GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Master Thesis

Multivariate Modelling of the Dependence 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 August 25, 2020

Processing time of 20 weeks

Confidentiality Clause

This Master thesis contains confidential data of adidas AG.
This work may only be made available to the first and second reviewers and authorized members of the board of examiners. Any publication and duplication of this Master thesis - even in part - is prohibited.
An inspection of this work by third parties requires the expressed permission of the author and adidas AG.
This confidentiality clause expires automatically after five years.

Acknowledgments

I would like express my deepest gratitude towards Dr. Alexander März, who guided me throughout this thesis with his expertise and has always been finding the time to help during these extraordinary times. I would also like to thank the entire adidas Data Science & AI team for contributing a great deal to my learning journey and giving me the opportunity to conduct this thesis.

I am also sincerely grateful towards Prof. Dr. Thomas Kneib for supervising my Master thesis and moreover for leading the study program of Applied Statistics and passing on his knowledge to his students in the best way possible.

Last but not least, I would like to thank my fellow students Patrick Neff and Malte Lehna who were playing a key role in my personal growth and with whom I had an unforgettable time in Göttingen.

Ευχαριστώ θερμά τους γονείς μου για την άνευ όρων υποστήριξη. Σας αγαπώ!

Abstract

While transactional data storage of the adidas eCom website is constantly increasing in size, analytical products generating data-driven insights became a vital part of the business. For this thesis, two years of weekly data were provided in order to discover dependence structures between demand quantities of sports and fashion articles.

Since eCom sales data are fairly sparse considering that there are tens of thousands of articles available online, targeting directly the articles is quite a challenge. Therefore, the focus will firstly lie in modelling dependence structures of summarized groups of articles called "key category clusters", as they represent aggregated sport/fashion categories at adidas. Demand quantities of three specific clusters undergo thorough analysis along with critical features such as promotion activities and markdowns. To model temporal dependencies between the clusters, a two-step approach is carried out where we first have the demand quantities of the marginal responses fitted flexibly with respect to the mentioned features over time, using GAMLSS with exponentially modified Gaussian families. Secondly, the quantile residuals from the first step were used as normally distributed responses to obtain the pairwise dependence structures of the clusters with the help of Bivariate Copula Additive Models for Location, Scale and Shape. Both steps were carefully realized; dependence structures in form of time-varying correlation coefficients over 109 weeks were discovered and validity of the results were reviewed.

As for the individual articles, some central data sufficiency issues were pointed out, highlighting the sparsity of the data. An exercise of dynamic Bayesian networks, which have the potential to detect causal effects between entities over time, was applied to distinct articles in one of the available key category clusters to demonstrate some of the advantages of these models against the data limitations. The thesis wraps up proposing such frameworks for future research.

Keywords: *GAMLSS - Structured Additive Conditional Copula Regression - Generalized Joint Regression Models - Dynamic Bayesian Networks*

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. The purpose of this thesis is to capture dependence structures between article demand quantities over time, applied to transactional eCom data. Primary modelling focus will lie on special clusters of article demand quantities, where we will analyze time-varying correlation structures. Another latent effect is sales cannibalization between newer and older articles or articles of similar traits. To identify such causal effects in a quantifiable way, frameworks applied to individual articles will be outlined afterwards serving as a suggestion for further investigation.

For 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 will be introduced to get familiar with the data. Chapter 2 introduces some notions on statistical methodologies relevant for this thesis. In Chapter 3 we will have a closer look into theoretical aspects of the copula framework, which comprises the major ingredient of this thesis' modelling part. In Chapter 4 some exploratory data analysis is performed to delve into the patterns of article sales and to step-wise investigate some hierarchical properties of the data. Chapter 5 analyses the results of modelling conditional copulas to clusters of the data based on some preliminary work on the marginal distributions via GAMLSS. Thereafter, some diagnostics will be examined. Moreover, in Section 5.3 a sketch for inferring causal effects between demand quantities of individual articles using dynamic Bayesian networks will be proposed for research beyond this thesis. A final conclusion will be summarized in Chapter 6, where we point out the main findings of this thesis and further potential study directions.

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

1.1 adidas

The *"Dassler Brothers Shoe Factory"* (German: *"Gebrüder Dassler Schuhfabrik"*), which was led by *Adolf Dassler* (aka *Adi Dassler*) and his brother Rudolph, was dissolved in 1947. 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 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 [adidas-group.com profile] and is nowadays sponsoring a vast range of athletes, artists and organizations across the globe (e.g. the FIFA World Cup™).

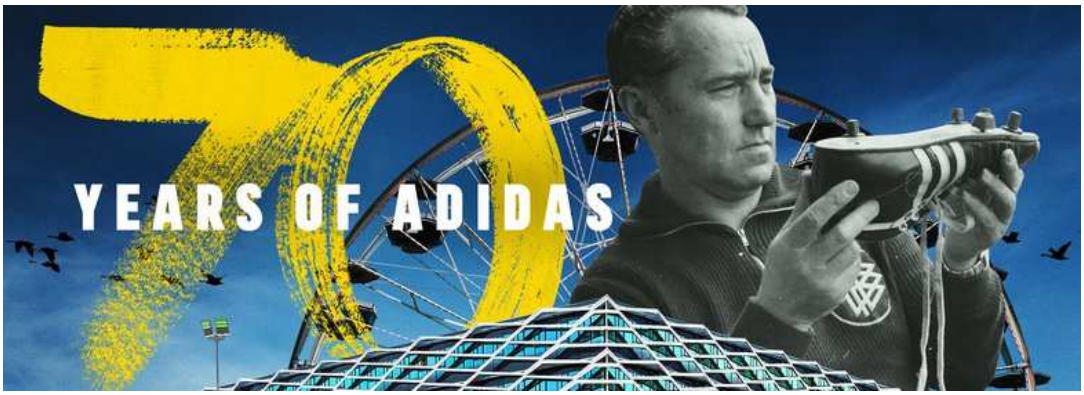


Figure 1.2: adidas celebrates its 70th anniversary and the opening of the ARENA building [adidas 70 years, 2019]

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 numerous DNA use cases.

1.2 Data Sources

Throughout each season, transactional data were 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 were 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

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

attributes by the article ID, this translates to a dataset of 587,127 instances including 26,203 distinct articles and over 30 variables.

2 Statistical Theory & Methods

This chapter introduces some statistical methods used during the conduction of this thesis. Basic notations regarding mathematical foundations of statistics (such as linear algebra, probability theory, hypothesis testing etc) are skipped. Theoretical aspects regarding copulas and dependence structures will be introduced separately in Chapter 3.

2.1 Shapiro-Wilk Test of Normality

The *Shapiro-Wilk test of normality* was first introduced in Shapiro and Wilk [1965] and is a method used to test the hypothesis whether a sample of observations $\boldsymbol{x} = x_1, \ldots, x_n$ was drawn from a normal distribution, i.e.

$$H_0 : \boldsymbol{x} \sim \mathcal{N}(\mu, \sigma) \quad \text{vs} \quad H_1 : \boldsymbol{x} \nsim \mathcal{N}(\mu, \sigma)$$

and the test statistic is

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

where the coefficients a_i are given by

$$(a_1, \ldots, a_n) = \frac{m^\top V^{-1}}{(m^\top V^{-1} V^{-1} m)^{1/2}}.$$

The expected values of the order statistics of independent and identically distributed (iid) random variables (RV), which are sampled from a standard normal distribution $\mathcal{N}(0, 1)$, are represented by the vector $m = (m_1, \ldots, m_n)^\top$ and V is the covariance matrix of those order statistics.

2.2 Generalized Linear Models

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \ldots, n \tag{2.1}$$

which in matrix notation can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.2}$$

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

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.3}$$

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

$$E(y_i) = h(\eta_i), \tag{2.4}$$

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

$$g = h^{-1}, \tag{2.5}$$

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

2.3 Additive Models

Additive Models [Fahrmeir et al., 2003] expand models with just a linear predictor

$$\eta_i^{lin} = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} \tag{2.6}$$

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

$$\sum_{i=1}^{n} f_1(z_{i1}) = \ldots = \sum_{i=1}^{n} f_q(z_{iq}) = 0. \tag{2.11}$$

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

$$\boldsymbol{Z}_0 = \begin{pmatrix} \mathbf{1}_1 & & & \mathbf{0} \\ & \ddots & & \\ & & \mathbf{1}_i & \\ & & & \ddots \\ & & & \mathbf{1}_m \end{pmatrix} \qquad \boldsymbol{Z}_1 = \begin{pmatrix} \boldsymbol{x}_1 & & & \mathbf{0} \\ & \ddots & & \\ & & \boldsymbol{x}_i & \\ & & & \ddots \\ & & & \boldsymbol{x}_m \end{pmatrix}. \tag{2.12}$$

More details and technicalities regarding mixed effects in additive models can be found in Fahrmeir et al. [2003].

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.

2.4 Generalized Additive Models for Location, Scale and Shape

Generalized Additive Models for Loacation, Scale & Shape (GAMLSS) [Rigby and Stasinopoulos, 2001, 2005] are a framework which surpass the limitations that come with GLMs and GAMs. Particularly, in GAMLSS the assumption that the response variable y belongs to a distribution of the exponential family is relaxed and a more general distribution family is permissible, including highly skewed and/or kurtotic distributions. In addition, other parameters besides the mean (or location) of the response's distribution can be modelled flexibly incorporating linear, non-linear and/or additive functions of covariates as well as random effects. By modelling the scale and shape of the variable, the issue of heteroscedasticity in the response is being adressed. Two algorithms can be used to fit the models, namely the CG and the RS algorithms, which can be looked upon in more detail in Rigby and Stasinopoulos [2005].

Independent observations y_i for $i = 1, 2, \ldots, n$ with probability (density) function $f\left(y_i \mid \boldsymbol{\theta}^i\right)$, where $\boldsymbol{\theta}^i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{ip})$ is assumed. Without loss of generality, the number of parameters p is at most 4 and the parameters are denoted as $(\mu_i, \sigma_i, \nu_i, \tau_i)$, where the parameter μ_i is the location parameter, σ_i is the scale parameter, and ν_i and τ_i are characterized as shape parameters.[3] Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ be the vector of the response variable and $g_k(.)$, $k = 1, 2, 3, 4$ be known monotonic link functions. Then

$$
\begin{aligned}
g_1(\boldsymbol{\mu}) &= \eta_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} h_{j1}\left(\boldsymbol{x}_{j1}\right) \\
g_2(\boldsymbol{\sigma}) &= \eta_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} h_{j2}\left(\boldsymbol{x}_{j2}\right) \\
g_3(\boldsymbol{\nu}) &= \eta_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} h_{j3}\left(\boldsymbol{x}_{j3}\right) \\
g_4(\boldsymbol{\tau}) &= \eta_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} h_{j4}\left(\boldsymbol{x}_{j4}\right),
\end{aligned}
\tag{2.13}
$$

where $\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}$ and $\boldsymbol{\eta}_k$ and $\boldsymbol{x}_{jk}$, for $j = 1, \ldots, J_k$ and $k = 1, 2, 3, 4$ are vectors of length n. The explanatory variable X_{jk} evaluated at x_{jk} is described by the additive function h_{jk}. X_k are fixed design matrices and β_k are the parameter vectors.[4]
Note that the model 2.13, also known as *semi-parametric GAMLSS* model, can be extended to allow random effect terms to be included for any parameter (more details can be found in the mentioned literature for this section).

[3] The model can be applied to distributions of any kind of parametric nature.

[4] According to Stasinopoulos et al. [2007], in typical applications a constant is often adequate for each of the two shape parameters.

2.5 Univariate & Multivariate Time Series

For a compact description of univariate & multivariate time series, the reference literature will be from *"Bayesian Networks in R - with Applications in Systems Biology"* [Nagarajan et al., 2013] to maintain the congruity with later appearing terms derived from the same source.

A sequence of Random Variables (RVs)

$$\{X(t)\} = \{\ldots, X(t-1), X(t), X(t+1), \ldots\} \tag{2.14}$$

measured at consecutive time-points with uniform time intervals is called a *Univariate Time Series (UTS)*. The time series is considered *covariance stationary*[5] if its first two moments are invariant over time, i.e.

$$\mathrm{E}(X(t)) = \mu, \quad \forall t \qquad \text{and} \tag{2.15}$$

$$\mathrm{COV}(X(t), X(t-i)) = \mathrm{E}((X(t)-\mu)(X(t-i)-\mu)) = \gamma_i, \quad \forall t, i. \tag{2.16}$$

A stationary UTS can be modelled as an *Auto-Regressive (AR)* process, where the value at time t can be written as a linear combination of its lagged values from previous time-points $X(t-i)$ for $i = 1, \ldots, p$:

$$X(t) = a_1 X(t-1) + \cdots + a_i X(t-i) + \cdots + a_p X(t-p) + b + \varepsilon(t), \quad \forall t \geqslant p, \tag{2.17}$$

where $X(t)$ is the RV observed at time t, p is the *lag* or *order* of the time series, $a_i \in \mathbb{R}$ with $i = 1, \ldots, p$ are the coefficients of the RVs observed at the previous p time-points $t-1, t-2, \ldots, t-p$, $b \in \mathbb{R}$ is the intercept and $\varepsilon(t)$ is a Gaussian white noise, i.e. $\boldsymbol{\varepsilon}(t) \sim N\left(0, \sigma^2\right)$.

A sequence of multivariate RVs measured at consecutive time-points is called a *Multivariate Time Series (MTS)*. MTS are commonly used to assess the associations between multiple individuals over time and can be modelled as Vector Auto-Regressive (VAR) processes. In a VAR(p) of order p, the variables observed at any time-point $t \geqslant p$ should satisfy

$$X(t) = A_1 X(t-1) + \cdots + A_i X(t-i) + \cdots + A_p X(t-p) + B + \varepsilon(t), \tag{2.18}$$

where $X(t) = (X_i(t))$ with $i = 1, \ldots, k$ is the vector of k variables observed at time t, A_i with $i = 1, \ldots, p$ are coefficient matrices of dimensions $k \times k$, B is an intercept vector of

[5] For brevity, we will use the term "stationary".

dimension k and $\varepsilon(t)$ is a white noise vector of dimension k with $\mathrm{E}(\boldsymbol{\varepsilon}(t)) = 0$ and time-invariant positive definite covariance matrix $\Sigma = \mathrm{COV}(\varepsilon(t))$. Similar to an AR process, a VAR(p) process assumes a linear correlation structure between the k variables observed at the t time-points and the k variables observed at the p previous time-points.

2.6 Elements of Graph Theory

The source literature for this section can be found in the first chapter of Nagarajan et al. [2013], where elements of graph theory are introduced.

A graph $G = (\boldsymbol{V}, A)$ consist of a non-empty set of *nodes* or *vertices* $\boldsymbol{V}$ and a finite set A of pairs of vertices calles *arcs*, *links*, or *edges*. Each arc $a = (u, v)$ comprises of as an ordered or unordered pair of nodes, which are *connected by* and *incident* on the arc a or *adjacent* to each other an therefore u and v can also be called *neighbors*.

For the purpose of this thesis, we will limit the theory to ordered node pairs. If (u, v) is an ordered pair, then u is said to be the *tail* of the arc a and v the head of a, i.e. a is called *directed* from u to v. The usual representation is $(u \to v)$. It is also said that arc a *leaves* or is *outgoing* for u and that it *enters* or is *incoming* for v. If a graph G consist of directed arcs only and has no cycles[6], it is called a *Directed Acyclic Graph (DAG)*.
In the example of Figure 2.1, the node set is $\boldsymbol{V} = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}\}$ and the graph is characterized by the arc set $A = \{(A \to B), (C \to A), (D \to B), (C \to D), (C \to E)\}$. As arcs are directed, $C \to D$ and $D \to C$ are different and due to acyclicity it is impossible for both arcs to be in the graph because there can be at most one arc between each pair of nodes. Hence, $C \to D \in A$ while $D \to E \notin A$.

[6] i.e. No node can be traversed back to itself.

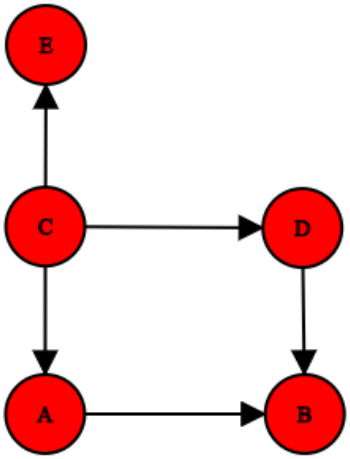


Figure 2.1: Example of a directed acyclic graph

A sequence of arcs connecting two nodes (*end-nodes*) is called a *path* and is denoted with the sequence of nodes $(v_1, v_2, \ldots, v_n)$ incident on those arcs. A path between two end-nodes is assumed to be unique. If node v_i precedes node v_j, there can be no arc from v_j to v_i. In this case, v_i is called an *ancestor* of v_j and v_j is called a *descendant* of v_i. If these two nodes are adjacent, v_i is called a parent of v_j and v_j is called a child of v_i.

2.7 Bayesian Networks

Let $\boldsymbol{X} = \{X_1, X_2, \ldots, X_p\}$ be a set of RVs. A *Bayesian network* is a DAG $G = (\boldsymbol{V}, A)$ where each node $v_i \in \boldsymbol{V}$ corresponds to a RV X_i, $i \in \{1, \ldots, p\}$ and each arc $a = (u, v)$ corresponds to a conditional dependence between two RVs. The *Markov property*, i.e. a node is conditionally independent of its non-descendants given its parents, enables the representation of the joint probability distribution of the RVs in $\boldsymbol{X}$ as a product of conditional distributions [Nagarajan et al., 2013], allowing simplification of the joint distribution as a result of the *chain rule* and reducing the conditional part to the parents of X_i. For discrete RVs, the joint probability is therefore given by

$$\mathrm{P}_{\mathbf{X}}(\boldsymbol{X}) = \prod_{i=1}^{p} \mathrm{P}_{X_i}\left(X_i \mid X_1, \ldots, X_{i-1}\right) = \prod_{i=1}^{p} \mathrm{P}_{X_i}\left(X_i \mid \Pi_{X_i}\right), \tag{2.19}$$

where Π_{X_i} is the set of the parents of X_i. For continuous RVs, we can write the joint density of $f_{\boldsymbol{X}}$ as

$$f_{\boldsymbol{X}}(\boldsymbol{X}) = \prod_{i=1}^{p} f_{X_i}\left(X_i \mid X_1, \ldots, X_{i-1}\right) = \prod_{i=1}^{p} f_{X_i}\left(X_i \mid \Pi_{X_i}\right). \tag{2.20}$$

In real-world settings, entities often represent variables which vary over time. *Dynamic Bayesian Networks (DBNs)* extend the framework of static Bayesian networks to model associations arising from temporal dynamics between such entities [Nagarajan et al., 2013]. Each RV in a DBN is represented by several nodes across time-points. In the DAG resulting from a DBN, arcs can be drawn between variables across successive time-points (or at the same time-point,[7] as long as no arc is enters an ancestor node). The arcs in the DAG describe exactly the conditional dependencies between any pair of variables given the past variables (or variables at the same time stage). Figure 2.2 represents a graphical example of such a DBN.

