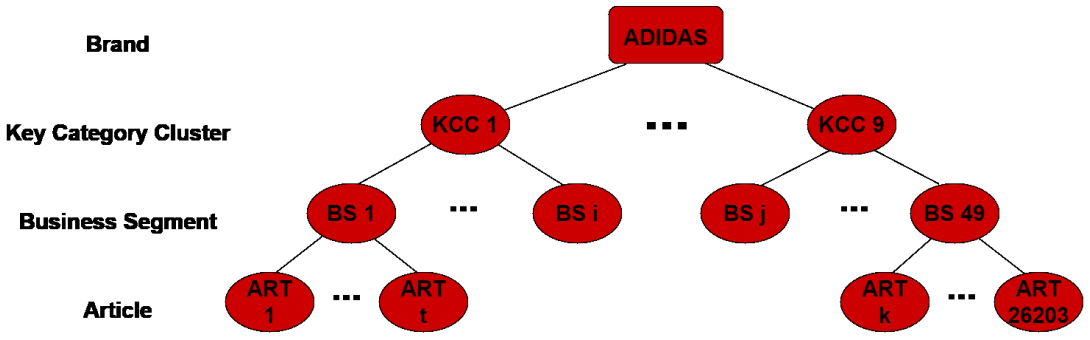


Figure 4.1: Illustration of a hierarchical article structure

Worth mentioning is that it is possible that some individual nodes might have only one single child node, meaning that the hierarchy level can stay consistent across multiple nodes. This phenomenon however is very rare and when it occurs, it affects usually two consecutive nodes only. For example, *Sub-Brand 4* has only one child node *KCC 6* (See Figure 4.2). Sub-Brands are visible for consumers through an own, not shared logo (See Figure 1.1).

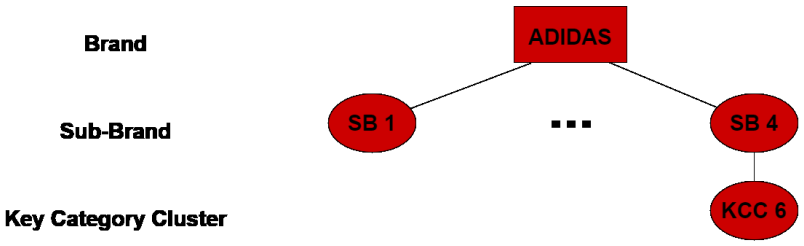


Figure 4.2: Example of a single child node

4.1 Universal Sales Patterns

As mentioned in Section subsection 1.2, our data contains the information about sold articles over the years 2017 and 2018. Figure 4.3 shows the weekly course for the quantities over those two years, highlighting active promotion weeks as vertical lines. We can undoubtedly recognize that *"Black Friday"* weeks (black lines) have an exceptional impact on sales, as they stand internationally for the most busy shopping periods. During these days in mid- to late November each year, large amounts of different products are heavily discounted.

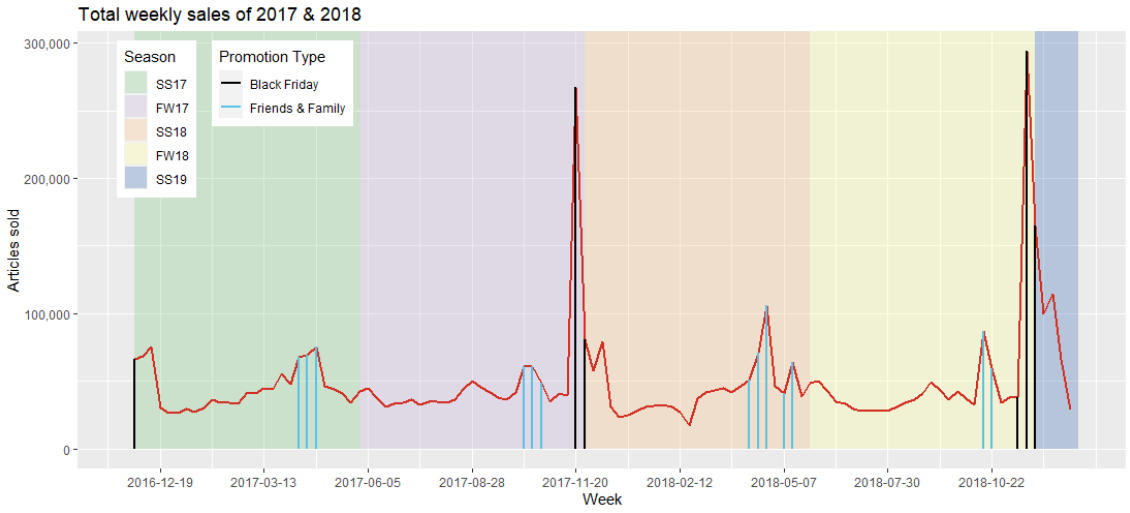


Figure 4.3: Course of article unit sales

Another promotion type we are interested in is *"Friends & Family"*, occurring yearly around April-May and October, where on the eCom website plenty of articles are on offer. On these weeks, we have elevated numbers of sold articles as well (Figure 4.3, blue lines). Tables 4.1 and 4.2 show the weeks were Black Friday and Friends & Family took place respectively. The dates indicate always the Monday of the respective week (according to European standards, a week starts on Monday).

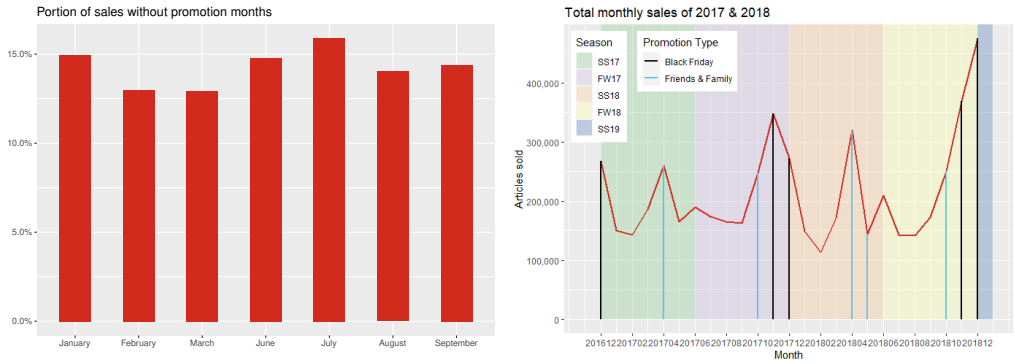
Black Friday weeks	2016-11-28	2017-11-20	2017-11-27	2018-11-12	2018-11-19	2018-11-26

Table 4.1: Black Friday weeks

Friends & Family weeks	2017-04-10	2017-04-17	2017-04-24	2017-10-09	2017-10-16	2017-10-23	2018-04-09
	2018-04-16	2018-04-23	2018-05-07	2018-05-14	2018-10-15	2018-10-22	

Table 4.2: Friends & Family weeks

The months in general don't seem to differ in their sale quantities, that is of course excluding the months of the two mentioned promotion types. The monthly sales portions for those months can be observed in Figure 4.4a. We have slightly higher sales portions on January, June and July. Most probably this is due to Christmas and "end of the year" shopping habits, which are carried forward from December to the very next month January. Regarding June and July, they indicate summer periods and frequent occurrence of big sports events, which may drive sales. Notice that, despite the fact that December is not explicitly present in Table 4.1, it is nonetheless a Black Friday month as the promotion is still activated moving from November to December (see Figure 4.4b).



(a) Sales distribution grouped by months without promos

(b) Monthly course of article unit sales with promo indicators

Figure 4.4: Monthly patterns of article unit sales

Moving forward, reviewing some sales summary statistics along with the findings so far, we detect a very high overdispersion in our data. The first two rows in Table 4.3 give as a first impression of the sales distribution. Considering the sold units of one article at a time within a week, there are lots of weeks where no single unit was sold. The median is at 2.71 units,[9] 75% of the "article-week" combinations take on a value of at most 20 and the minority exceeds 100 pieces (99%-quantile). The third row of the table shows

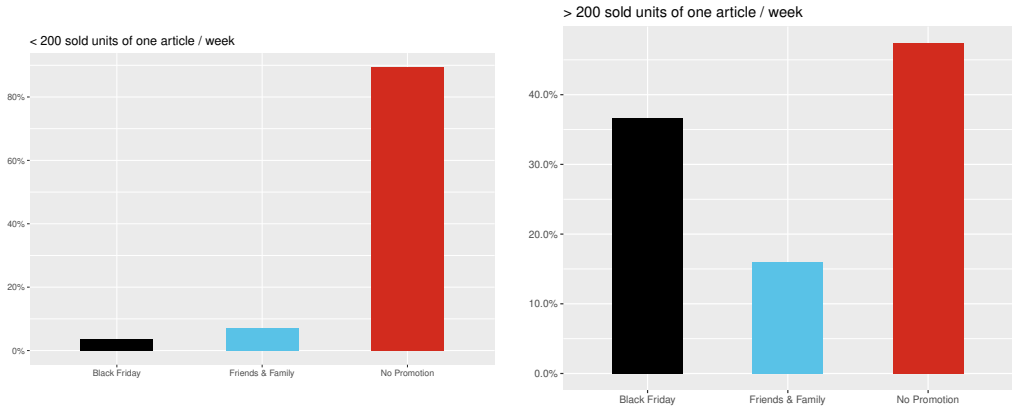
[9]Reminder from Section 1.2: the values are in reality discrete, but due to anonymization the were transformed into real numbers.

how many distinct articles fall under the respective quantile of sales and there is a visible anti-proportional behaviour towards the number of sold units, which is of course intuitive. Remarkable though is the quantity of affected articles even for incredibly large quantiles.

Quantile	Min	25%	50%	75%	90%	95%	99%	99.9%	Max
# Sold units / week	0	0.45	2.71	8.14	19.45	34.39	102.71	360.12	6,816.74
# Affected articles	26,203	26,195	23,797	17,014	10,275	6,458	1,800	273	1

Table 4.3: Number of sold units per week & number of affected articles for various quantiles of sales

Conscientiously, we want to inspect the number of weekly sold units above and below a certain (large) threshold to find out how promotions influence these vast sales numbers. In Figure 4.5a we can see how the sales are distributed over Black Friday, Friends & Family and regular weeks for below a threshold of 200 units. Most high sales occurrences are not attached to any of the two big promotions. They might be due to other events or unrelated to any campaigns altogether. Observations above that threshold of 200 can be seen in Figure 4.5b and we can clearly see a change from Figure 4.5a. As expected, Black Friday is the dominating promotion type, although the majority remains in not promoted sales.



(a) Below 200 units

(b) Above 200 units

Figure 4.5: Distribution of sold units of articles per week split at 200 units

To validate our exploration on this, we may look at the empirical CDF in Figure 4.6 using all observations now. Instances with no promotions have a steeper curve (red line) and reach their maximum faster compared to articles tagged with a promotion in a certain week. The less concave curve of Friends & Family promoted sales (blue line) implies that there are more instances with a larger amount of sold units overall. The same behaviour is even more pronounced for Black Friday (black line), having considerably more high

quantity instances. Along these lines, promotions might somewhat explain this pattern better.

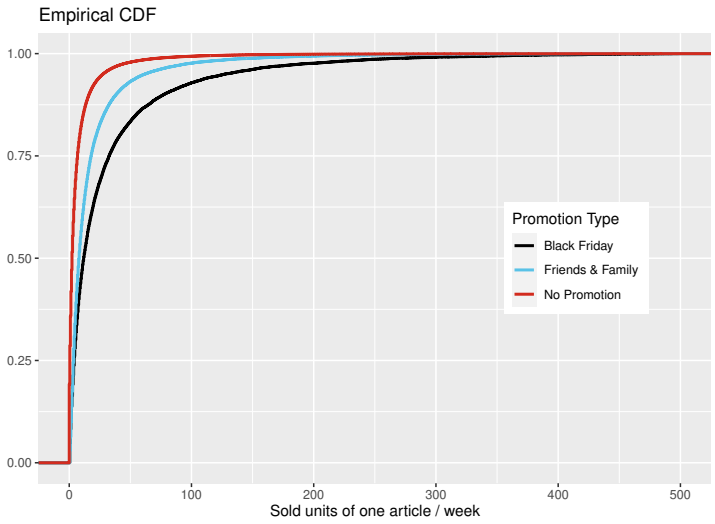


Figure 4.6: Empirical CDF of all sold units per week; x-axis cut at 500

Unfortunately, we cannot just remove such outliers with extreme sale numbers from the dataset, as it would produce gaps in the time series of some specific articles. Removing entire articles from the analysis is also not an option at this point, since we would be forced to remove a lot of articles. There will be a data delimitation process down the line to deal with this issue (see Subsection 5.3.1). For example, 273 articles alone would have to be removed to get rid of the highest 0.01% of quantities (see Table 4.3). Besides, these extreme values might be too informative for the underlying data generating process, so we decide to keep them all.