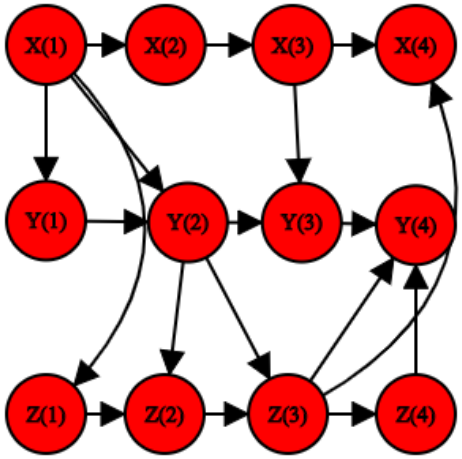


Figure 2.2: Graphical representation of a time-varying dynamic Bayesian network of three random variables (X, Y and Z) with four time periods

Dependence relationships in DBNs are often represented by a VAR process as in Equation 2.18. Assuming a VAR(1) process with k variables, each variable X_i, $i = 1, \ldots, k$ satisfies

$$X_i(t) = \sum_{j=1}^{k} a_{ij} X_j(t-1) + b_i + \varepsilon_i(t) \quad \text{where} \quad \varepsilon_i(t) \sim N\left(0, \sigma_i(t)\right), \tag{2.21}$$

where all arcs are defined between two consecutive time-points. The set of non-zero coefficients a_{ij} in an auto-regressive matrix A define the arc set, meaning that if element a_{ij} is non-zero, the network includes an arc from X_i to X_j. Repeated measurements can be used to perform linear regression, however the ordinary least square estimates of the regression coefficients a_{ij} and b_i can be computed only when the number of time-points $n \gg k$.

[7] In Nagarajan et al. [2013]) conditional dependencies can occur only between successive time-points. However, this assumption is relaxed here due to practical alignments coming up in Section 5.3.

3 Copulas & Dependence Structures

Multivariate distributions consist of the marginal distributions and the dependence structure between those margins. 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 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 margins 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 margins.

Invariance Principal

Suppose the RVs $X_1, \ldots, X_d$ have continuous margins 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 margins 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

Without loss of generality, 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

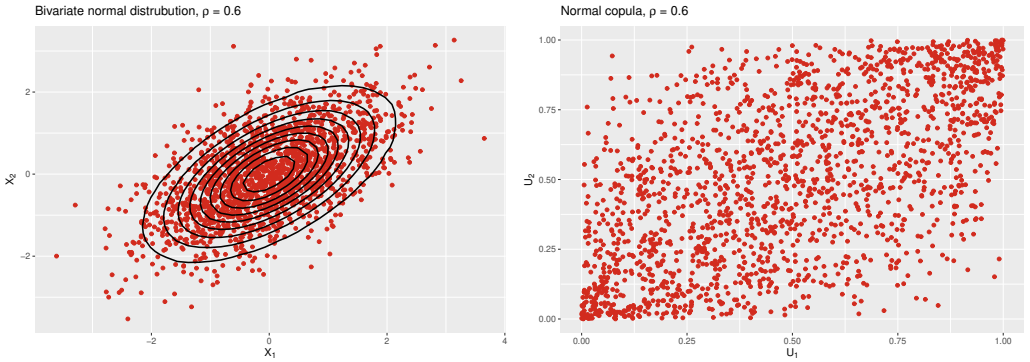
- the independence copula Π if $\rho = 0$,
- the comonotonicity copula M if $\rho = 1$ and
- the countermonotonicity copula W if $\rho = -1$

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 margins

t-Copula

Consider without loss of generality $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{0}, \boldsymbol{P})$ (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.).

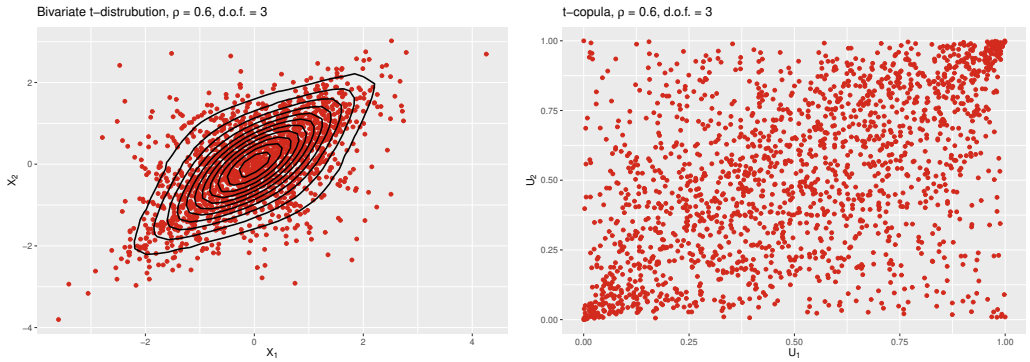
For the *bivariate t-copula* ($d = 2$), the special cases are the same as for the Gaussian copula except that $d = 0$ does not yield the independence copula (unless $\nu \to \infty$ in which case $C_{\nu,\rho}^{t} = C_{\rho}^{Ga}$).

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 margins

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.[8] The advantage of such copulas lies in the fact that they interpolate between certain fundamental dependence structures.

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

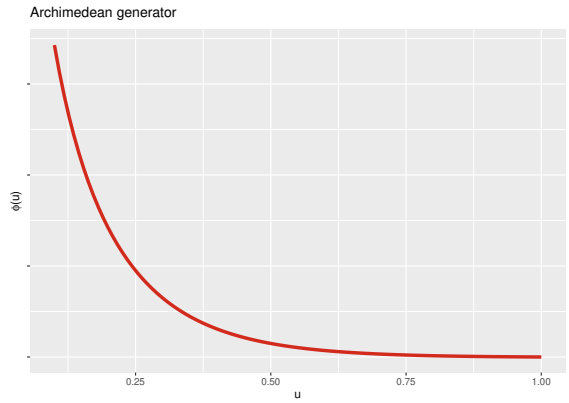


Figure 3.3: Shape of a generator function

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

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

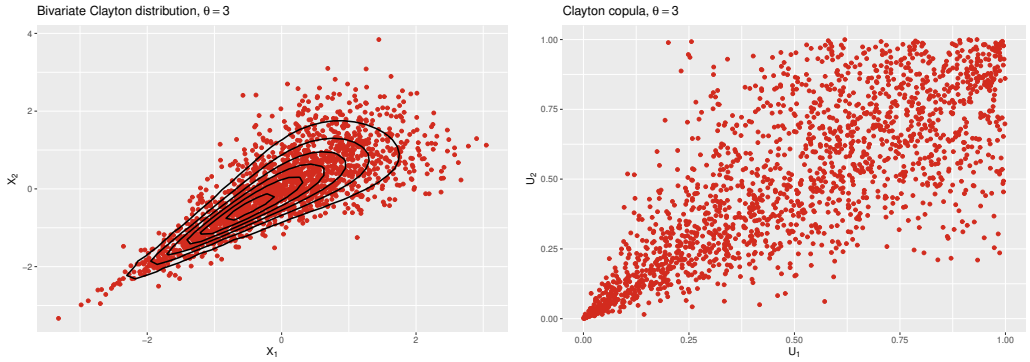
where $\theta \in [-1, \infty)\backslash\{0\}$.

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 margins

Gumbel Copula

If the generator takes on the form

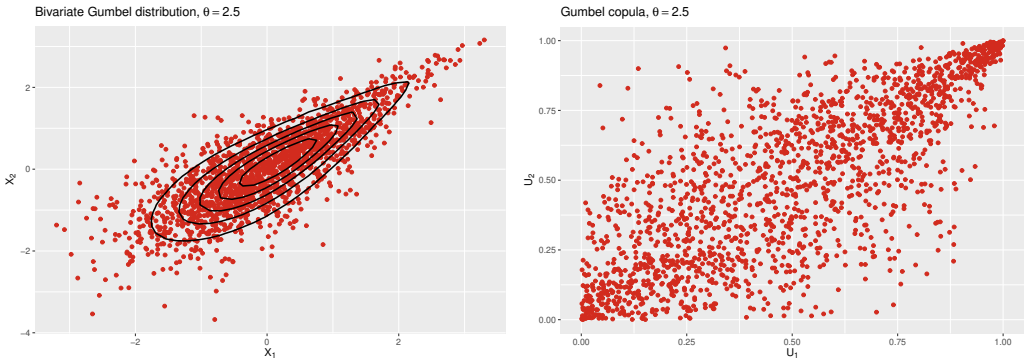
$$\phi_{Gu}(u) = (-\ln u)^{\theta}, \quad \theta \in [1, \infty), \tag{3.20}$$

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 margins

Frank Copula

If the generator takes on the form

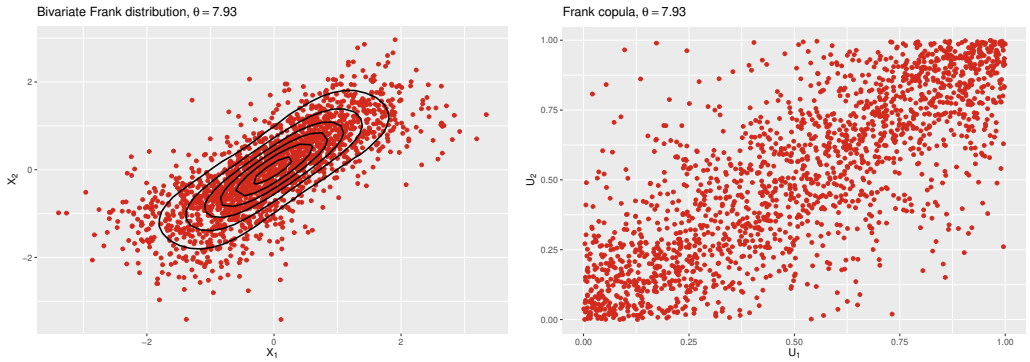
$$-\ln\left(\frac{\mathrm{e}^{-\theta u} - 1}{\mathrm{e}^{-\theta} - 1}\right), \quad \theta \in \mathbb{R}\backslash\{0\}, \tag{3.22}$$

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 margins

3.3 Dependence Measures

Dependence measures allow us to summarize a particular kind of dependence into a single number.[9] 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.[10]

[9] In the bivariate case

[10] 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 margins and correlation ρ, one is able to construct a joint distribution only for the class of elliptical distributions.
- Given the margins, 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[11] of (X_1, X_2). Then *Kendall's tau* is defined by

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

[11] 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,[12] 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)

[12] 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.

Measure / Copula	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 margins 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*[13] and for extensive literature, good references to view are Klein and Kneib [2016] & Marra and Radice [2016]. Note that notations (such as symbols or letters) may differ from some of the reference literatures in order to align them with this paper.

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

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

[13] In Klein and Kneib [2016] they are referred to as "structured additive conditional copula regression models".

in conjunction with all facets of Section 3.1 [Patton, 2006].

The marginal CDFs $F_d\left(y_{id}|\boldsymbol{\nu}_i\right)$ 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 margins 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.

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 elements of the hierarchical structure depicted in Figure 4.1[14] and in particular our KCCs of interest are *"KCC 2"*, *"KCC 6"* and *"KCC 8"*.

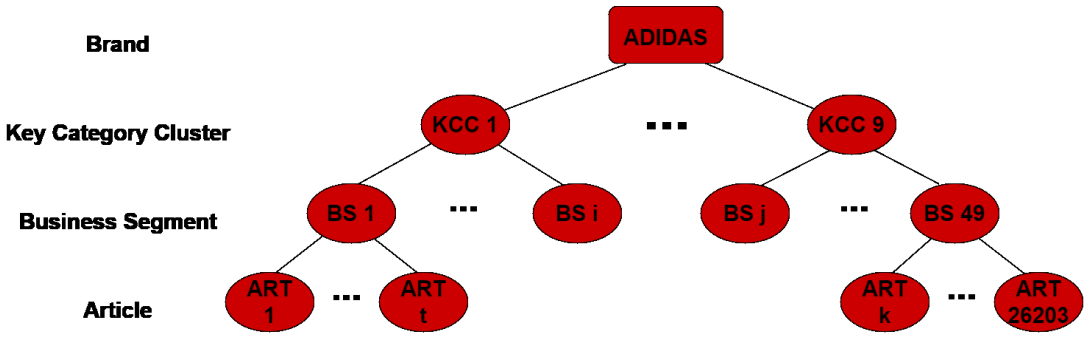


Figure 4.1: Illustration of a hierarchical article structure

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).

[14] For intermediate levels, we will focus on key category cluster only.

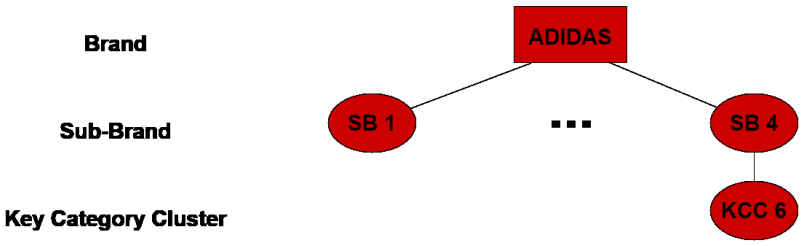


Figure 4.2: Example of a single child node

4.1 Universal Sale Patterns

As mentioned in Section subsection 1.2, the data contain 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 vertical 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.

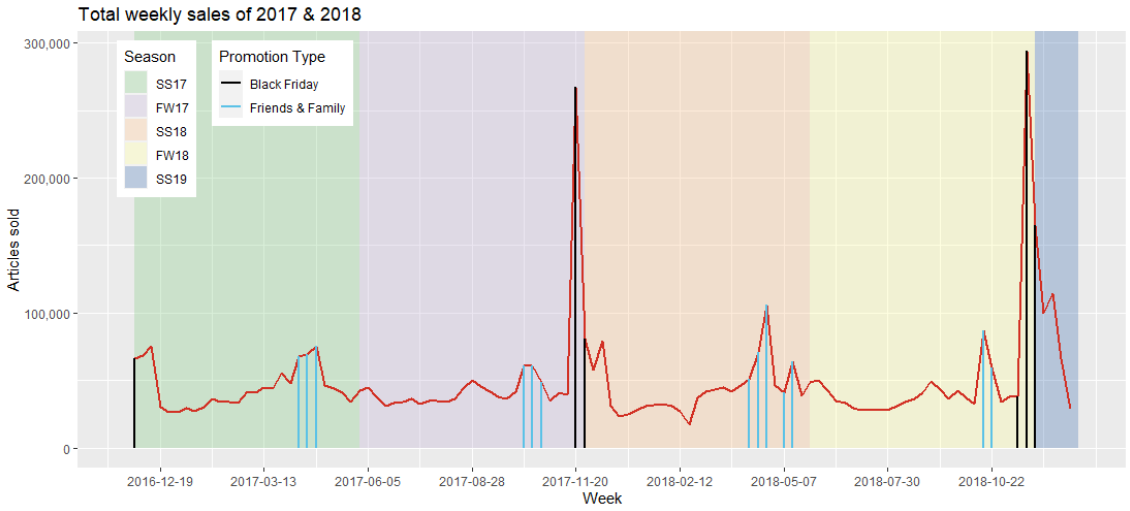


Figure 4.3: Course of article unit sales

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 (blue vertical lines).

Tables 4.1 and 4.2 show the weeks were Black Friday and Friends & Family took place respectively. The dates on those tables are one week apart each and always indicate the

Monday of the respective week[15].

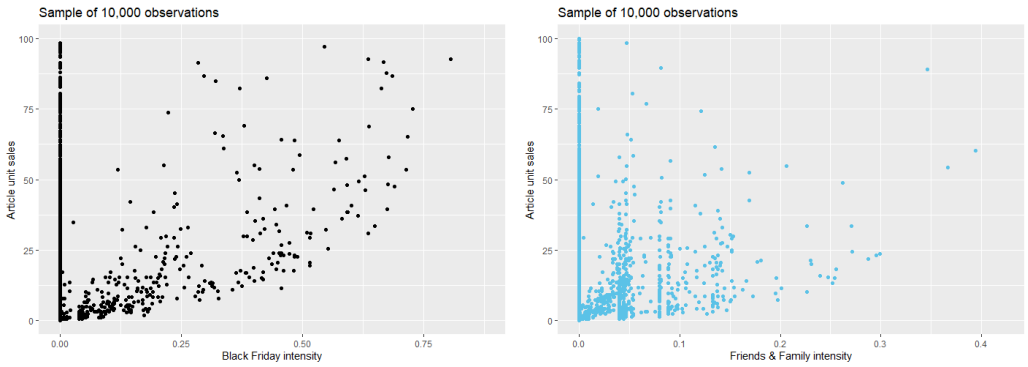
Black Friday weeks	2016-11-28	2017-11-20	2017-11-27	2018-11-12	2018-11-19	2018-11-26

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

Figure 4.4 depicts scatterplots of 10,000 randomly chosen observations in the dataset, where the two main promotion intensities are plotted against article unit sales. An overall positive relationship is visible as we would expect, even more so for Black Friday. Due to the huge noise persisting in these relationships however, Pearson's correlation coefficient for Black Friday against sales and Friends & Family against sales take on values of 0.32 and 0.11 respectively. Notice the high concentration of different sale quantities when there is neither Black Friday nor Friends & Family promotion activity (vertical points on the y-axes of Figure 4.4). Similar behavioural conclusions can be made about the total markdown percentage, with a correlation coefficient of 0.27 (see Figure 4.5). The type of season (Fall-Winter (FW) or Spring-Summer (SS)) doesn't seem to make an overall difference in the sale quantities, as a glimpse at Figure 4.6 points out (at least for modelling purposes as will be discussed in Chapter 5).



(a) Black Friday against unit sales (b) Friends & Family against unit sales

Figure 4.4: Scatterplots of promotion intensities against article unit sales; The y-axes are cut at 100

[15]According to European standards, a week starts on Monday.