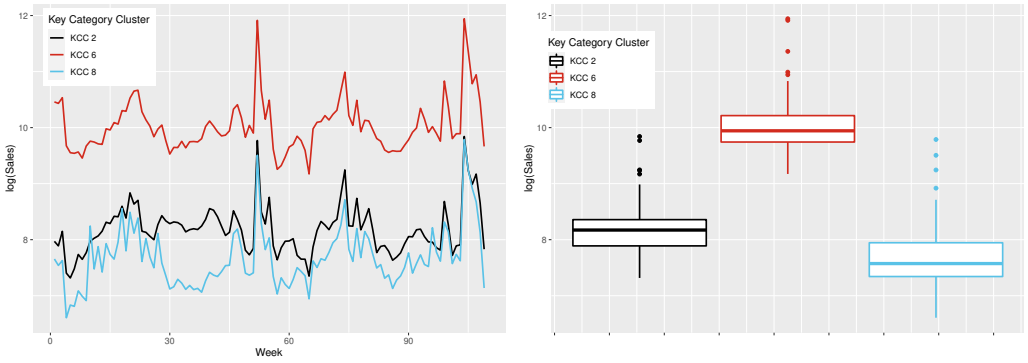
4.2 Grouped Patterns

To gain some insights on the hierarchical levels of interest, we perform some quick analysis on different groups of the upper levels of the tree in Figure 4.1. We start out with a broad picture on the key category clusters in Subsection 4.2.1 and eventually reach a better understanding for the the article behaviour.

4.2.1 Exploring Key Category Cluster

By viewing the sales trends separately for each key category cluster, we can observe in Figure 4.7a that, among the weekly noise, they climax similarly. Just like in the previous section, those peaks come about primarily during the big promotions weeks. To put it into

perspective, we can see logarithmic sales behaviour in Figure 4.7b, were patterns are quite similar although they differ strongly in volume.



(a) Time series of log(Sales) per KCC (b) Boxplots of log(Sales) per KCC

Figure 4.7: Time series and boxplot showing logarithmized sales of the key category clusters

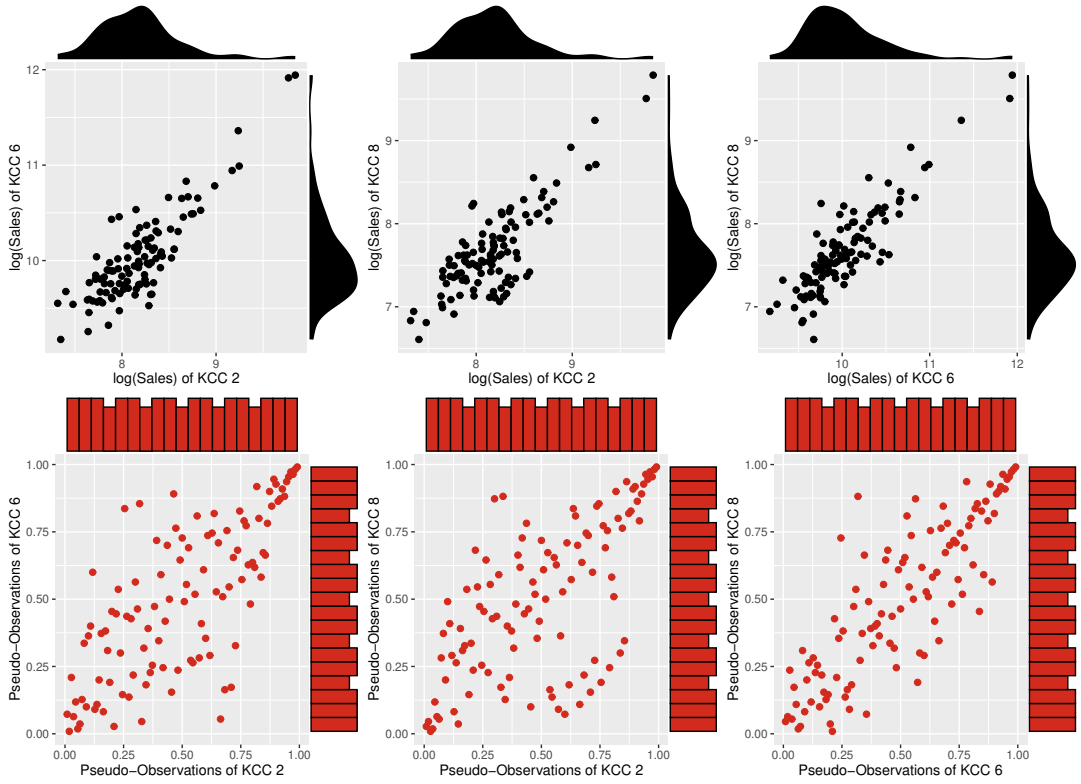


Figure 4.8: Pairwise scatterplots of sales on KCC level. First row: Logarithmic sales with marginal densities, Second row: Pseudo sales observation with marginal histograms

All the more interesting are the joint distributions of our KCCs. Figure 4.8 shows scatterplots of KCC pairs. In the first row we can see some isolated points on the upper tails representing outliers. We took the logarithmized sales to spot differences that would be otherwise hard to see. The outliers produced by the promotions still remain outliers in the log-scale. Also, by checking the densities for the marginals, pertinent marginal distribu-

tions are hardly determined but not to be ruled out.

The second row displays the pairs of the according *pseudo observations*. Pseudo observations are calculated by taking the data ranks and dividing them by (1 + number of observations), which makes them robust against outliers and restricts the value range to $(0, 1)$. Here we are faced with a strange behaviour of the histograms. They practically look uniformly distributed, however there are seemingly regular step patterns in the pseudo data. This might be traced back to the fact that we are dealing in reality with discrete data of not necessarily unique occurrence.

For the above reasons, on KCC level we will attempt modelling parametric distributions to the marginals as well as directly use the pseudo data. The reason being for the latter is that our primary objective is to capture a dependence structure, whereas the marginals per se are of secondary interest. In addition, looking at both rows of Figure 4.8, we suspect tail dependence and there is an obvious strong positive correlation among all three pairs, which is confirmed by viewing Figure 4.9.

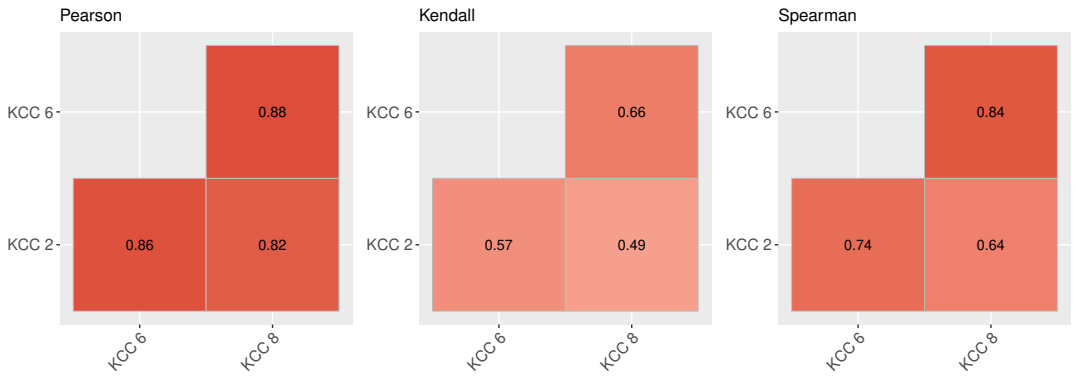


Figure 4.9: Correlation plots of the three KCC log-sales with different correlation coefficients. Left: Pearson's rho, Middle: Kendall's tau, Right: Spearman's rho

5 Modelling

In this chapter, we will explore various ways on how to model, evaluate and interpret the dependence structure of different hierarchy levels. We first start at key category cluster level and compare the capabilities of the tried and tested methods on article unit sales.