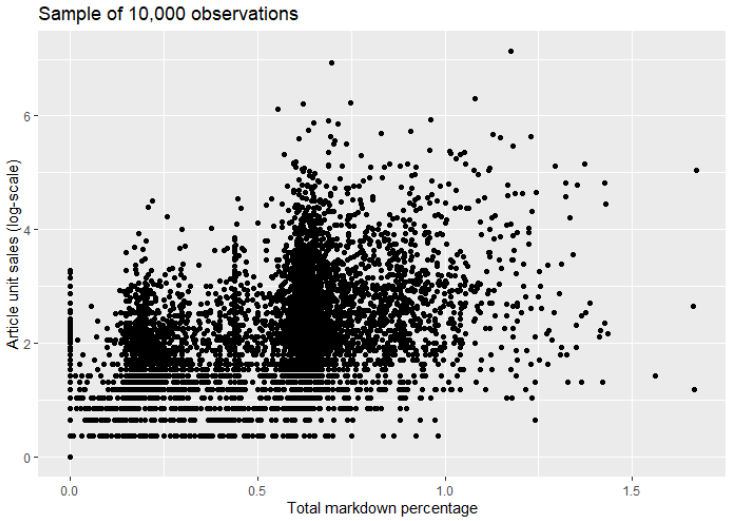


Figure 4.5: Unit sales in log-scale against total markdown percentage

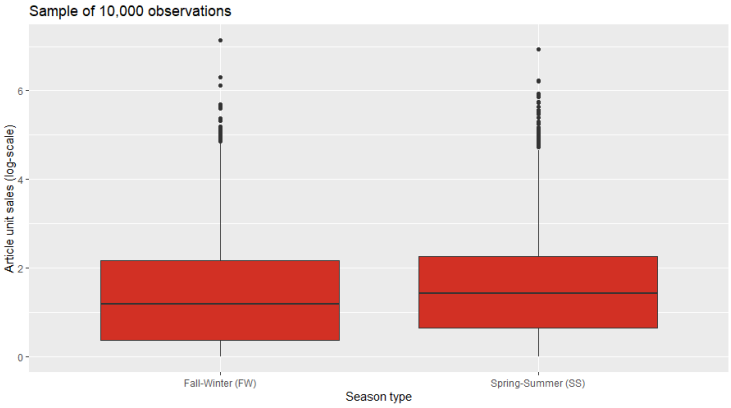
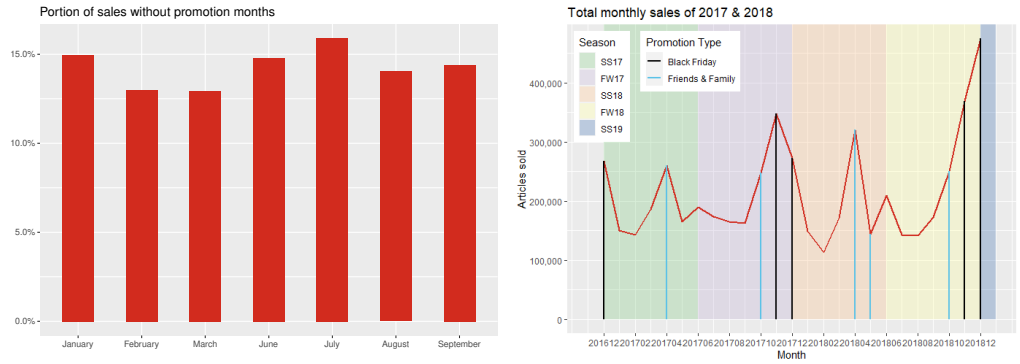


Figure 4.6: Unit sales in log-scale against season type

Regarding the months, they don't seem to differ much 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 4.7a. We have slightly higher sale 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 4.7b).



(a) Sales distribution grouped by months without promos

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

Figure 4.7: Monthly patterns of article unit sales

Moving forward, reviewing some sale 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,[16] 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 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.

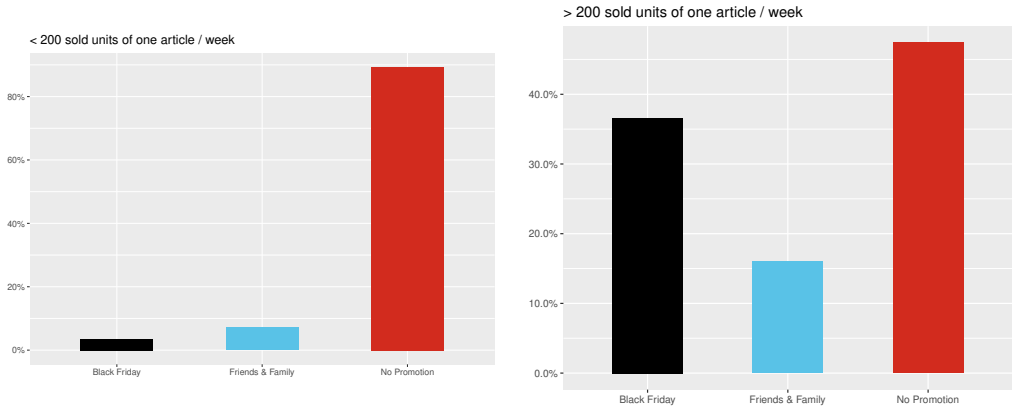
Quantile	Min	25%	50%	75%	90%	95%	99%	99.9%	Max
# Sold units / week	0	0.45	2.71	8.14	19.45	34.39	102.71	360.12	6,816.74
# Affected articles	26,203	26,195	23,797	17,014	10,275	6,458	1,800	273	1

Table 4.3: Number of sold units per week & number of affected articles for different 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 sale numbers. In Figure 4.8a 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 sale 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 4.8b and we can clearly see a change from Figure 4.8a. As expected, Black Friday is the dominating promotion type, although the majority remains in not promoted article

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

sales.



(a) Below 200 units

(b) Above 200 units

Figure 4.8: Distribution of sold units of articles per week split at 200 units

To validate the exploration on this, we may look at the empirical CDF in Figure 4.9 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.

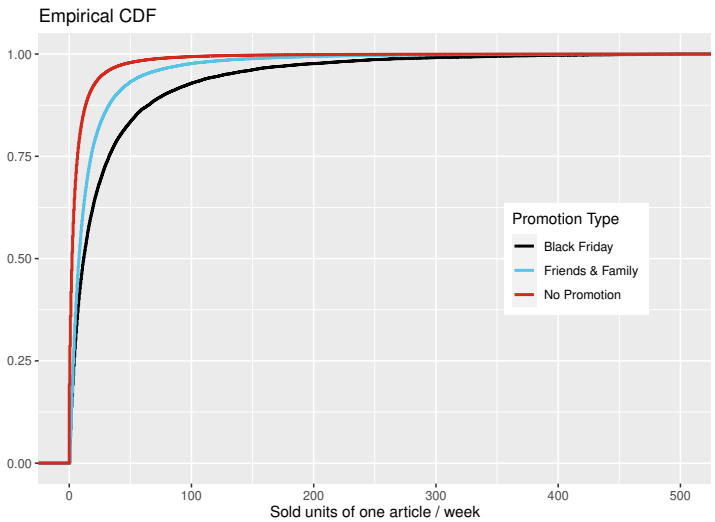


Figure 4.9: Empirical CDF of all sold units per week; x-axis cut at 500

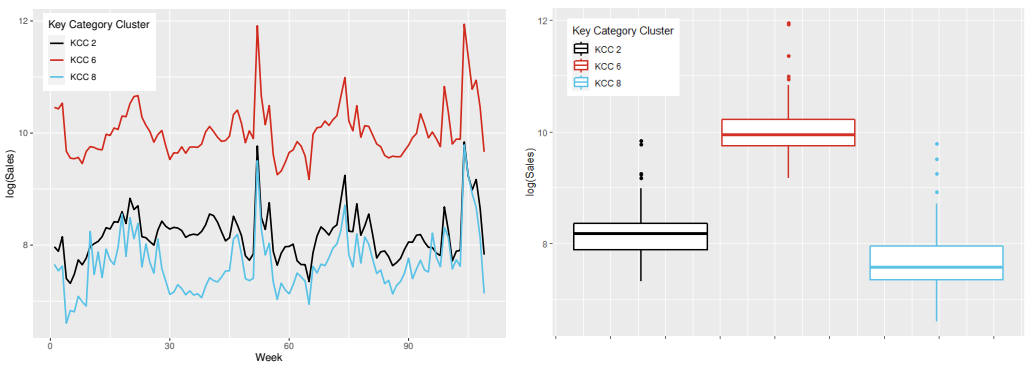
Unfortunately, such outliers with extreme sale numbers should not be removed from the

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. 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 no removal of instances will be employed to the dataset.

4.2 Grouped Sale Patterns - Key Category Cluster

To gain some insights on a hierarchical level of interest, some quick analysis is performed on different groups (nodes) of the key category cluster level in the tree (see Figure 4.1). We start out with a broad picture on this upper level and eventually reach a better understanding for the the article behaviour.

By viewing the sale trends separately for each key category cluster, we can observe in Figure 4.10a 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 sale behaviour in Figure 4.10b, 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 4.10: Time series and boxplot showing logarithmized sales of the key category clusters

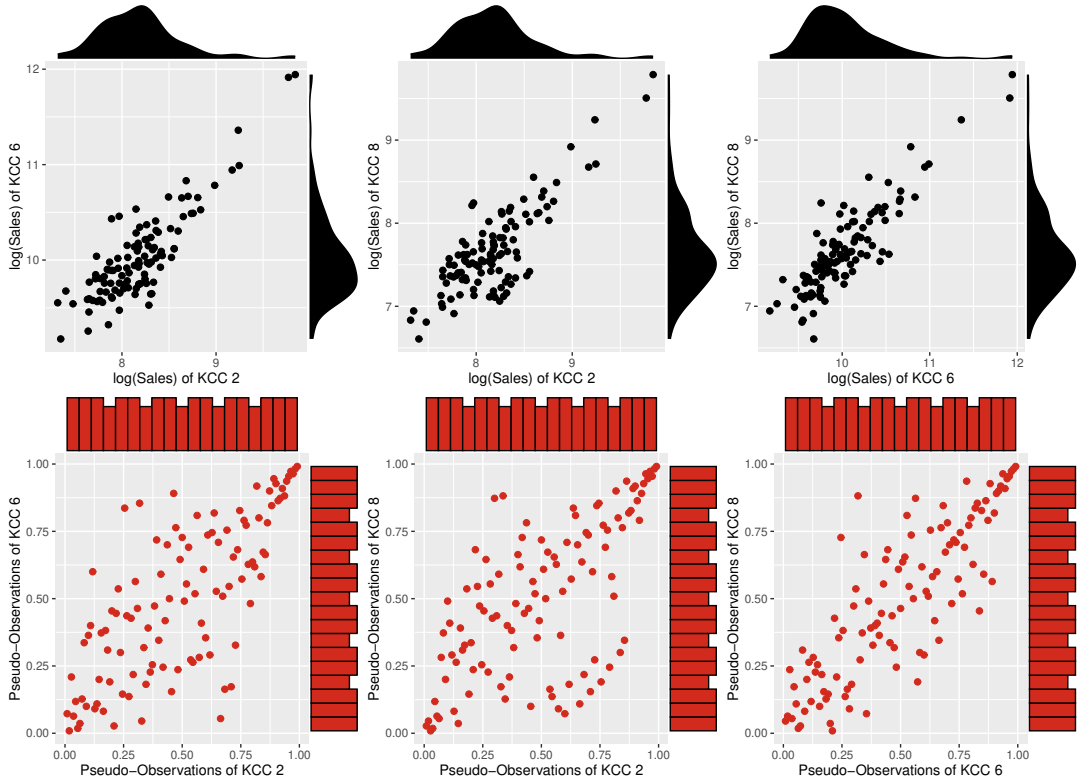


Figure 4.11: Pairwise scatterplots of sales on KCC level. First row: Logarithmic sales with marginal densities, Second row: Pseudo sales observation with marginal histograms

All the more interesting are the joint distributions of our KCCs. Figure 4.11 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 margins, pertinent marginal distributions 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 margins (see Chapter 5). In addition, looking at both rows of Figure 4.11, we suspect tail dependence and there is an obvious strong positive correlation among all three pairs, which is confirmed by viewing Figure 4.12 displaying three correlation metrics; Pearson'

rho, Kendall's tau and Spearman's rho.

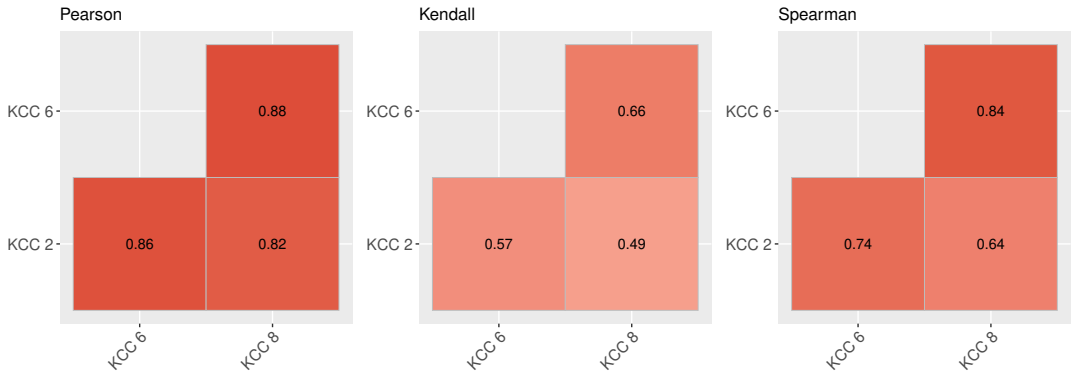


Figure 4.12: Correlation plots of the three KCC log-sales with different correlation coefficients. Left: Pearson's rho, Middle: Kendall's tau, Right: Spearman's rho

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. The boxplots of the two promotion types depicted in Figure 4.13 and Figure 4.14 point out how they affect the sales. Though, one shall keep in mind that, out of 109 weeks, only the minority include promo activation. Precisely, Black Friday is activated over 6 weeks out of 109 and Friends & Family is activated over 13 weeks[17] (see Tables 4.1 and 4.2). In Figure 4.14, one shall also be aware of large outliers being present during no Friends & Family weeks, which is fairly aligned with what we figured out in the previous subsection (i.e. many sale occurrences do not have any promotion).

[17] And of course not in a row, as can be clearly observed in Figure 4.3.

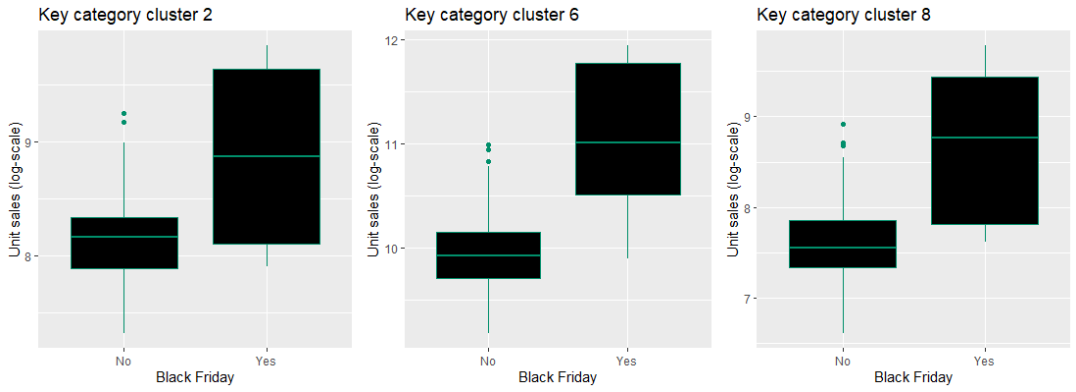


Figure 4.13: Boxplots showing log-sales of KCCs against presence of Black Friday

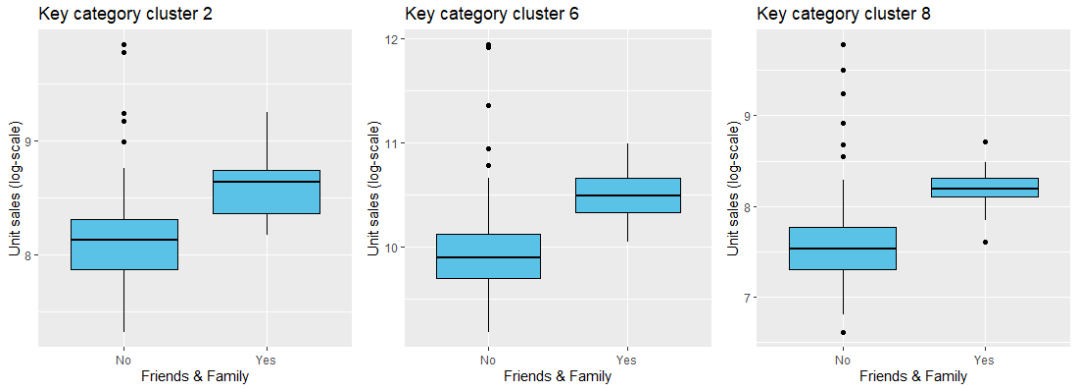


Figure 4.14: Boxplots showing log-sales of KCCs against presence of Friends & Family

The scatterplots in Figure 4.15 clearly reveal the strong positive relationship between the log-sales and the KCCs' respective total markdown percentages.

Regarding the season type (SS vs FW), visual exploration is not sufficient to conclude existence of effects on the unit sales. The effect of season type as well as other features on the unit sales shall be discussed in Chapter 5.

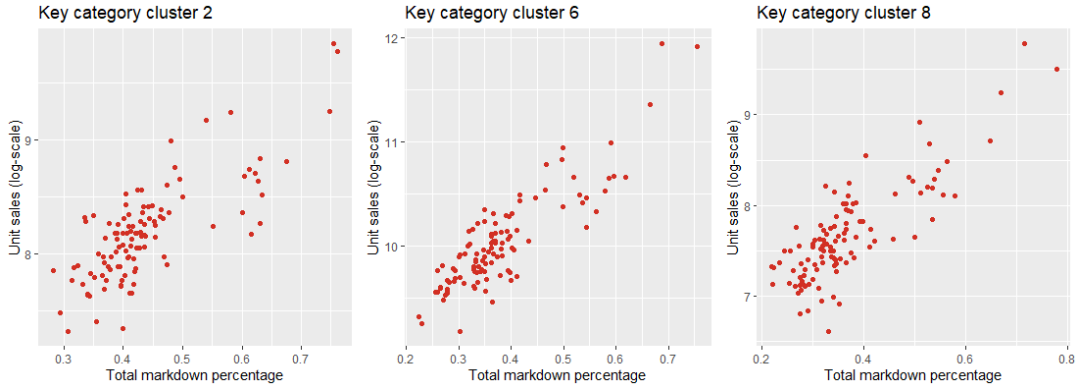


Figure 4.15: Scatterplots of KCC log-sales against total markdown percentage

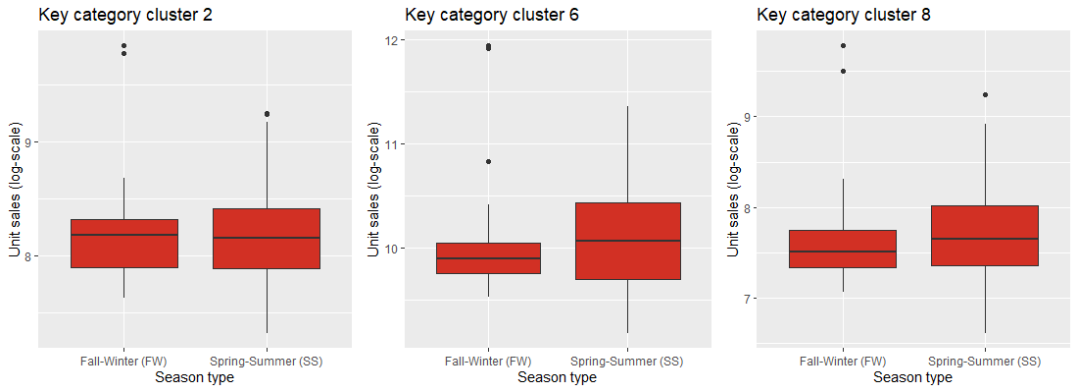


Figure 4.16: Boxplots of KCC log-sales against the two season types

4.3 Individual Sale Patterns

So far we explored the data pinpointing multiple characteristics of article unit sales along with other features in an aggregate or grouped frame. This section will briefly highlight some additional aspects of demand quantities regarding individual articles and the associated limitations.

Straight away, the biggest challenge is the life cycle of the articles. A sample of seven articles will list some these challenges. The course of those article sales is plotted in Figure 4.17. Article "5040" for example has only a lifespan of a 18 weeks and there are 12,679 out of 26,203 distinct articles having even less than that, which is almost half of the dataset. There are thousands of articles which have a lifespan of single digit (in weeks). The barplot in Figure 4.18 makes the extent of this issue visible. Articles that were observed during the entire observation window (or at least the majority of it) are the exception to the rule. It should be mentioned also that many articles started their selling journey on eCom before 2017 where the data for this task were acquired and other articles are still being in stock after 2018 (probably the case for article "21928"). These troublesome facts make it very hard to apply any kind of promising quantitative methods. Nevertheless, this issue will be discussed and tackled in later parts of this thesis (see Section 5.3). It should be pointed out that for sales of many individual article, high peaks can be regularly observed during Black Friday periods too.

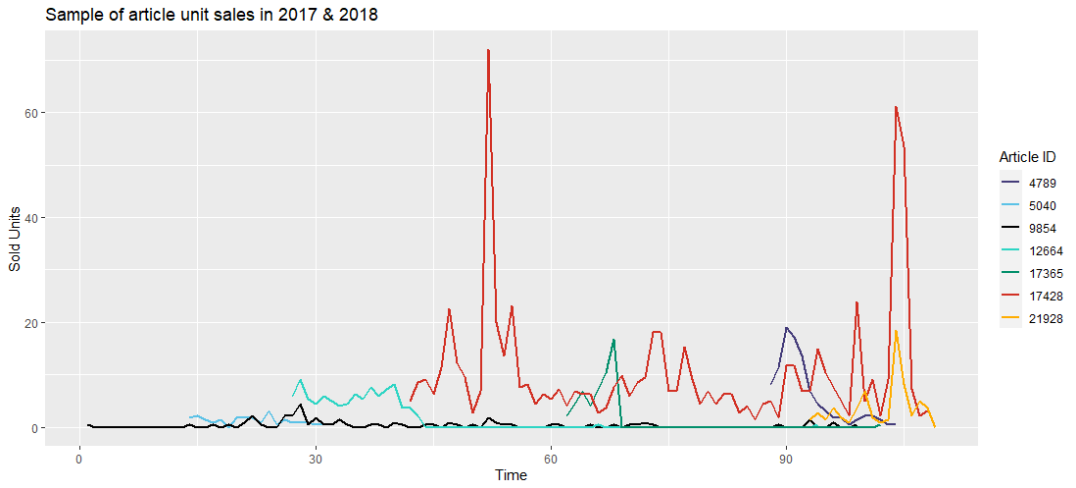


Figure 4.17: Sample of seven articles and their demand quantity life cycles

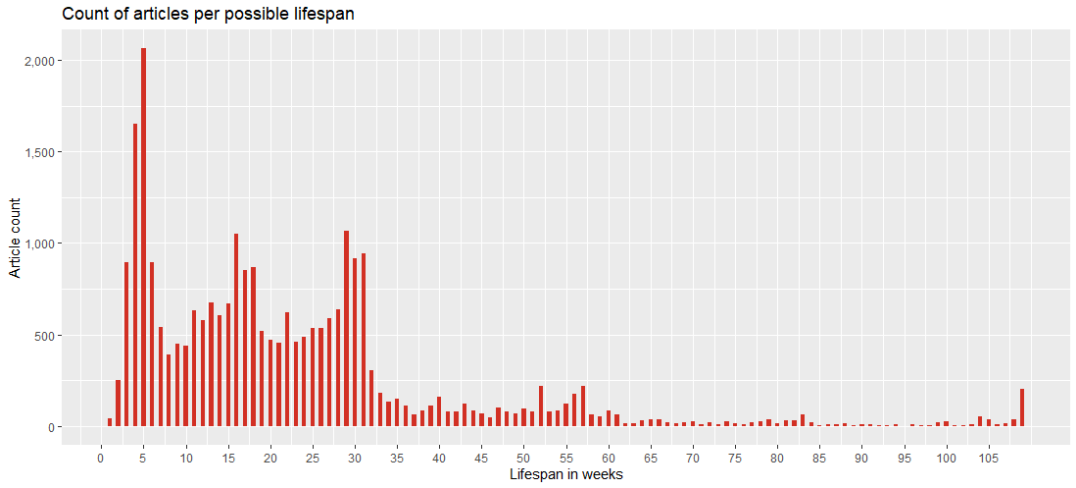


Figure 4.18: Number of articles for each possible lifespan of 1 to 109 weeks

Yet another immediate remark is the zero inflation persisting in the data. A lot of articles spend their time on eCom for several weeks without being sold once, as can be seen e.g. in Figure 4.17 for articles "9854", "12664" and "17365". To be precise, for 100,732 instances we have a gross demand quantity of zero, which makes up almost a fifth of the dataset. Combining this inflation of zeros with the short lifespan increases the level difficulty even more.

5 Modelling

In this chapter, we will explore diverse ways on how to model, evaluate and interpret the dependence structure of unit sales.

An 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].
As the observation period is only 109 weeks and the highest peaks occur during Black Friday twice, the second time being very late (104th week, see e.g. Figure 4.3), all of the modelling will be performed in-sample. Thus, evaluation and diagnostics will be based on the entire observation period so that both extreme peaks can be included in model trainings.

5.1 Key Category Cluster - Marginal Distributions

Figure 4.11 (Section 4.2) is hinting that the marginal distributions come with a noticeable skewness, which are best to take into account. The logarithmic scale is the preferred transformation due to variance stabilization of the margins. Among a pool of possible parametric distributions, we pick an appropriate one for each margin. Several distributions would theoretically be justifying the shape of our data, e.g. Weibull, Gamma, Box-Cox-Cole-Green or Dagum distribution. After screening those parametric distributions, we find that the exponentially modified Gaussian distribution (or ex-Gaussian distribution which will be used from here on) fits all three margins fairly well [Grushka, 1972].[18] Before entering the dependence structures between the three KCC pairs, the marginal distribution of each cluster will be analyzed individually in the next three subsections. First, the ex-Gaussian distribution will be fitted to the margins and in a second step covariate effects will be included to obtain flexible estimations on the distribution parameters.

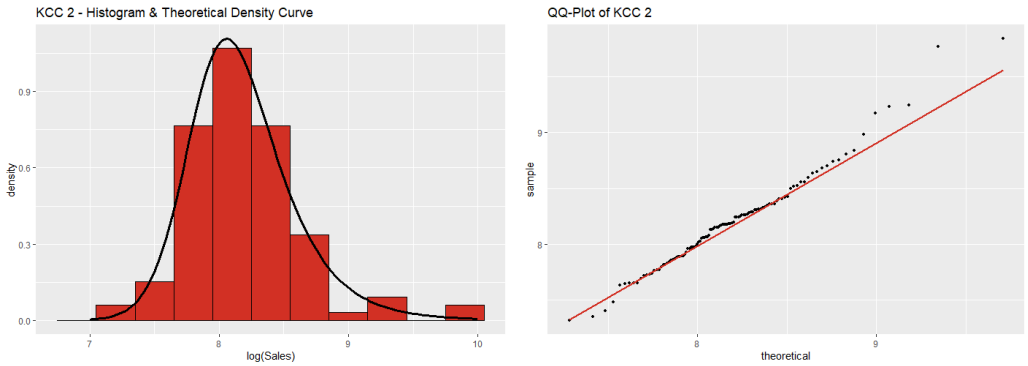
[18] Another appropriate distribution would be the Dagum distribution [Dagum, 1975], however interpretability of the parameters is difficult to comprehend.

5.1.1 Key Category Cluster 2

Simple maximum likelihood estimation based on the log-sales of the margins allow us to estimate the ex-Gaussian distribution parameters. Figure 5.1a describes how the histogram of the data match to the theoretical density of an ex-Gaussian distribution considering the estimated parameter values represented in Table 5.1.

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\nu}$
7.85	0.26	0.33

Table 5.1: Estimated parameters for log-sales of KCC 2 fitted to ex-Gaussian distribution with no covariate effects



(a) Histogram & theoretical density

(b) QQ-Plot

Figure 5.1: ex-Gaussian distribution fitted to log-sales of KCC 2

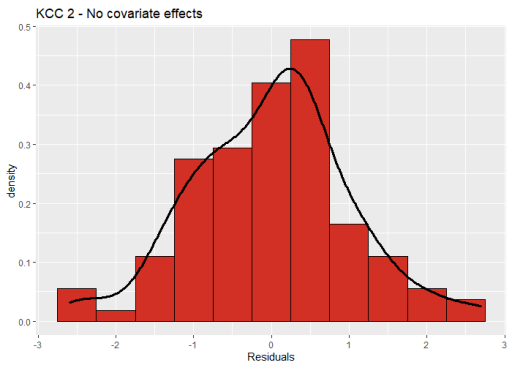


Figure 5.2: Residuals of KCC 2 log-sales fitted to an ex-Gaussian distribution with no covariate effects together with their density curve

As can be seen in Figure 5.2, the residuals of the fitted distribution are not too far from a normal distribution. In fact, a Shapiro-Wilk normality test (see Section 2.1) will fail to reject the null hypothesis, returning a p-value of 0.6645. There still exist skewness in the

distribution of the residuals and correction will be attempted in the following.

The findings so far indicate that the overall fit for this cluster is quite satisfiable and also supports interpretability of the estimated parameters.
To make the estimation more precise and to get a better understanding of what is driving sales, flexible estimation of the distribution parameters is required and thus covariate effects shall be included.

After multiple equation setups for the marginal distribution of KCC 2, the chosen GAMLSS model specification (see Section 2.4) is as follows:

$$\begin{aligned} \mu &= \beta_{01} + f_{11}(\textit{time}) + f_{12}(\textit{total\_markdown\_pct}) + \textit{promo\_type*total\_markdown\_pct} \\ log(\sigma) &= \beta_{02} + f_{21}(\textit{time}) \\ log(\nu) &= \beta_{03}, \end{aligned} \tag{5.1}$$

where *promo_type* is a factor variable with levels *"No Promotion"* (reference category), *"Black Friday"* and *"Friends & Family"*. The smooth functions f_{11}, f_{21} and f_{12} are build upon P-splines with 40 knots each. For the scale and shape parameters (σ and ν respectively) the logarithmic link function is used to ensure that they are mapped to the real positive line as the ex-Gaussian distribution family can only capture positive skewness (ν). For σ we just use time as the only regressor and we set the skewness ν as a constant.
The estimated location and scale parameters over time are depicted in Figure 5.3 and in Table 5.2 the estimated skewness parameter value with its 95% confidence interval is shown. The fitted values approximate the actual values quite adequately given the present limitations. The sale peaks are also well captured, especially for Black Friday. The scale parameter has an increasing trend up until roughly mid-January (60th observation) and then decreases again. However, the overall range of deviation is kept to a minimum, ranging from 0.05 to 0.20. The estimated skewness parameter is closer to zero than in the maximum likelihood approach without any covariate effects.

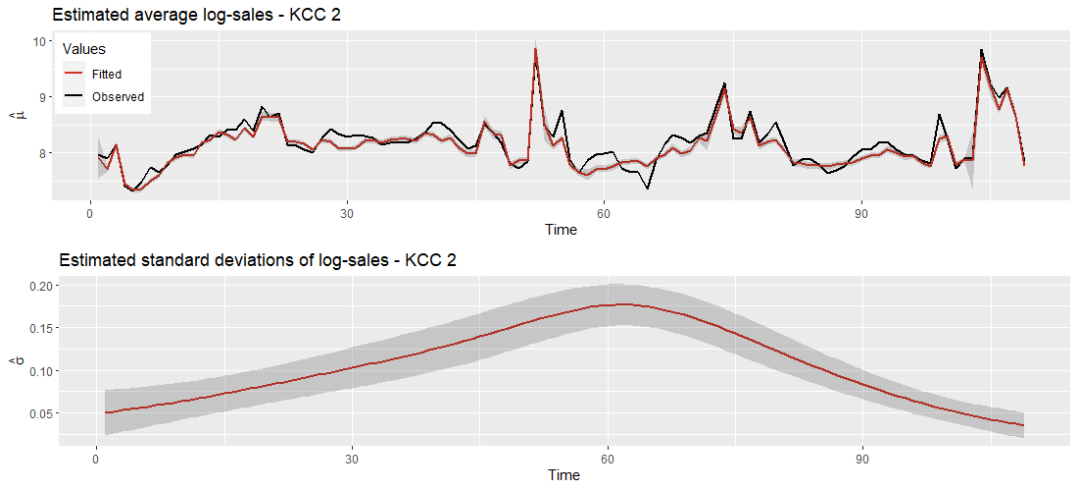


Figure 5.3: Estimated location parameter $\hat{\mu}$ compared to the observed values and scale parameter $\hat{\sigma}$ with confidence bands of GAMLSS fit - KCC 2

Lower	$\hat{\nu}$	Upper
0.067	0.058	0.076

Table 5.2: Estimated skewness parameter $\hat{\nu}$ of GAMLSS fit with 95% confidence interval bounds - KCC 2

Below we can observe the summary output from the R console (R output 5.1 as well as the visual covariate effects on the expected value (Figure 5.4).

R output 5.1: GAMLSS Fit on KCC 2

```
******************************************************************
Family:  c("exGAUS", "ex-Gaussian")

Call:  gamlss(formula = logsales_kcc_2 ~ pb(time_obs) + pb(total_markdown_pct) +
    promo_type * total_markdown_pct, sigma.formula = ~pb(time_obs),      nu.formula = ~1, family = "exGAUS",
         data = data_agg_KCC)

Fitting method: RS()

------------------------------------------------------------------
Mu link function:  identity
Mu Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                5.5420839  0.0730169  75.901  < 2e-16 ***
pb(time_obs)                               0.0010396  0.0002714   3.831 0.000251 ***
pb(total_markdown_pct)                     6.7884739  0.1846136  36.771  < 2e-16 ***
promo_typeBlack Friday                     1.3325479  0.1897891   7.021 6.22e-10 ***
promo_typeFriends & Family                 2.3073358  0.4146677   5.564 3.33e-07 ***
promo_typeBlack Friday:total_markdown_pct -3.8745823  0.3674925 -10.543  < 2e-16 ***
promo_typeFriends & Family:total_markdown_pct -6.0992694  0.7498619  -8.134 4.18e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

------------------------------------------------------------------
Sigma link function:  log
Sigma Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.838160   0.232620 -12.201  < 2e-16 ***
```

```
pb(time_obs)  0.032702   0.003654   8.949 1.03e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-------------------------------------------------------------------
Nu link function:  log
Nu Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.7038     0.2471  -10.94   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-------------------------------------------------------------------
NOTE: Additive smoothing terms exist in the formulas:
 i) Std. Error for smoothers are for the linear effect only.
ii) Std. Error for the linear terms maybe are not accurate.
-------------------------------------------------------------------
No. of observations in the fit:  109
Degrees of Freedom for the fit:  28.27367
      Residual Deg. of Freedom:  80.72633
                      at cycle:  20

Global Deviance:     -148.3187
            AIC:     -91.77135
            SBC:     -15.67708
*******************************************************************
```

From the visual representation of the covariate effects in (Figure 5.4, we can see that the temporal effect on log-sales varies within a certain range. Higher total markdown percentages enhance higher response values, which can also be observed when looking at the R output. The estimated coefficient for the markdown obtains a value of 6.79 (without considering promotions). The log-sales increase by 1.33 and 2.31 units on average when Black Friday and Friends & Family respectively is activated, assuming no total markdown percentages exist. That is of course counterintuitive, as promotion go hand in hand with markdowns. The summary in R output 5.1 reveals such kind of peculiar effects when we look at the interaction coefficients of promotions and total markdown percentage. When any promo type is activated, the total markdown percentage has a negative effect on the log-sales. Interpretation of the results shall be expressed with caution. Season type (SS & FW) was not included in the model as it had no significant effect and distorts the output and the effect plots. All curves of smooth effects are centered around zero due to identifiability constraints (see Section 2.3).