One side note on the promotion intensities of Black Friday and Friends & Family (see Table 1.1) is that on higher levels such as key category cluster, as we aggregate our data, promotion intensities become binary values indicating whether the respective promotion took place in those respective weeks. Also note that Black Friday and Friends & Family weeks do not overlap. When we aggregate our data, not only promotion intensities need to be adapted. In the course of this, we average the total markdown percentage of all articles belonging to a respective aggregation group, e.g. each key category cluster obtains its own mean total markdown percentage. However, when we are trying to model dependence measures of pairwise groups as functions of covariates, we need a unifying feature for those groups. Luckily, we can readily average the aggregated total markdown percentages. Figure 5.1 clarifies the strong linear correlations of those aggregated markdown percentages, therefore this simple solution should be rational.

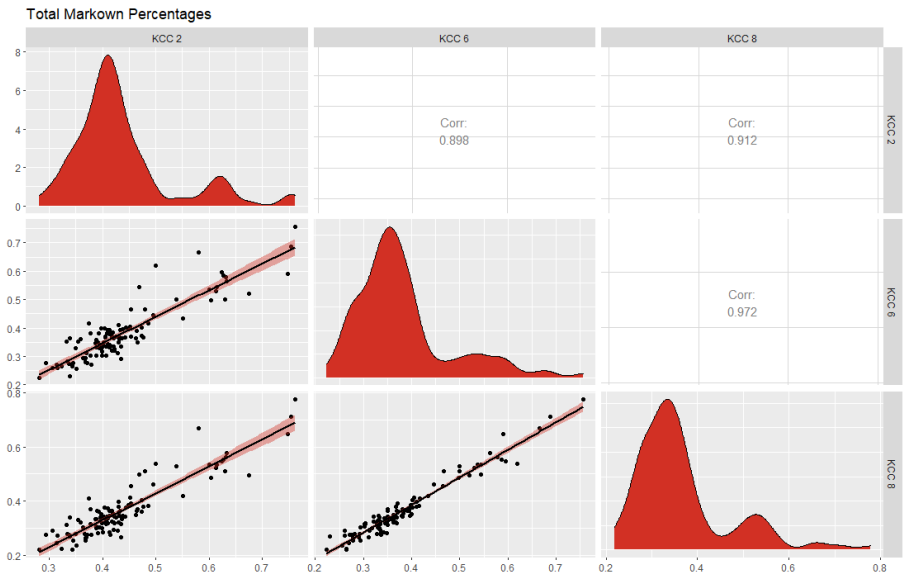


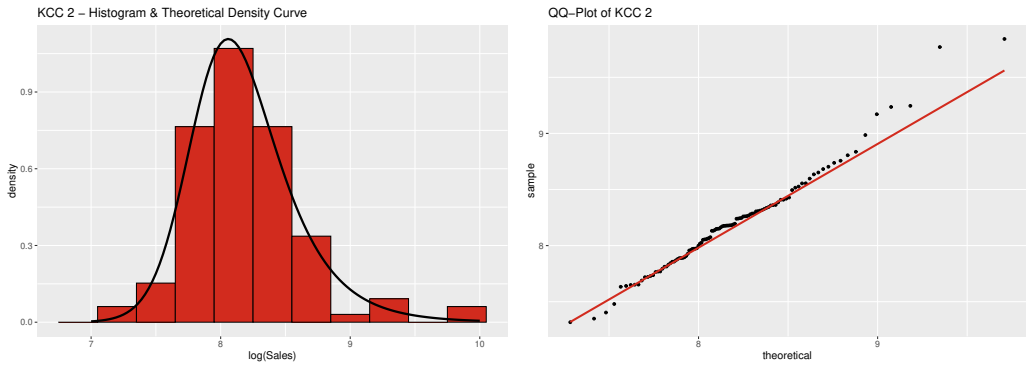
Figure 5.1: Pairwise total markdown percentages of key category clusters

Another important aspect of the modelling part is that usually, continuous responses are implied in the literature. Nevertheless, discrete responses (like in our "real" case) are also

justified when explanatory variables are involved. A detailed explanation can be found in Trivedi and Zimmer [2017].

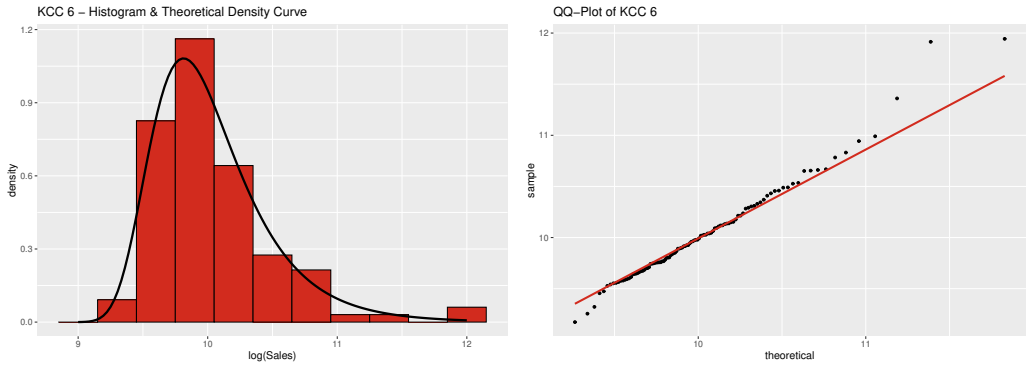
5.1 KCC Marginals

Figure 4.8 is hinting that the marginal distributions come with a noticeable skewness, which are best to take into account. Among a pool of possible parametric distributions, we pick an appropriate one for each marginal. Several distributions would theoretically be justifying the shape of our data, e.g. Weibull, Gamma, Fisk (or Log-logistic), Singh-Maddala or Dagum distribution. After screening those parametric distributions, we find that the Dagum distribution fits all three marginals fairly well [Dagum, 1975]. Although the distribution assumptions are aligned with the empirical distributions of the data, there are still some expected outliers. This is quite visible in Figures 5.2 - 5.4, which depict the marginal empirical distributions along with the theoretical assumptions, especially when looking at the Quantile-Quantile Plots (right figures).



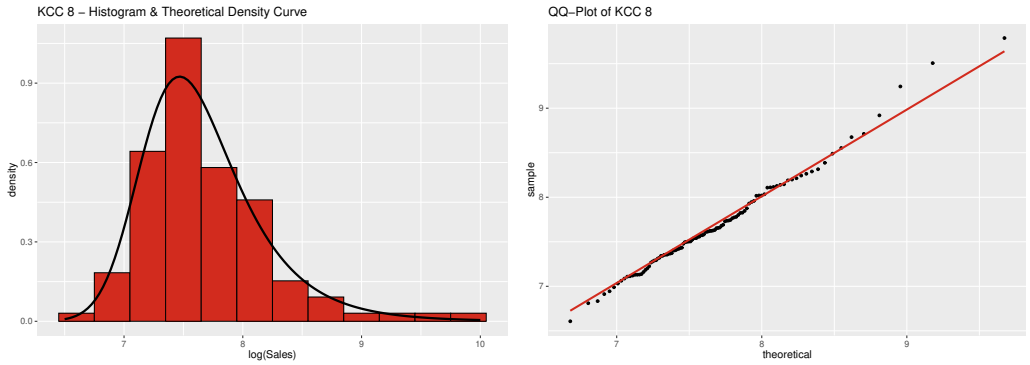
(a) Histogram & theoretical density (b) QQ-Plot

Figure 5.2: Dagum distribution fitted to log-sales of KCC 2; Estimated parameters: scale = 7.83, shape1.a = 29.15, shape2.p = 2.42



(a) Histogram & theoretical density (b) QQ-Plot

Figure 5.3: Dagum distribution fitted to log-sales of KCC 6; Estimated parameters: scale = 8.32, shape1.a = 29, shape2.p = 122.95



(a) Histogram & theoretical density

(b) QQ-Plot

Figure 5.4: Dagum distribution fitted to log-sales of KCC 8; Estimated parameters: scale = 7, shape1.a = 21.03, shape2.p = 4.13

For the approaches in the below Section 5.2 requiring parametric distribution assumptions for the marginals, we will consider the Dagum distribution with parameters specified as in the captions of the above three figures, which represent estimated maximum likelihood values (that is, if they do not have to be estimated over again within a model like e.g. Generalized Additive Models for Loacation, Scale & Shape (GAMLSS)).

5.2 KCC Correlations

Starting the dependence modelling on the key category cluster level, we will be aiming for obtaining separate estimations of pairwise correlations over time, as we only have 3 nodes on this level and manual approximation is feasible on this low number of dimensions. For the next three Subsections (5.2.1, 5.2.2 & 5.2.3), we will compare resulting dependence structures inheriting from different approaches introduced in this Section.

There is an inevitable challenge when it comes to estimating and projecting summary metrics like correlation coefficients onto every observation point; they cannot be tested and thus fully reliable assessments are not guaranteed. Therefore, we will be interpreting the results by comparing several frameworks *under feasibly similar conditions*.

Generalized Additive Models for Conditional Dependence structures (gamCopula)
We will be taking advantage of the framework of *Generalized Additive Models for Conditional Dependence Structures* introduced in Vatter and Chavez-Demoulin [2015] to account for predictor effects on the dependence measure.[10] The R implementation for this

[10] In the scope of this thesis, by dependence we mean concordance and do not distinguish between those terms as opposed to the reference.

framework and its extension to Pair-Copula Constructions [Vatter and Nagler, 2018] can be utilized by the *gamCopula* package [Vatter and Nagler, 2019].

The copula parameter estimates are obtained by maximum likelihood estimation, where each iteration is reformulated as a generalized ridge regression solved using the *mgcv* package [Wood, 2017]. The Fisher scoring method is employed, as the results turn out to be more stable and in greater alignment with other approaches (for more details, see Vatter and Nagler [2018]). We choose Akaike Information Criterion (AIC) for the bivariate copula selection and set the maximal number of Newton-Raphson iterations to 100 so that the algorithm converges for all cases.
The model setup is as follows:

$$\tau(z) = \frac{exp(z) - 1}{exp(z) + 1}, \tag{5.1}$$

where

$$z = \beta_0 + \beta_1 \textit{bf} + \beta_2 \textit{ff} + f(\textit{time}) + f(\textit{total\_markdown\_pct}) \tag{5.2}$$

when using the bivariate (pairwise) pseudo observations z of the marginals as model responses. This allows for flexible estimation of the copula parameter θ and therefore also Kendall's tau (see Section 3.3.2) which is dependent on time. For all three pairs, the selected copula is the Student's t-copula. The t-copula is indeed found to be the most appropriate copula family after testing other families as well (like Normal or Gumbel copula) and is capturing tail dependencies adequately well (see Section 3.2.2), not only in the scope of this approach.
As stated in Equation 3.29, the copula is sufficient for directly deriving the correlation coefficient. The smooth functions for time and total markdown percentage in Equation 5.2 are constructed using thin plate regression splines and by setting 26 and 10 knots respectively, whereas the promotion indicators are set as linear covariates. Results should however be treated with caution, as the algorithm does not converge for all cases.

□

Generalized Joint Regression Models (GJRM)

This approach is based on the framwork of *Generalized Joint Regression Models (GJRM)* and its implementations in R [Marra and Radice, 2016, 2020], where the scope of GAMLSS, which was first introduced by Rigby and Stasinopoulos [2005], is extended by a bivariate copula model. The framework represents the frequentist counterpart of Klein and Kneib [2016] and parameter estimation is achieved within a penalized likelihood framework by

using a trust region algorithm (see Marra and Radice [2020] for more details).

To set similar conditions on the fitting part, we use again a Student's t copula which again provides the most stable results. The marginals are of course assumed to be Dagum distributed (see Section 5.1). As we are primarily interested in the dependence structure, we will not be using covariate effects for any model parameters except for the copula parameter (and consequently the dependence measure). So essentially, besides the constant model parameters, we have

$$\theta = \beta_0 + \beta_1 \textit{bf} + \beta_2 \textit{ff} + f(\textit{time}) + f(\textit{total\_markdown\_pct}) \tag{5.3}$$

with link function for θ

$$\theta = arctanh(\rho) = \frac{1}{2} ln(\frac{1+\rho}{1-\rho}). \tag{5.4}$$

Optionally, we can convert ρ to Kendall's tau via

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

(see Table 3.1).

The parameter θ from Equation 5.3 is in this case being mapped to Pearson's correlation coefficient ρ via Equation 5.4 due to the t-copula. The additive model structure of Equation 5.3 with the copula parameter as response matches Equation 5.2 regarding the linear predictor, i.e. linear effects as well as smooth functions and their respective number of knots. Again, we should treat the outcome with caution as the algorithm fails to converge for all cases. Convergence failure may have various reasons, e.g. low sample size compared to complexity of the model or model misspecification. Unfortunately, not a single model configuration accomplished convergence so we carefully interpret the outcomes in the below Subsections.