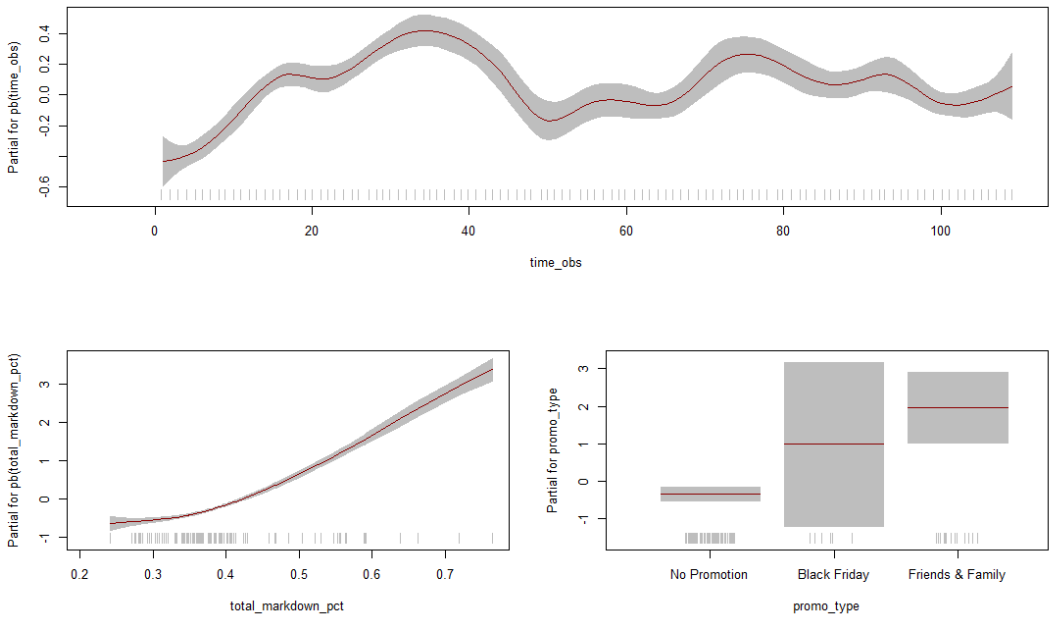


Figure 5.4: Covariate effects on the expected response variable (log-sales) of GAMLSS fit - KCC 2

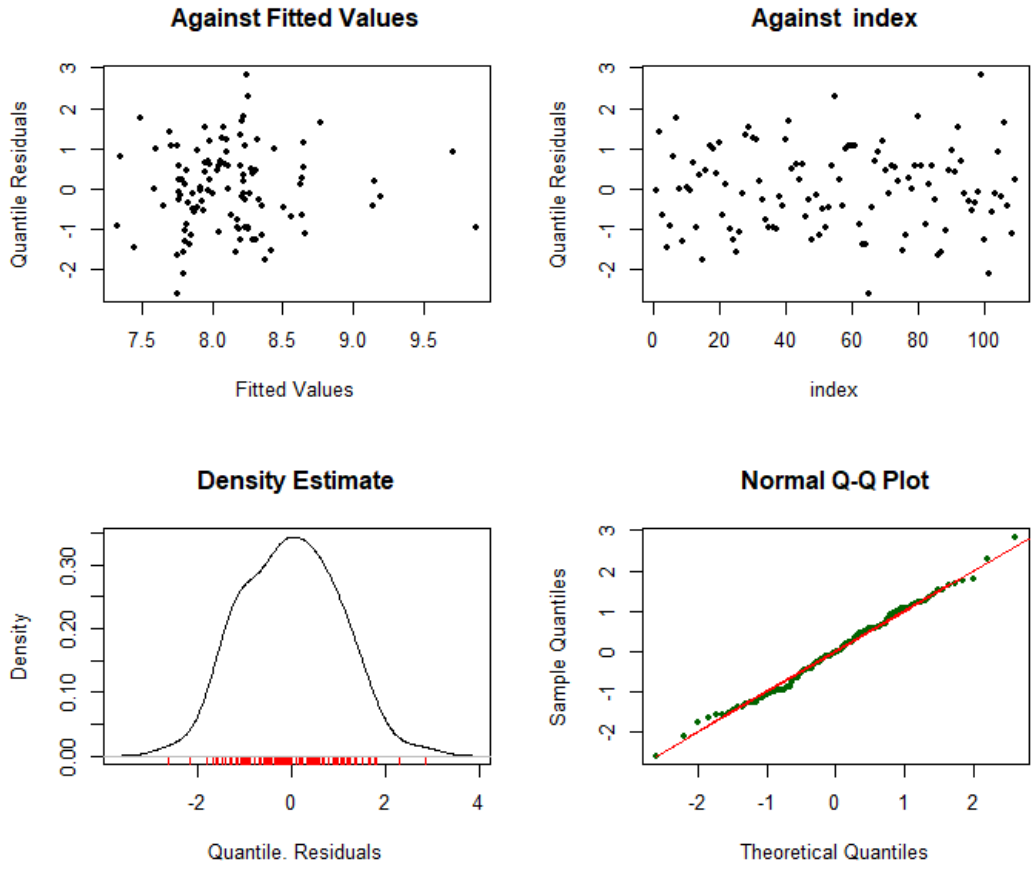


Figure 5.5: Residuals of GAMLSS fit - KCC 2

Fitting the data to a GAMLSS model like this can be considered favourable. The emerged residuals are closer to a standard normal distribution than the maximum likelihood esti-

mation without any covariate effects. Diagnostic plots of quantile residuals in Figure 5.5 confirm this. R output 5.2 below returns a detailed summary of the estimated location, scale and shape parameters of the residuals' distribution. The mean, variance, skewness and kurtosis are close enough to zero, one, zero and three respectively, which are sufficient conditions to assume normality for practical purposes. A Shapiro-Wilk test for the newly emerged residuals returns a p-value of 0.92, which is also in plain favour of a normal distribution.

R output 5.2: Residuals of GAMLSS Fit on KCC 2

```
1 ******************************************************************
2                Summary of the Quantile Residuals
3                             mean   =  -0.01811135
4                         variance   =  1.042057
5                 coef. of skewness  =  0.08348086
6                 coef. of kurtosis  =  2.645021
7 Filliben correlation coefficient   =  0.996831
8 ******************************************************************
```

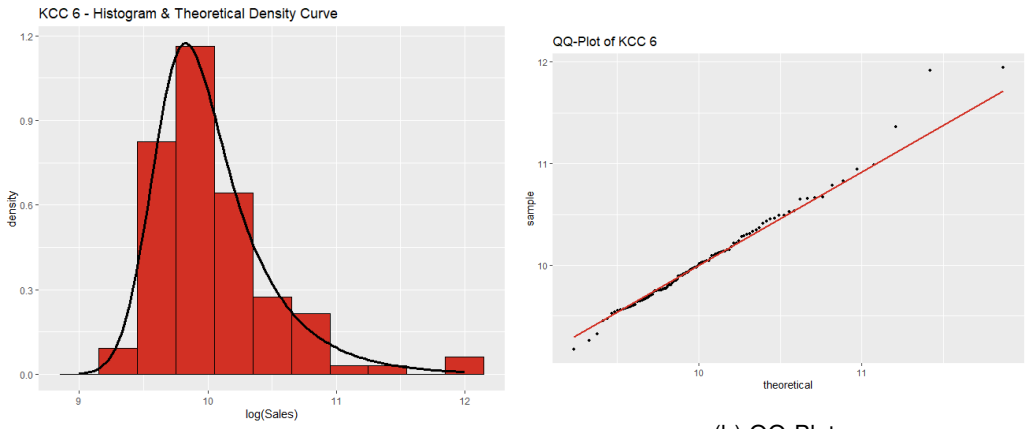
5.1.2 Key Category Cluster 6

The same procedure as in Subsection 5.1.2 for key category cluster 2 will be applied here to analyze the marginal distribution of the log-sales in key category cluster 6.

Excluding all covariates, simple maximum likelihood estimation results can be summarized within Table 5.3, Figure 5.6 and Figure 5.7. A Shapiro-Wilk test on the residuals returns a p-value of 0.87 and the fails to reject the null hypothesis of non-normality.

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\nu}$
9.62	0.20	0.41

Table 5.3: Estimated parameters for log-sales of KCC 6 fitted to ex-Gaussian distribution with no covariate effects



(a) Histogram & theoretical density

(b) QQ-Plot

Figure 5.6: ex-Gaussian distribution fitted to log-sales of KCC 6

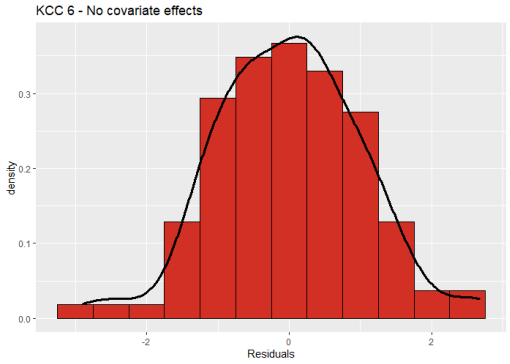


Figure 5.7: Residuals of KCC 6 log-sales fitted to an ex-Gaussian distribution with no covariate effects together with their density curve

Reviewing different model specifications, an equivalent model as in the previous Subsection is chosen (Model 5.1) with an ex-Gaussian distribution family for the response variable. A summary is printed below in R output 5.3. The estimated time-varying location and scale parameters can be seen in Figure 5.8 and the skewness parameter $\hat{\nu}$ with 95% CI in Table 5.4. Fitted values are close to real values, capturing the promotion peaks fairly well. The standard deviation fluctuates throughout time within a range between 0.1 and 0.3.

R output 5.3: GAMLSS Fit on KCC 6

```
1  ******************************************************************
2  Family:  c("exGAUS", "ex-Gaussian")
3
4  Call:  gamlss(formula = logsales_kcc_6 ~ pb(time_obs) + pb(total_markdown_pct) +
5      promo_type * total_markdown_pct, sigma.formula = ~pb(time_obs),      nu.formula = ~1, family = "exGAUS",
           data = data_agg_KCC)
6
7  Fitting method: RS()
8
9  ------------------------------------------------------------------
```

```
Mu link function:  identity
Mu Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                8.3111193  0.1157520  71.801  < 2e-16 ***
pb(time_obs)                               0.0024889  0.0004789   5.197 1.15e-06 ***
pb(total_markdown_pct)                     4.0074977  0.3016532  13.285  < 2e-16 ***
promo_typeBlack Friday                    -0.8054399  0.2951362  -2.729  0.00756 **
promo_typeFriends & Family                 0.8703618  0.4802656   1.812  0.07309 .
promo_typeBlack Friday:total_markdown_pct  1.0176236  0.5352771   1.901  0.06031 .
promo_typeFriends & Family:total_markdown_pct -2.4686730  0.8798442  -2.806  0.00608 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-------------------------------------------------------------------
Sigma link function:  log
Sigma Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.255651   0.149038  -8.425  3.7e-13 ***
pb(time_obs) -0.010311   0.003173  -3.250   0.0016 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-------------------------------------------------------------------
Nu link function:  log
Nu Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -3.126      1.164  -2.686  0.00852 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-------------------------------------------------------------------
NOTE: Additive smoothing terms exist in the formulas:
 i) Std. Error for smoothers are for the linear effect only.
ii) Std. Error for the linear terms maybe are not accurate.
-------------------------------------------------------------------
No. of observations in the fit:  109
Degrees of Freedom for the fit:  13.62438
      Residual Deg. of Freedom:  95.37562
                      at cycle:  19

Global Deviance:     -105.8531
            AIC:     -78.6043
            SBC:     -41.93636
*******************************************************************
```

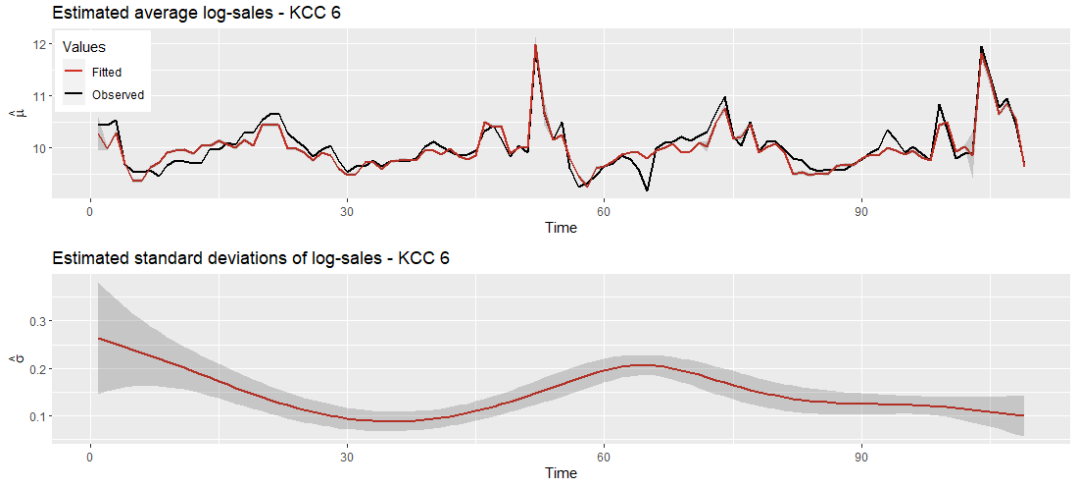


Figure 5.8: Estimated location parameter $\hat{\mu}$ compared to the observed values and scale parameter $\hat{\sigma}$ with confidence bands of GAMLSS fit - KCC 6

Lower	$\hat{\nu}$	Upper
0.044	0.031	0.057

Table 5.4: Estimated skewness parameter $\hat{\nu}$ of GAMLSS fit with 95% confidence interval bounds - KCC 6

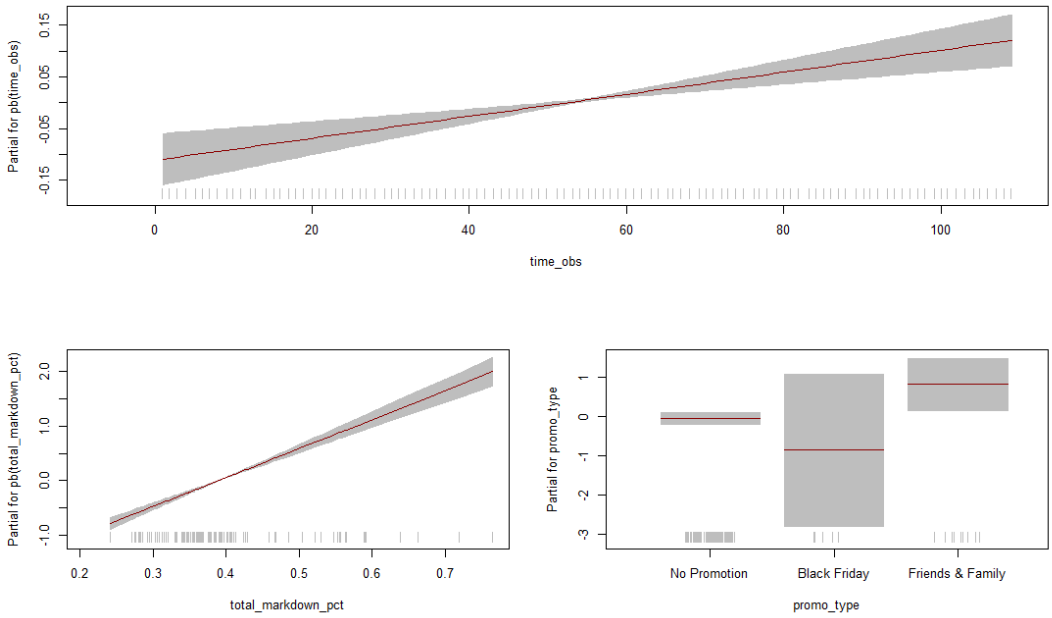


Figure 5.9: Covariate effects on the expected response variable (log-sales) of GAMLSS fit - KCC 6

Figure 5.9 reveals some interesting points regarding covariate effects. The temporal effect, just like the total markdown percentage, collapses to an increasing straight line. As opposed to the GAMLSS fit for KCC 2, Friends & Family seems to have the highest effect of all promos for this KCC. Controversial results are again the case here, as can be seen in R output 5.3. The two promotions seem to negatively interact with the total markdown percentage. Those kinds of behaviour might need further investigation.

Inspecting the diagnostic plots for the residuals in Figure 5.10 along with the associated R output 5.4, we can again confirm an appropriate fit. A Shapiro-Wilk test returns a p-value of 0.7, which is below the p-value of the fit without covariate effects. Nevertheless, normality is a steady assumption for the quantile residuals.

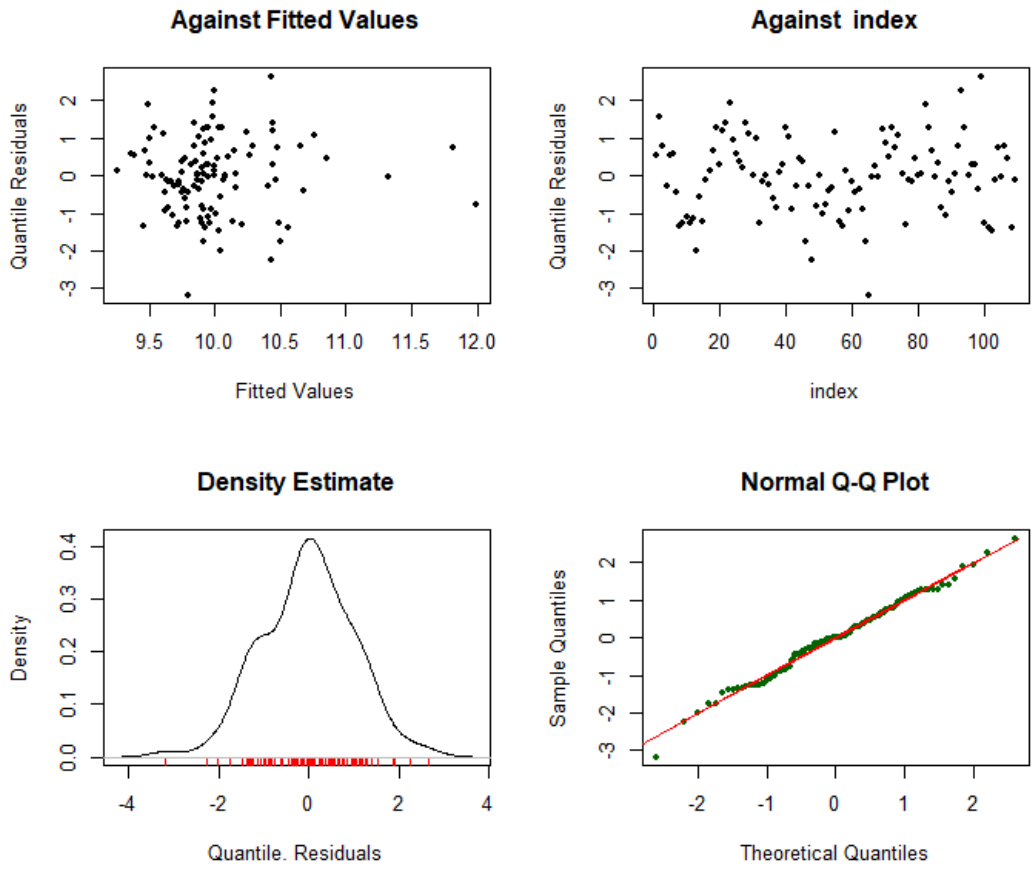


Figure 5.10: Residuals of GAMLSS fit - KCC 6

R output 5.4: Residuals of GAMLSS Fit on KCC 6

```
1 ******************************************************************
2                Summary of the Quantile Residuals
3                             mean   =  -0.01692598
4                         variance   =  1.011483
5                 coef. of skewness  =  -0.1401063
6                 coef. of kurtosis  =  3.144701
7 Filliben correlation coefficient   =  0.9948268
8 ******************************************************************
```

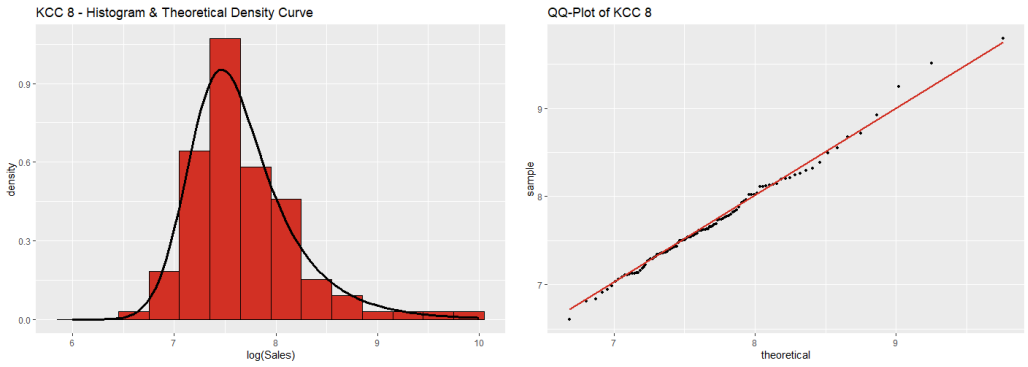
5.1.3 Key Category Cluster 8

The procedure is continued also for key category cluster 8 with the same conditions, since we have similar issues as in the previous two clusters.