□

5.2.1 KCC 2 & KCC 6

We will start the pairwise analysis with inspecting the GJRM approach. As mentioned above, the marginals are specified as Dagum distributed RVs and we choose the Student's t copula (the degrees of freedom do not affect the outcome, as they just serve as

starting values for GAMLSS estimation). All model parameters are set to constant, except the copula parameter (which is defined as in Equation 5.3). These settings also hold for Subsections 5.2.2 and 5.2.3. We can see in the diagnostic plots of Figure 5.5 that this type of model fitting returns quite successful results.

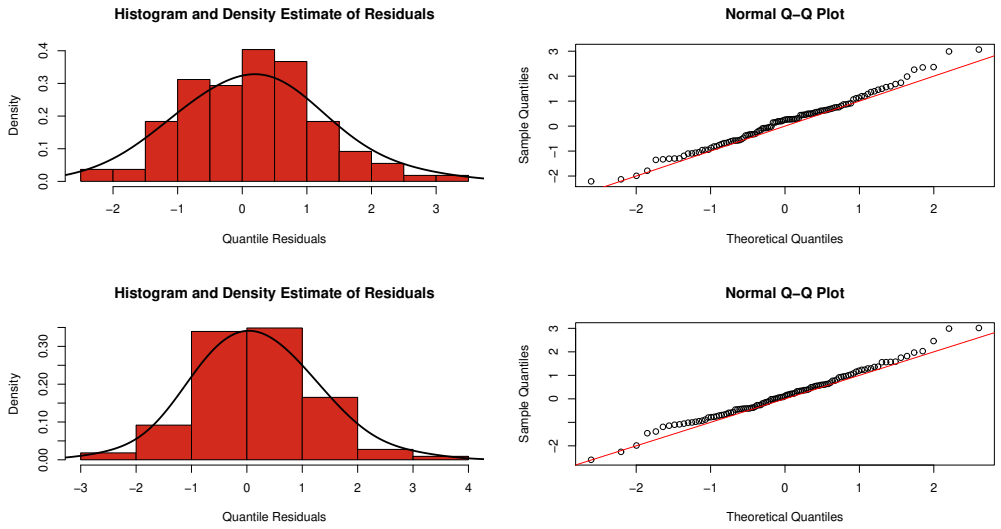


Figure 5.5: Estimation diagnostics for the response marginals KCC 2 & KCC 6; GJRM approach

Regarding the fitting of the copula parameter ρ, time-varying estimation is achieved. The red line in Figure 5.6 shows the estimated time-dependent sequence of $\hat{\rho}$, where an instant positive conclusion is that the outcome is very close to the gamCopula approach. This indicates that the model setup of these two approaches both agree for the most part on the dependence structure. (Confidence intervals of the lines are not shown to retain clarity in the figure).

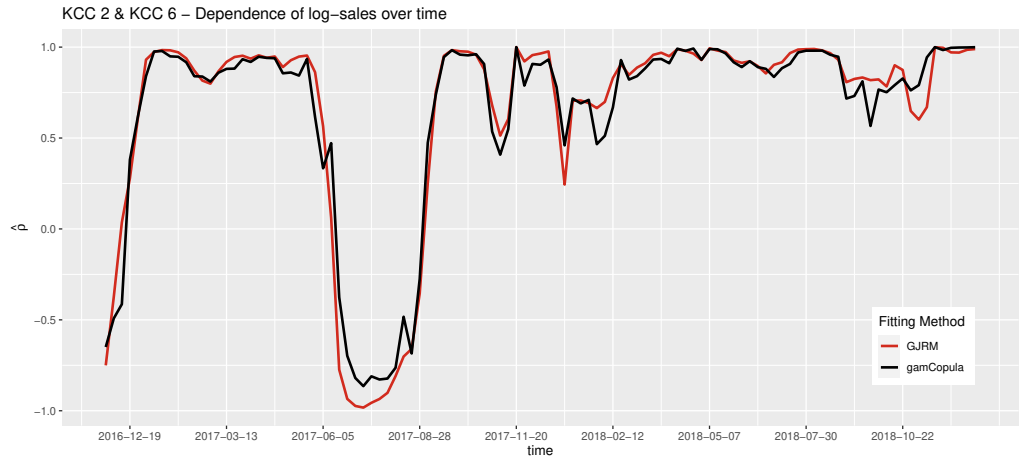


Figure 5.6: Estimated time-varying Pearson's correlation coefficients for the pair KCC 2 & KCC 6

By large, the dependence is constantly at a very high positive level, with a noticeable turn in high negativity during the summer months of 2017. The gamCopula approach seems to pick up slightly more regional fluctuations than the GJRM approach.

5.2.2 KCC 2 & KCC 8

On this KCC pair, the GJRM approach is not as favorable as in the pair KCC 2 & KCC 6, though it might be considered acceptable (Figure 5.7).

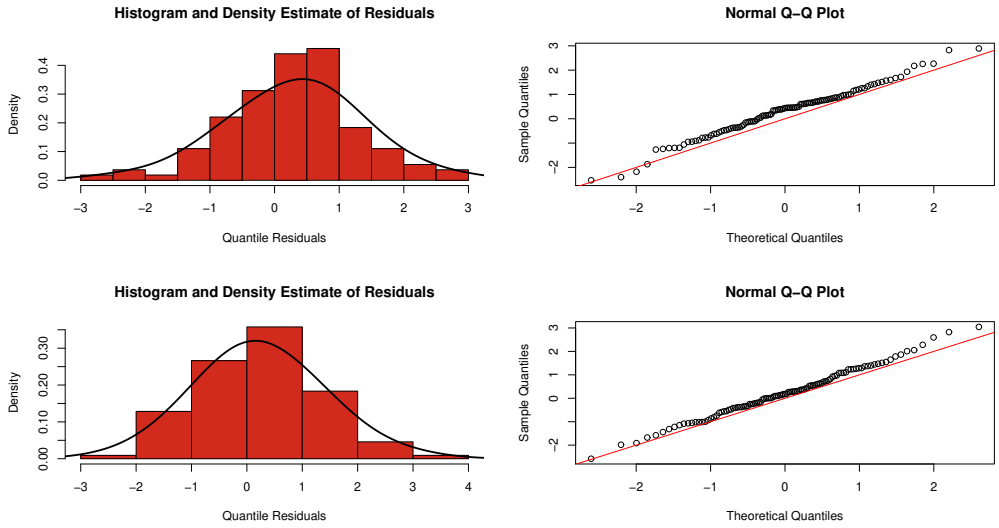


Figure 5.7: Estimation diagnostics for the response marginals KCC 2 & KCC 8; GJRM approach

This circumstance is reflected within the comparison of the "gamCopula" estimated dependence measures. The overall similarity between the two lines in Figure 5.8, which depict the time-dependent correlation parameters of this pair, is given. However, the details reveal that there are severe differences between the estimation methods. As the gamCopula approach yields converge in contrast to the GJRM method, it might be preferential. Nevertheless, the results should be treated with adequate caution.

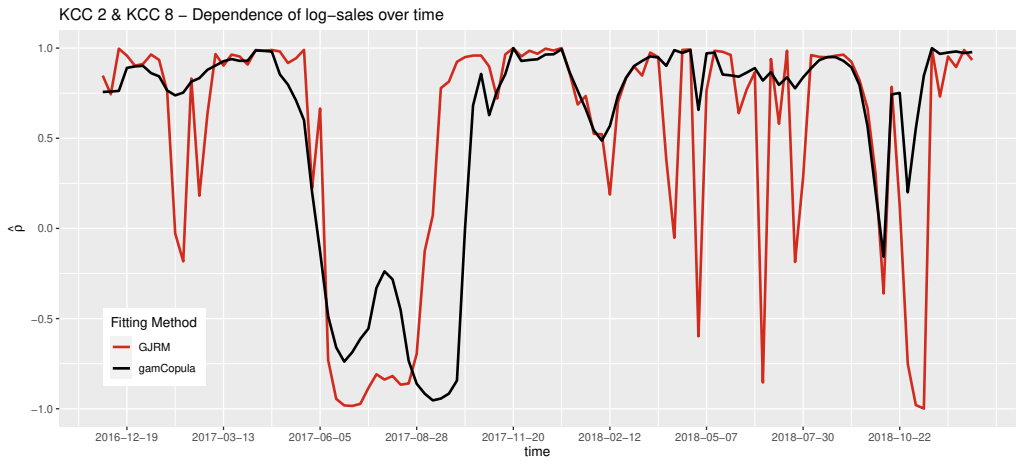


Figure 5.8: Estimated time-varying Pearson's correlation coefficients for the pair KCC 2 & KCC 8

The stretched sign change of the correlation over time happens, just like in the previous pair, from June to August 2017. The curves here are surprisingly different in terms of their stability. The gamCopula method seems to have a more stable course compared to the GJRM approach. The red line in Figure 5.8 appears very sensitive in very short time intervals.

5.2.3 KCC 6 & KCC 8

The GJRM fit on this pair seems to work just fine, which is underlined when observing Figure 5.9. As for the estimation of the dependence, we obtain a similar scheme as in the pair KCC 2 & KCC 8. The main course of the correlation coefficient is common for both approaches, however the gamCopula approach seems to reveal heavier fluctuations. As always, it shall be stressed that these conclusions should be taken into account with extra caution, as both algorithms converge.

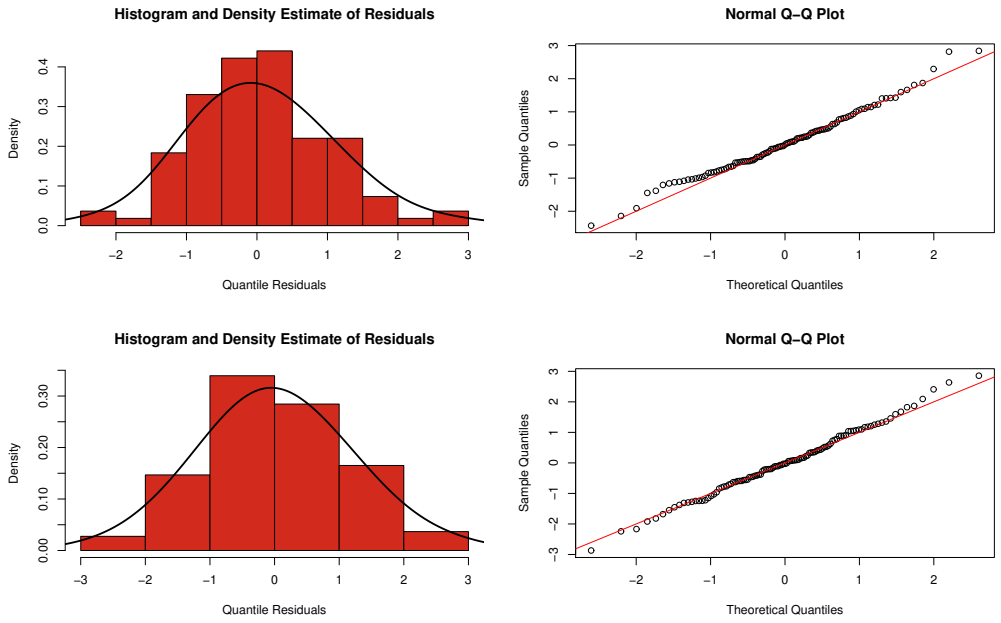


Figure 5.9: Estimation diagnostics for the response marginals KCC 6 & KCC 8; GJRM approach

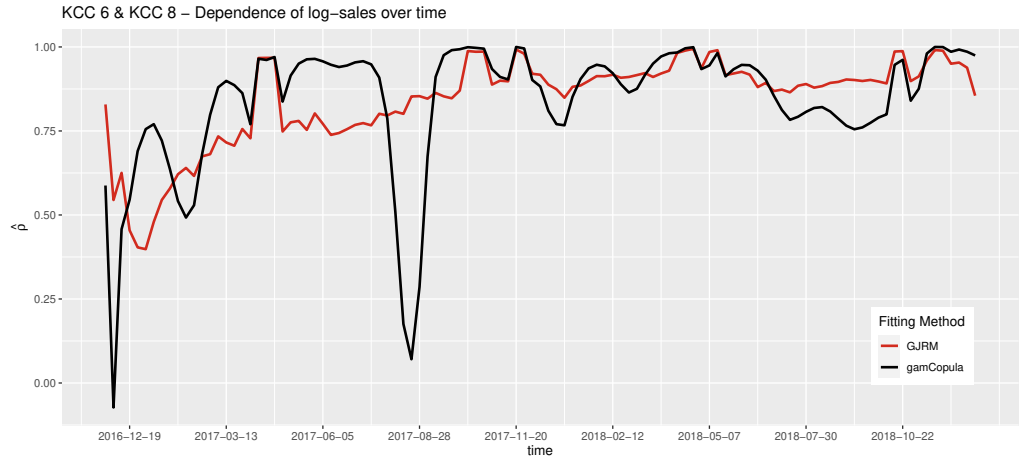


Figure 5.10: Estimated time-varying Pearson's correlation coefficients for the pair KCC 6 & KCC 8

In both correlations in Figure 5.10 show an overall increasing trend from the start, which stabilizes somewhere in September of 2017. Both curves return positive values throughout the entire time span (except for the very start of the black one). Also, wider and heavier fluctuations occur in the gamCopula approach as opposed to the GJRM method, which looks much smoother.

5.3 Article Correlations

In this subsection we attempt to discover a dependence structure on the individual article unit sales. To do so, we will first delimit our data on article level (Subsection 5.3.1) to

achieve lower dimensionality.