Table 5.5, Figure 5.11 and Figure 5.12 summarize the findings when the log-sales of KCC 8 are fitted to an ex-Gaussian distribution with no regressors. Normality of the residuals can be assumed, as the Shapiro-Wilk test returns a p-value of 0.99 and thus fails to reject the null hypothesis of non-normality.

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\nu}$
7.21	0.27	0.46

Table 5.5: Estimated parameters for log-sales of KCC 8 fitted to ex-Gaussian distribution with no covariate effects



(a) Histogram & theoretical density

(b) QQ-Plot

Figure 5.11: ex-Gaussian distribution fitted to log-sales of KCC 8

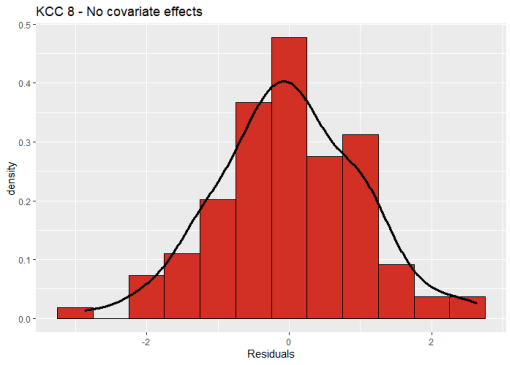


Figure 5.12: Residuals of KCC 8 log-sales fitted to an ex-Gaussian distribution with no covariate effects together with their density curve

Again, Model 5.1 is applied to the log-sales. The summary output and estimated parameters with confidence intervals can be seen in Figure 5.13 and in Table 5.6. We notice that models fit is acceptable matching the elevated sales during promo weeks and that the variability is strictly decreasing with the overall range having a range between 0.1 and 0.3. Skewness remains low in the data.

R output 5.5: GAMLSS Fit on KCC 8

```
1 ******************************************************************
2 Family:  c("exGAUS", "ex-Gaussian")
3
4 Call:  gamlss(formula = logsales_kcc_8 ~ pb(time_obs) + pb(total_markdown_pct) +
```

```
    promo_type * total_markdown_pct, sigma.formula = ~pb(time_obs),      nu.formula = ~1, family = "exGAUS",
         data = data_agg_KCC)

Fitting method: RS()

------------------------------------------------------------------
Mu link function:  identity
Mu Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                5.5055060  0.1744217  31.564  < 2e-16 ***
pb(time_obs)                               0.0042645  0.0008424   5.063 1.91e-06 ***
pb(total_markdown_pct)                     4.5349719  0.4527266  10.017  < 2e-16 ***
promo_typeBlack Friday                    -0.6541342  0.3738172  -1.750   0.0832 .
promo_typeFriends & Family                -0.0367560  0.6637042  -0.055   0.9559
promo_typeBlack Friday:total_markdown_pct  0.8760110  0.7295349   1.201   0.2327
promo_typeFriends & Family:total_markdown_pct -0.7643296  1.2393106  -0.617   0.5388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

------------------------------------------------------------------
Sigma link function:  log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.347871   0.271081  -4.972 2.78e-06 ***
pb(time_obs) -0.008544   0.005004  -1.708   0.0908 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

------------------------------------------------------------------
Nu link function:  log
Nu Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.7105     0.2198  -7.783    7e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

------------------------------------------------------------------
NOTE: Additive smoothing terms exist in the formulas:
 i) Std. Error for smoothers are for the linear effect only.
ii) Std. Error for the linear terms maybe are not accurate.
------------------------------------------------------------------
No. of observations in the fit:  109
Degrees of Freedom for the fit:  10.0084
      Residual Deg. of Freedom:  98.9916
                      at cycle:  7

Global Deviance:     -11.3607
            AIC:     8.656098
            SBC:     35.59218
******************************************************************
```

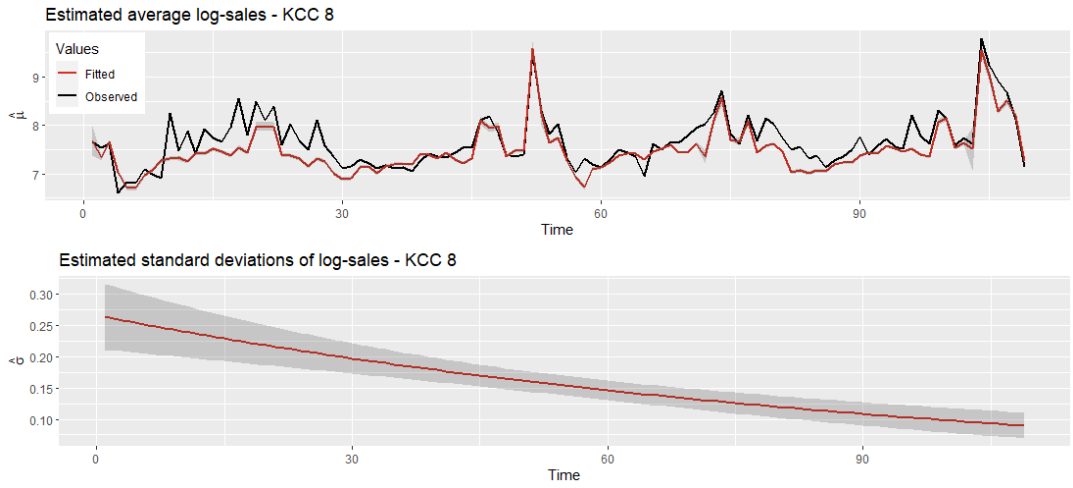


Figure 5.13: Estimated location parameter $\hat{\mu}$ compared to the observed values and scale parameter $\hat{\sigma}$ with confidence bands of GAMLSS fit - KCC 8

Lower	$\hat{\nu}$	Upper
0.181	0.158	0.203

Table 5.6: Estimated skewness parameter $\hat{\nu}$ of GAMLSS fit with 95% confidence interval bounds - KCC 8

Time as well as the total markdown percentage exhibit similar positive effects on the response as in KCC 6, collapsing into a straight line. The promo effects seem to be more balanced, with different interquartile ranges for all cases.

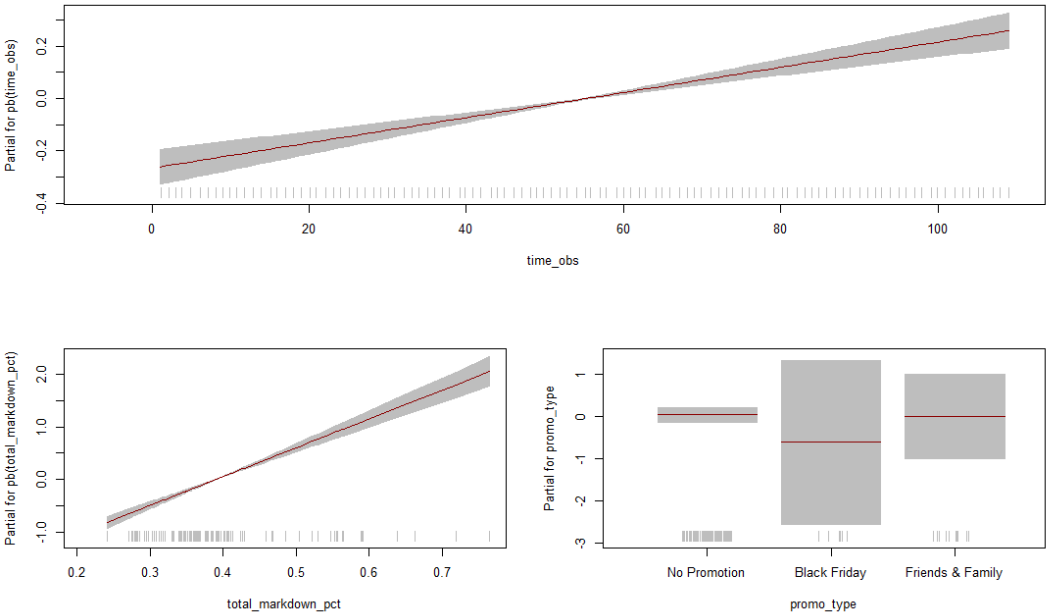


Figure 5.14: Covariate effects on the expected response variable (log-sales) of GAMLSS fit - KCC 8

Looking at Figure 5.15 and R output 5.6 for the distribution of the quantile residuals, the fitting method for this cluster is also justified. The Shapiro-Wilk test is "weaker" in comparison to the fit without covariate effects with a p-value of 0.86, which is still a solid reason to retail the null hypothesis of normality.

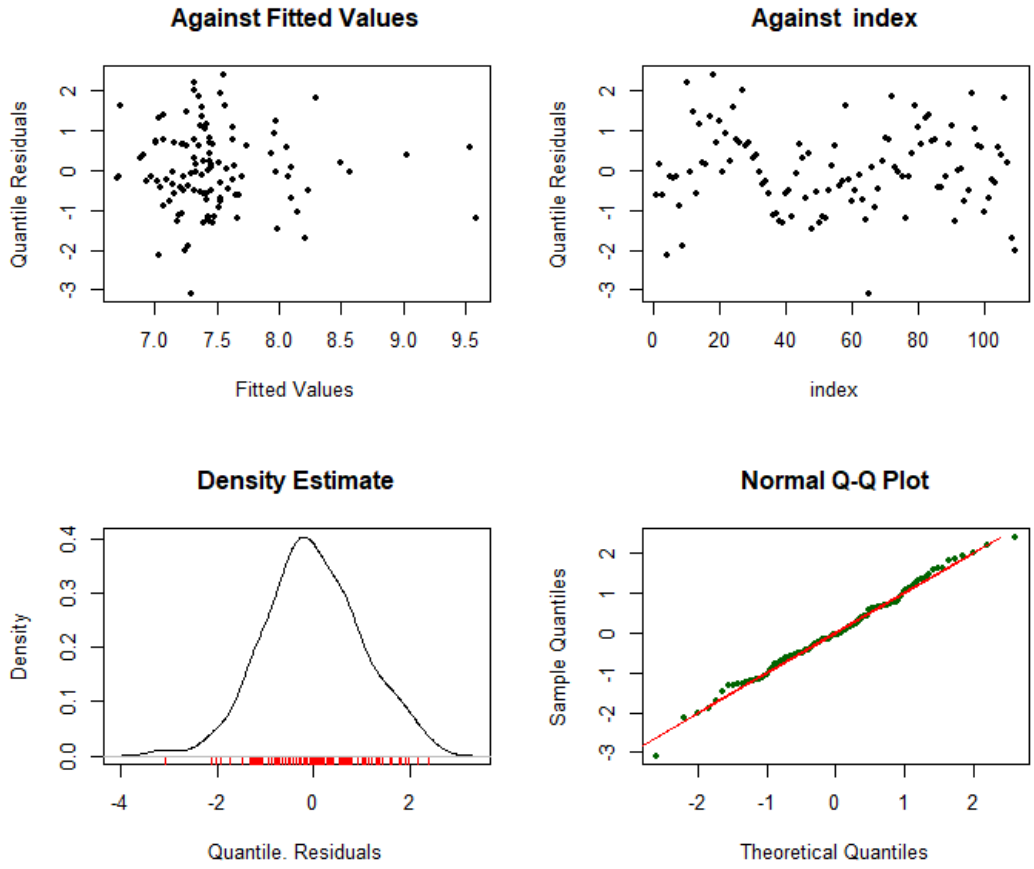


Figure 5.15: Residuals of GAMLSS fit - KCC 8

R output 5.6: Residuals of GAMLSS Fit on KCC 8

```
*****************************************************************
              Summary of the Quantile Residuals
                           mean   =  0.0004673878
                       variance   =  1.008849
                 coef. of skewness  =  -0.02817569
                 coef. of kurtosis  =  3.04467
Filliben correlation coefficient  =  0.9963667
*****************************************************************
```

5.2 Key Category Cluster - Pairwise Copulas

The next step, after fitting proper models to the marginal distributions of the log-sales aggregated on key category cluster level, is to enter the pairwise modelling of the clusters. Specifically, we are interested in capturing the dependence structure over time which is

delineated in the next three Subsections (5.2.1, 5.2.2 and 5.2.3).

Recalling the previous Section 5.1, the residuals from the fitting method are depicted in Figure 5.16 for each time-point. In a second step, the residuals will be used in a pairwise fashion after the marginal modelling to estimate the dependence structure with the help of conditional copulas (see Section 3.4).[19]

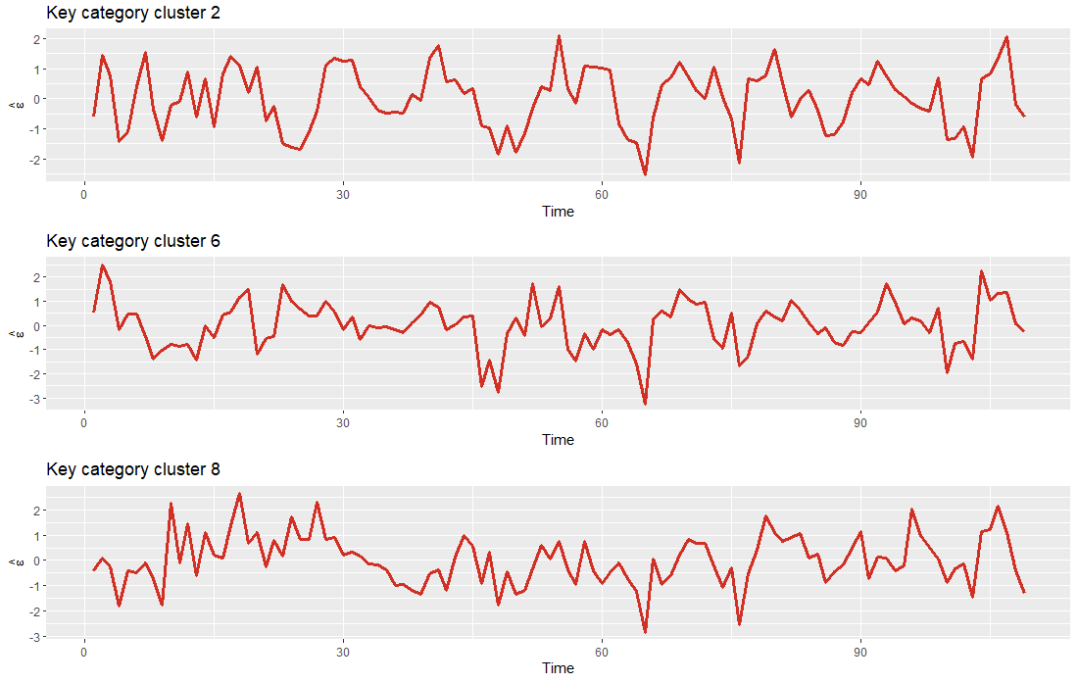


Figure 5.16: Estimated residuals of GAMLSS fits for the three key category clusters

R output 5.7: AIC values of distribution fits on KCC with *fitDist()* function of the *gamlss* package

```
[1] "Fit distribution to KCC 2"
      PE      PE2     SEP1    SHASH  SHASHo2   SHASHo     SEP2     SEP3     SEP4      SN2       GT       NO
311.5332 311.5332 311.6930 311.7580 311.8708 311.8708 311.8838 312.0426 312.2981 313.0165 313.5115 313.5151

[1] "Fit distribution to KCC 6"
      LO       PE      PE2       TF      TF2    SHASH       NO     SEP4  SHASHo2   SHASHo     SEP2     SEP1
310.7827 312.5686 312.5686 313.0641 313.0641 313.5135 313.5174 313.6367 313.6676 313.6676 313.6834 313.8563

[1] "Fit distribution to KCC 8"
      NO       LO      PE2       PE      TF2       TF      SN2      SN1   exGAUS     SEP2     SEP1     SEP3
312.9735 313.2301 314.5925 314.5925 314.6209 314.6209 314.7309 314.8506 314.8713 315.5988 315.7642 316.1121
```

In the previous Section 5.1, the model fits resulted in quantile residuals that we were able to approximate parametrically by normal distributions. When the three sets of residuals are passed into the *fitDist()* function of the *gamlss* package, it returns multiple suggestions of parametric distribution that can be used to fit to the data, ordered by ascending

[19] Currently, the *GJRM* package supports trivariate copula functionalities for binary responses only [Marra and Radice, 2020].

Akaike Information Criterion (AIC). An excerpt is given in R output 5.7. The only distributions that are also at the disposal of the *gjrm* package[20] are the normal and the logistic distributions and the AIC values are close, indicating that both are good fits. Thus, the normal distribution is chosen for all three residual sets in the following.

5.2.1 Key Category Clusters 2 & 6

To start things off, a look at the scatterplot in Figure 5.17 between the residuals of KCC 2 & KCC 6 provides a first hint of a potential correlation structure. For the most part, the data look quite elliptically scattered. Thus, a normal or a t-copula might be appropriate copulas to be used.

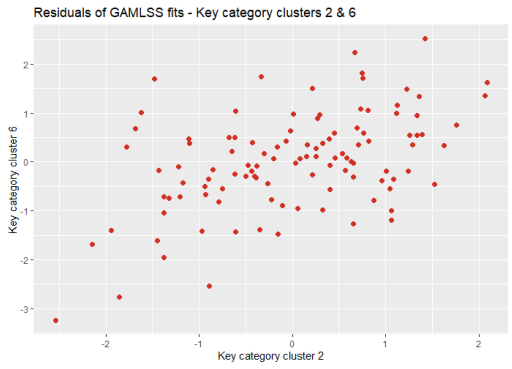


Figure 5.17: Scatterplot of estimated residuals of GAMLSS fits for key category clusters 2 & 6

With the help of the function *BiCopSelect()* of the R package *VineCopula* [Nagler et al., 2019], we can select an appropriate bivariate copula family for the data. A set of possible copula families is considered and maximum likelihood estimations for each are carried out. Selection of a bivariate copula is based on the lowest AIC. The normal and the t-copula are the first in the family set. The selected copula is a t-copula with 3.17[21] d.o.f. and the Pearson's rho as well as Kendall's tau correlation coefficients coincide with the empirical sample coefficients of 0.49 and 0.3 respectively.

Furthermore, to estimate the copula parameter flexibly over time, the Generalized Joint Regression Models (GJRM) framework introduced by Marra and Radice [2016] will be utilized[22]. Essentially, the scope of the GAMLSS framework is extended to a bivariate copula additive model, where all parameters of the margins as well as the copula parameter

[20] Elaboration on this in the following subsections.

[21] Estimation returns degrees of freedom as positive real values above 2 instead of just integer values.

[22] See Section 3.4 as well as the referred paper.

can be estimated simultaneously using structured additive predictors. The parameters are estimated within a penalized likelihood framework using a trust region algorithm with integrated automatic smoothing parameter selection (More details can be found in the referenced literature). This framework has been implemented in R by Marra and Radice [2020] within the scope of the *gjrm* package.

With the GAMLSS fit residuals of the key category clusters 2 & 6 acting as marginal inputs, normal distributions for both and a t-copula with the above mentioned degrees of freedom as starting value, the formula specification is as follows:

$$\begin{aligned} \mu_{KCC2} &= \beta_{\mu,KCC2} \qquad \mu_{KCC6} = \beta_{\mu,KCC6} \\ log(\sigma_{KCC2}) &= \beta_{\sigma,KCC2} \qquad log(\sigma_{KCC6}) = \beta_{\sigma,KCC6} \\ log(\nu) &= \beta_{\nu} \\ tanh^{-1}(\theta) &= \beta_{\theta} + \textit{promo\_type} + f(\textit{time}), \end{aligned} \tag{5.2}$$

where ν denotes the degrees of freedom of the t-copula, $tanh^{-1} =$ is the inverse hyperbolic tangent function[23] relating the expected value of the copula parameter $\hat{\theta}$ to the additive predictor $\beta_\theta + f\left(t_{time}\right)$ and f being a smooth function build upon thin plate regression splines and 30 equidistant knots. Note that as the chosen copula is a Student's t-copula, θ equates Pearson's correlation coefficient ρ. The rest of the model parameters are all set as constant values as the model seems to be robust to any changes including covariates. Incorporating covariates within smooth functions with high numbers of knots on the contrary lead to less reliable estimations (especially due to increasing number of parameters compared to this sample size). The promo types don't seem to be significant for the correlation structure.

Model 5.2 along with the mentioned prespecified settings in the *gjrm()* function generates a good overall fit, which is confirmed when checking the histogram and the QQ-Plots of the quantile residuals for the two margins (see Figure 5.18). A summary can be found in R output 5.8.

[23] Defined as $\tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$ on the domain $\{x \in \mathbb{R} : -1 < x < 1\}$.

R output 5.8: Summary of GJRM fit on key category clusters 2 & 6

```
COPULA:   Student-t (dof = 249)
MARGIN 1: Gaussian
MARGIN 2: Gaussian

EQUATION 1
Link function for mu.1: identity
Formula: res_gamlss_kcc_2 ~ 1

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.15455    0.09825  -1.573    0.116


EQUATION 2
Link function for mu.2: identity
Formula: res_gamlss_kcc_6 ~ 1

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.01163    0.08804   0.132    0.895


EQUATION 3
Link function for sigma.1: log
Formula: ~1

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.05296    0.07928   0.668    0.504


EQUATION 4
Link function for sigma.2: log
Formula: ~1

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.02427    0.06745   -0.36    0.719


EQUATION 5
Link function for dof: log(. - 2)
Formula: ~1

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   86.207      3.044   28.32   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


EQUATION 6
Link function for theta: atanh
Formula: ~promo_type + s(time_obs, k = 30)

Parametric coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                 0.8820     0.1198   7.360 1.83e-13 ***
promo_typeBlack Friday     -0.8718     0.7352  -1.186    0.2357
promo_typeFriends & Family -0.6513     0.3807  -1.711    0.0871 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Smooth components' approximate significance:
              edf Ref.df Chi.sq  p-value
s(time_obs) 20.44  23.79  63.55 1.81e-05 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

sigma.1 = 1.05(0.917,1.24)  sigma.2 = 0.976(0.844,1.09)
dof = 249(249,249)
theta = 0.486(-0.0467,0.82)  tau = 0.371(-0.0434,0.679)
n = 109  total edf = 28.4
```

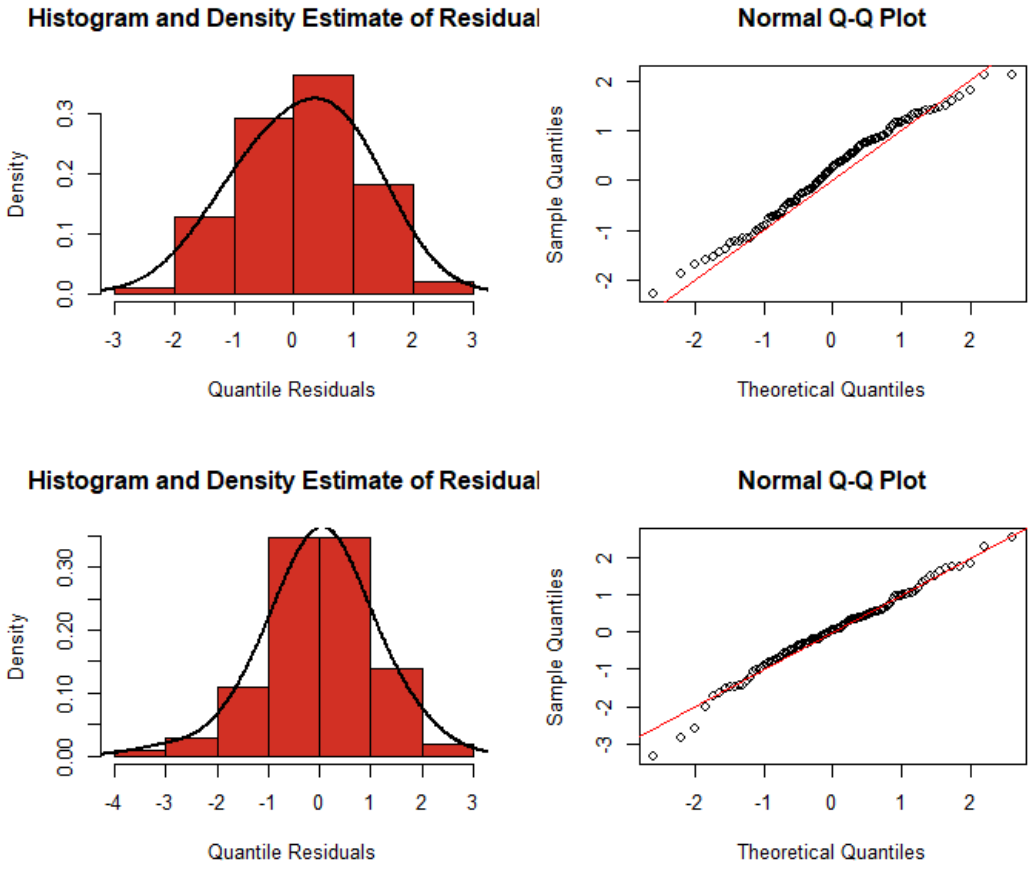


Figure 5.18: Diagnostic plots of quantile residuals based on GJRM models for key category clusters 2 & 6

The effect of time on the copula parameter turns out to be significant (R output 5.8 and can be viewed in Figure 5.19, where the number 26.64 in brackets on the y-axis denotes the effective degrees of freedom of the smooth curve, which is centered around zero due to identifiability constraints.

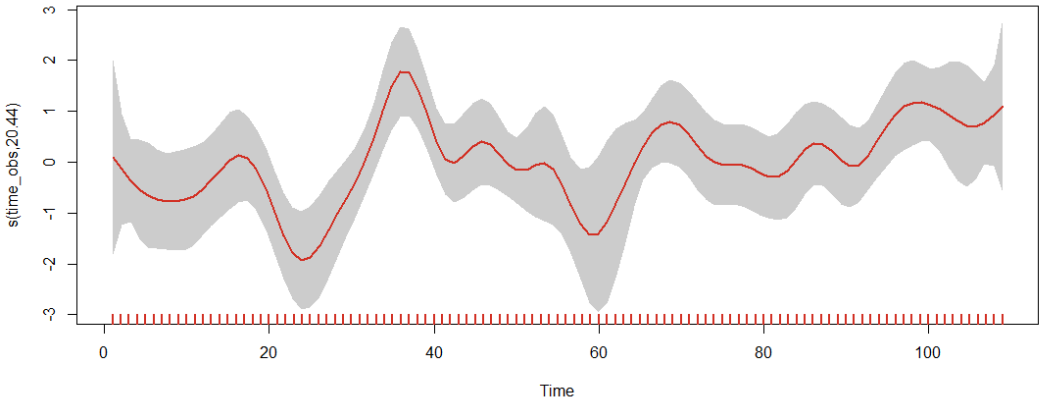


Figure 5.19: Estimated Smooth effect of time on the copula parameter θ with 95% confidence bands for key category clusters 2 & 6

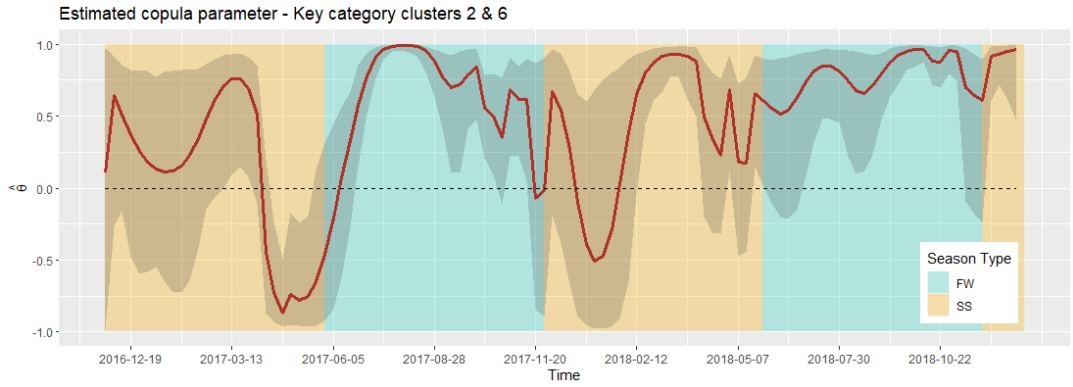


Figure 5.20: Estimated copula parameter over time and 95% confidence bands from a GJRM t-copula model with normal margins for key category clusters 2 & 6 - Season type in the background - Dashed horizontal line at $\hat{\theta} = 0$

Figure 5.20 displays the correlation between the two margins which highly fluctuates between (almost) perfect negative and positive dependence, suggesting that only copulas which can account for both of those dependence (concordance) types should be considered. Candidate copulas were the Frank copula, the Ali-Mikhali-Haq copula and Fari-Gumbel-Morgenstern copula, where the last two can only account for weak dependencies [Marra and Radice, 2016]. The Student's t-copulas provides the most appropriate fit among those types of copulas so we stick to it. The width of the confidence intervals also does vary a lot across different parts of the observation period.

Positive correlation in a business context would mean that for both clusters sales are increasing conjointly. Negative correlations (below dashed line of Figure 5.20) indicate that as one's cluster's sales increase the other one's drop and further questioning should

be carried out. We see such negative correlations during the Spring-Summer seasons and somewhat during Black Friday towards the end of Fall-Winter 2017.

5.2.2 Key Category Clusters 2 & 8

The same strategy as in the previous Subsection 5.2.1 shall be followed. In Figure 5.21 a lower tail dependence can be suspected. Whether both positive and negative dependence are hiding in the the data will be tested beforehand, so that we can narrow down the number of plausible copula families.

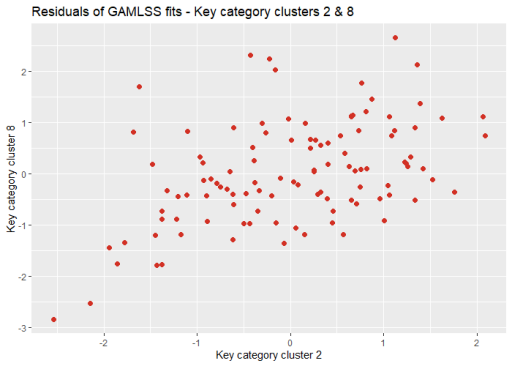


Figure 5.21: Scatterplot of estimated residuals of GAMLSS fits for key category clusters 2 & 8

A GJRM approach equivalent to Model 5.2 will be applied here, with the only difference that we leave out the promo type (see Model 5.3) since it not only has a non-significant effect, but unlike for the previous pair it also heavily distorts the outcome for this pair.

$$
\begin{aligned}
\mu_{KCC2} &= \beta_{\mu,KCC2} \qquad \mu_{KCC6} = \beta_{\mu,KCC6} \\
log(\sigma_{KCC2}) &= \beta_{\sigma,KCC2} \qquad log(\sigma_{KCC6}) = \beta_{\sigma,KCC6} \\
log(\nu) &= \beta_{\nu} \\
tanh^{-1}(\theta) &= \beta_{\theta} + f(\textit{time})
\end{aligned}
\tag{5.3}
$$

We can see the summary in R output 5.9. All model parameters but the copula parameter are set as constants. We first examine copulas which account for positive as well as negative dependencies with the help of the *BiCopSelect()* function. The t-copula is again the most appropriate within a set of copula families including the independence copula, the Gaussian copula, the Student's t-copula and the Frank copula. Looking at equation 6 of the output, we can see that time has a significant effect on $\hat{\theta}$ and its partial smooth effect can be checked visually in Figure 5.23.

R output 5.9: Summary of GJRM fit on key category clusters 2 & 8

```
1
2  COPULA:   Student-t (dof = 249)
3  MARGIN 1: Gaussian
4  MARGIN 2: Gaussian
5
6  EQUATION 1
7  Link function for mu.1: identity
8  Formula: res_gamlss_kcc_2 ~ 1
9
10 Parametric coefficients:
11             Estimate Std. Error z value Pr(>|z|)
12 (Intercept) 0.008837   0.098263    0.09    0.928
13
14
15 EQUATION 2
16 Link function for mu.2: identity
17 Formula: res_gamlss_kcc_8 ~ 1
18
19 Parametric coefficients:
20             Estimate Std. Error z value Pr(>|z|)
21 (Intercept)  0.08536    0.09780   0.873    0.383
22
23
24 EQUATION 3
25 Link function for sigma.1: log
26 Formula: ~1
27
28 Parametric coefficients:
29              Estimate Std. Error z value Pr(>|z|)
30 (Intercept) -0.005018   0.066775  -0.075     0.94
31
32
33 EQUATION 4
34 Link function for sigma.2: log
35 Formula: ~1
36
37 Parametric coefficients:
38             Estimate Std. Error z value Pr(>|z|)
39 (Intercept) -0.05304    0.06315   -0.84    0.401
40
41
42 EQUATION 5
43 Link function for dof: log(. - 2)
44 Formula: ~1
45
46 Parametric coefficients:
47             Estimate Std. Error z value Pr(>|z|)
48 (Intercept)   16.952      3.879    4.37 1.24e-05 ***
49 ---
50 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
51
52
53 EQUATION 6
54 Link function for theta: atanh
55 Formula: ~s(time_obs, k = 30)
56
57 Parametric coefficients:
58             Estimate Std. Error z value Pr(>|z|)
59 (Intercept)   0.6381     0.1096   5.825 5.72e-09 ***
60 ---
61 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
62
63 Smooth components' approximate significance:
64               edf Ref.df Chi.sq p-value
65 s(time_obs) 18.5  21.75  41.45 0.00681 **
66 ---
67 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
68
69 sigma.1 = 0.995(0.896,1.13)  sigma.2 = 0.948(0.837,1.06)
70 dof = 249(249,249)
71 theta = 0.466(-0.151,0.847)  tau = 0.34(-0.112,0.684)
72 n = 109  total edf = 24.5
```

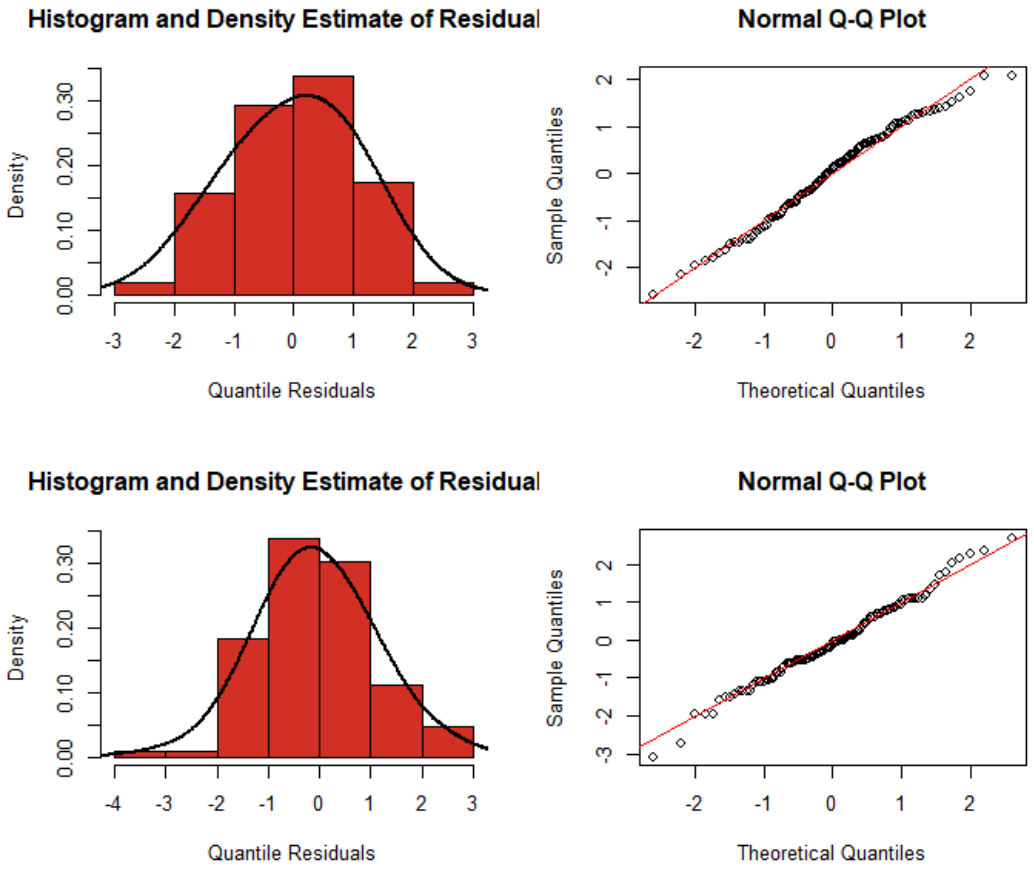


Figure 5.22: Diagnostic plots of quantile residuals based on GJRM models for key category clusters 2 & 8

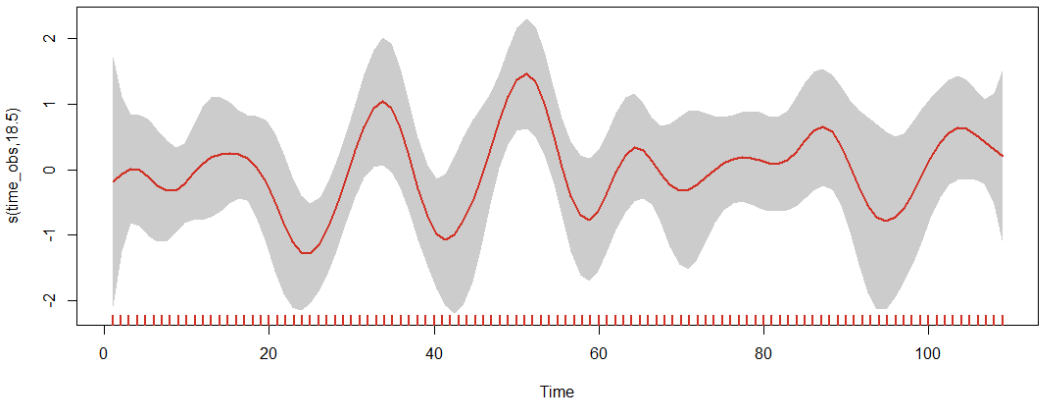


Figure 5.23: Estimated Smooth effect of time on the copula parameter θ with 95% confidence bands for key category clusters 2 & 8

The parsimonious model setup produces a nice fit (see Figure 5.22) so we proceed to analyzing the correlation parameter. Once again, $\hat{\theta}$ takes on positive as well as negative values interchangeably over time periods. Therefore considering further copula families like the Clayton copula would be arguably inappropriate.