MAYBE WRITE STH ON USAGE OF VINES HERE..

5.3.1 Data Delimitation

On the lowest product hierarchy level, the number of unique articles is quite large for models to handle. In Table 5.1, we can see that our key category clusters of interest, 2, 6 and 8, have 1587, 9746 and 2012 articles respectively. Although we only work on those KCCs, even advanced techniques like vine copulas are struggling to handle such dimensions.

Key Category Cluster	1	2	3	4	5	6	7	8	9
Unique Articles	14	1,587	4,151	278	1,713	9,746	374	2,012	6,328

Table 5.1: Articles per Key Category Cluster

We have to take into consideration different models for each group. On top, further data delimitation shall reduce the dimensionality even further. Referring back to Table 1.1, we know that our last week of observation is "2018-12-24". We will restrict ourselves to articles that were not taken off the assortment before that date because they will most probably not be put back for sale.[11] After setting these boundaries, we arrive at a considerably smaller magnitude shown in Table 5.2.

Key Category Cluster	2	6	8
Unique Articles	39	256	58

Table 5.2: Articles per Key Category Cluster after data delimitation

[11] We neglect the possibility of an item to return in the future.

6 Conclusion

List of Abbreviations

AIC Akaike Information Criterion

BS Business Segment

CDF Cumulative Distribution Function

d.o.f. Degrees of Freedom

GAM Generalized Additive Model

GAMLSS Generalized Additive Models for Loacation, Scale & Shape

GAMM Generalized Additive Mixed Model

GLM Generalized Linear Model

GLMM Generalized Linear Mixed Model

GJRM Generalized Joint Regression Models

KC Key Category

KCC Key Category Cluster

LM Linear Regression Model

LMM Linear Mixed Model

PDF Probability Density Function

RV Random Variable

STAR Structured Additive Regression Model

w.l.o.g. without loss of generality

List of Figures

List of Tables

References

[adidas 70 years 2019] adidas 70 years: *adidas celebrates its 70th anniversary and the opening of the ARENA building.* 2019. – https://www.adidas-group.com/en/media/news-archive/press-releases/2019/adidas-celebrates-70th-anniversary, Last accessed on 2020-04-01

[adidas-group.com] adidas-group.com: *adidas-group*. – https://www.adidas-group.com/en/, Last accessed on 2020-04-01

[adidas.com media-center] adidas.com media-center: *adidas media-center. Pictures and Videos*. – https://www.adidas-group.com/en/media/media-center/, Last accessed on 2020-04-01

[Dagum 1975] Dagum, Camilo: A model of income distribution and the conditions of existence of moments of finite order. In: *Bulletin of the International Statistical Institute* 46 (1975), S. 199–205

[Embrechts et al. 2001] Embrechts, Paul ; Lindskog, Filip ; McNeil, Alexander: Modelling dependence with copulas. In: *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich* 14 (2001)

[Fahrmeir et al. 2003] Fahrmeir, L. ; Kneib, T. ; Lang, S. ; Marx, B.: *Regression; Models, Methods and Applications. 2013*. 2003

[Hastie and Tibshirani 1986] Hastie, Trevor ; Tibshirani, Robert: Generalized Additive Models. In: *Statist. Sci.* 1 (1986), 08, Nr. 3, S. 297–310. – URL https://doi.org/10.1214/ss/1177013604

[Hastie and Tibshirani 1990] Hastie, Trevor J. ; Tibshirani, Robert J.: Generalized additive models, volume 43 of. In: *Monographs on statistics and applied probability* 15 (1990)

[Hofert et al. 2019] Hofert, Marius ; Kojadinovic, Ivan ; Mächler, Martin ; Yan, Jun: *Elements of copula modeling with R*. Springer, 2019

[Klein and Kneib 2016] Klein, Nadja ; Kneib, Thomas: Simultaneous inference in structured additive conditional copula regression models: a unifying Bayesian approach. In: *Statistics and Computing* 26 (2016), Nr. 4, S. 841–860

[Marra and Radice 2016] Marra, G. ; Radice, R.: A Bivariate Copula Additive Model for Location, Scale and Shape. Cornell University Library (2017). In: *arXiv preprint arxiv:1605.07521* (2016)

[Marra and Radice 2020] Marra, Giampiero ; Radice, Rosalba: GJRM-package: Generalised Joint Regression Modelling. (2020)

[McNeil et al. 2015] McNeil, Alexander J. ; Frey, Rüdiger ; Embrechts, Paul: *Quantitative risk management: concepts, techniques and tools-revised edition*. Chap. Copulas and Dependence, Princeton university press, 2015

[Patton 2006] Patton, Andrew J.: Modelling asymmetric exchange rate dependence. In: *International economic review* 47 (2006), Nr. 2, S. 527–556

[R Core Team 2018] R Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (Veranst.), 2018. – URL `https://www.R-project.org/`

[Rigby and Stasinopoulos 2005] Rigby, Robert A. ; Stasinopoulos, D. M.: Generalized additive models for location, scale and shape. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54 (2005), Nr. 3, S. 507–554

[Ruppert and Matteson 2015] Ruppert, David ; Matteson, David S.: *Copulas*. S. 183–215. In: *Statistics and Data Analysis for Financial Engineering: with R examples*. New York, NY : Springer New York, 2015. – URL `https://doi.org/10.1007/978-1-4939-2614-5_8`. – ISBN 978-1-4939-2614-5

[Schmidt 2007] Schmidt, Thorsten: Coping with copulas. In: *Copulas-From theory to application in finance* (2007), S. 3–34

[Sklar 1959] Sklar, M.: Fonctions de repartition an dimensions et leurs marges. In: *Publ. inst. statist. univ. Paris* 8 (1959), S. 229–231

[stanfordphd] stanfordphd: *Copula*. – `https://stanfordphd.com/Copula.html`, Last accessed on 2020-04-27

[Trivedi and Zimmer 2017] Trivedi, Pravin ; Zimmer, David: A note on identification of bivariate copulas for discrete count data. In: *Econometrics* 5 (2017), Nr. 1, S. 10

[Vatter and Chavez-Demoulin 2015] Vatter, Thibault ; Chavez-Demoulin, Valérie: Generalized additive models for conditional dependence structures. In: *Journal of Multivariate Analysis* 141 (2015), S. 147–167

[Vatter and Nagler 2018] Vatter, Thibault ; Nagler, Thomas: Generalized additive models for pair-copula constructions. In: *Journal of Computational and Graphical Statistics* 27 (2018), Nr. 4, S. 715–727

[Vatter and Nagler 2019] Vatter, Thibault ; Nagler, Thomas: gamCopula-package: Generalized Additive Models for Bivariate Conditional... (2019)

[Wood 2017] Wood, Simon N.: *Generalized additive models: an introduction with R*. CRC press, 2017