Looking at the correlation structure in Figure 5.24, we can see that during both mid-seasons of Fall-Winter 2017 and 2018 there is some negative dependence between the two clusters. End of Spring-Summer 2017 and beginning of Spring-Summer 2018 also show some negative correlation.

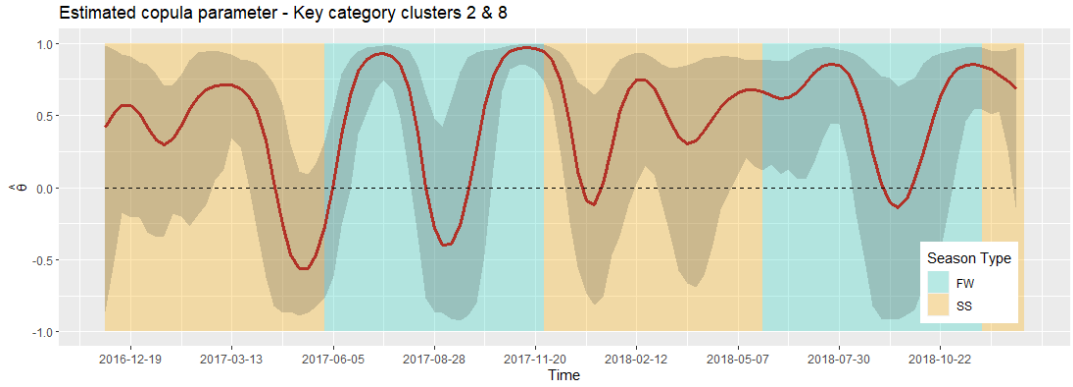


Figure 5.24: Estimated copula parameter over time and 95% confidence bands from a GJRM t-copula model with normal margins for key category clusters 2 & 8 - Season type in the background - Dashed horizontal line at $\hat{\theta} = 0$

5.2.3 Key Category Clusters 6 & 8

This pair of clusters also let's us speculate whether tail dependence exists (see Figure 5.25).

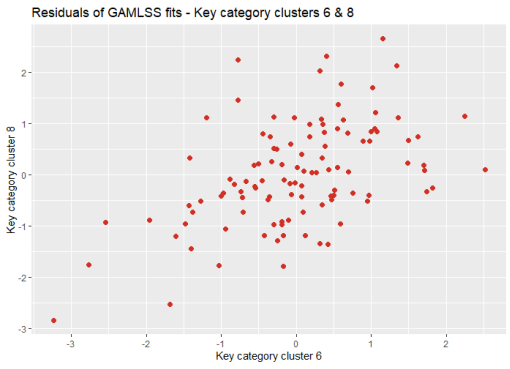


Figure 5.25: Scatterplot of estimated residuals of GAMLSS fits for key category clusters 6 & 8

This time we fast forward to the outcome when a t-copula is applied (which is the proposed family of copulas allowing both concordance directions) under similar conditions

as in Model 5.3 (the reasons for leaving out promo type being the same as in Subsection 5.2.2). Interestingly, the correlation parameter seems to be purely non-negative[24] (see Figure 5.26).

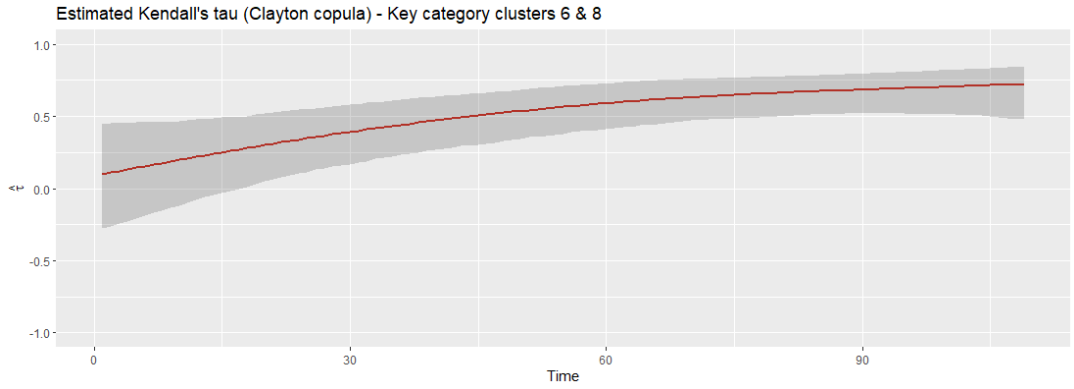


Figure 5.26: Estimated copula parameter over time from a GJRM t-copula model with normal margins for key category clusters 6 & 8 with 95% confidence bands

Keeping the complexity of the model to a minimum, the model setup will remain as is, with the only minor differences. Other copula families will be considered in addition, particularly those that account for positive (and not negative) dependencies. Out of all implemented copula families in the *VineCopula* package (rotated versions included), the Clayton copula yields the most fitting family for the given target variables. This result is sensible, as the Clayton copula is able to capture lower tail dependencies (Figure 5.25) and accounts for positive dependence structures only (see Subsections 3.2.3 and 3.3.3). Note that the new model specification is

$$\begin{aligned} \mu_{KCC6} &= \beta_{\mu,KCC6} \qquad & \mu_{KCC8} &= \beta_{\mu,KCC8} \\ log(\sigma_{KCC6}) &= \beta_{\sigma,KCC6} \qquad & log(\sigma_{KCC8}) &= \beta_{\sigma,KCC8} \\ log(\theta) &= \beta_\theta + f\left(t_{time}\right), \end{aligned} \tag{5.4}$$

where now there are no degrees of freedom to determine and the copula parameter receives a logarithmic link function to ensure that $\hat{\theta}$ is well defined on the non-negative real line. Equation 5 in R output 5.10 shows the significance of time on $\hat{\theta}$ and its smooth effect is visualized in Figure 5.28. The diagnostic plots of quantile residuals in Figure 5.27

[24] Without considering the lower confidence bounds in the first part of the time window.

demonstrate the proper fit qualities once more.

R output 5.10: Summary of GJRM fit on key category clusters 6 & 8

```
1
2  COPULA:   Clayton
3  MARGIN 1: Gaussian
4  MARGIN 2: Gaussian
5
6  EQUATION 1
7  Link function for mu.1: identity
8  Formula: res_gamlss_kcc_6 ~ 1
9
10 Parametric coefficients:
11             Estimate Std. Error z value Pr(>|z|)
12 (Intercept) -0.07752    0.08888  -0.872    0.383
13
14
15 EQUATION 2
16 Link function for mu.2: identity
17 Formula: res_gamlss_kcc_8 ~ 1
18
19 Parametric coefficients:
20             Estimate Std. Error z value Pr(>|z|)
21 (Intercept)  0.03416    0.12033   0.284    0.776
22
23
24 EQUATION 3
25 Link function for sigma.1: log
26 Formula: ~1
27
28 Parametric coefficients:
29             Estimate Std. Error z value Pr(>|z|)
30 (Intercept) -0.03599    0.06464  -0.557    0.578
31
32
33 EQUATION 4
34 Link function for sigma.2: log
35 Formula: ~1
36
37 Parametric coefficients:
38             Estimate Std. Error z value Pr(>|z|)
39 (Intercept)  0.02258    0.06529   0.346    0.729
40
41
42 EQUATION 5
43 Link function for theta: log
44 Formula: ~s(time_obs, k = 30)
45
46 Parametric coefficients:
47             Estimate Std. Error z value Pr(>|z|)
48 (Intercept)   -8.238      3.765  -2.188   0.0287 *
49 ---
50 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
51
52 Smooth components' approximate significance:
53                edf Ref.df Chi.sq p-value
54 s(time_obs) 21.88  23.85   1560  <2e-16 ***
55 ---
56 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
57
58 sigma.1 = 0.965(0.842,1.11)  sigma.2 = 1.02(0.892,1.18)
59 theta = 3.4(1.47,44.2)  tau = 0.342(0.145,0.771)
60 n = 109  total edf = 26.9
```

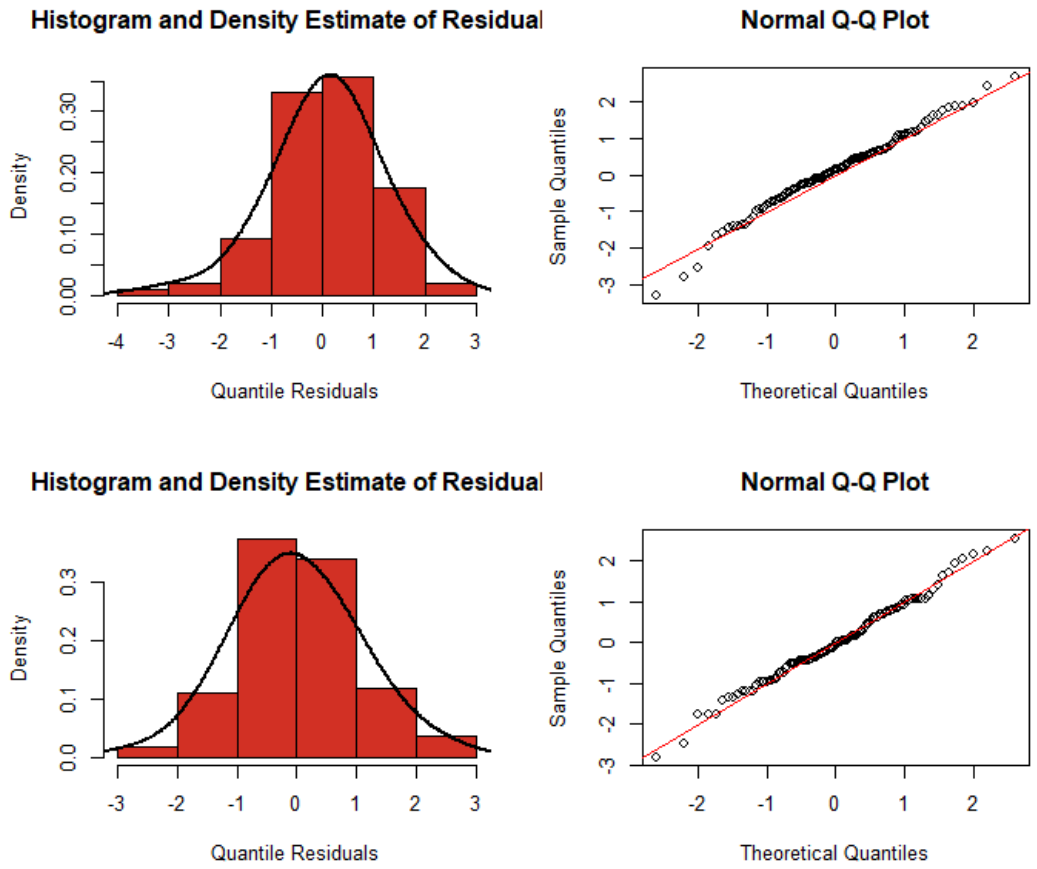


Figure 5.27: Diagnostic plots of quantile residuals based on GJRM models for key category clusters 6 & 8

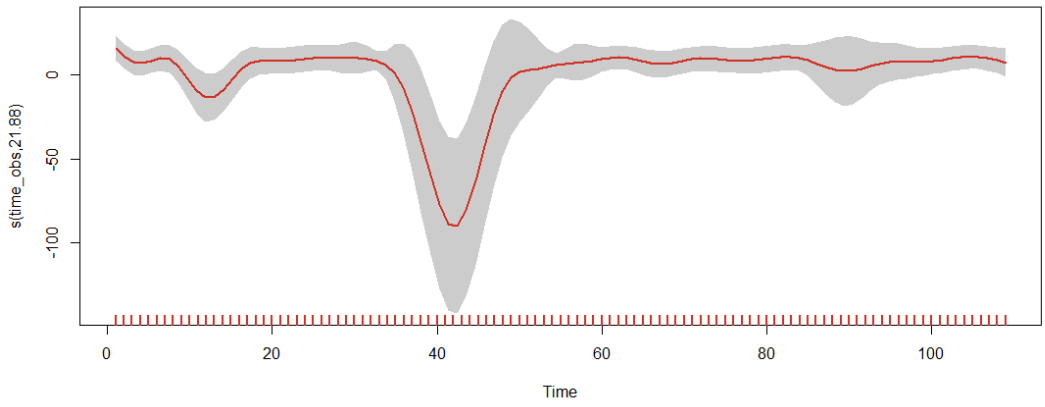


Figure 5.28: Estimated Smooth effect of time on the copula parameter θ with 95% confidence bands for key category clusters 6 & 8

Since Kendall's tau is more intuitive as a dependence measure of two random variables than the copula parameter θ (see Subsection 3.3.2), this is what we will inspect. The relationship between Kendall's tau and the parameter of the Clayton copula is $\tau = \frac{\theta}{\theta+2}$ (see Table 3.1). The estimated time-varying $\hat{\tau}$ is displayed in Figure 5.29, where we can

observe that there are some individual time windows of zero correlation.

Any model configuration of this pair regarding covariates and copulas produce such kind of strictly non-negative dependence. These results should be interpreted with extraordinary caution if at all.

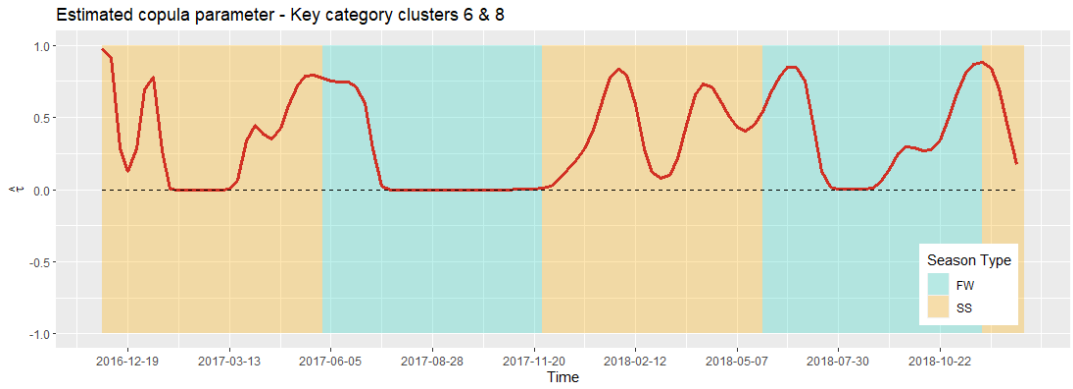


Figure 5.29: Estimated Kendall's tau over time and 95% confidence bands from a GJRM Clayton copula model with normal margins for key category clusters 6 & 8 - Season type in the background - Dashed horizontal line at $\hat{\theta} = 0$

5.3 Article Dependencies

Dependence structures over time for aggregate sets of sales data have been analyzed so far and measures of dependence (or concordance) were sought after, usually in the context of correlation with values between -1 and 1 indicating the magnitude. Parametric approaches were performed to obtain comprehensible results. However, entering the lowest level of granularity of target objects, namely the individual articles themselves (see Figure 4.1), there are some heavy drawbacks as discussed in Section 4.3.

Within the scope of Section 5.2, an important aspect needs to be taken into consideration. That is the well known fact that correlation does not imply causation. Therefore, alternative frameworks which enable us to detect causal effects shall be briefly discussed in this Section and serve as possible directions for further research.

To illustrate an application of a dynamic Bayesian network (DBN, see Section 2.7) with the help of the R package *dbnR*, which is an implementation of Gaussian dynamic Bayesian networks (GDBN) structure learning and inference based on Marco Scutari's package bnlearn [Scutari, 2010], we will delimit ourselves to key category cluster 1 since it contains only 14 distinct articles (see Table 5.7) and assures clear (graphical) overview. For details

on the implementation of the structure learning algorithm, one can refer to the package documentation.

10327	10328	13450	13451	19	19137	19138	21	24558	26191	26192	26193	26194	615

Table 5.7: Articles in key category cluster 1

The *dbnR* package allows Markovian orders higher than 1, i.e. VAR(p) processes with $p \geq 1$, and an optional setting for forbidding arcs between nodes. For this exercise, such constraints will be dismissed and the structure of a Markovian DBN of order 2 will be learned by the algorithm. R output 5.11 displays a summary of the net's structure. Note that the latest (current) time-point is denoted by t_0 in this context, so t_1 is one lag and t_2 are two lags. Graphical representations are depicted in Figures 5.30 and 5.31, the latter highlighting the node of article with ID "10328" at time-point t_0 all adjacent nodes.

R output 5.11: Learned structure of a DBN from articles in KCC 1

```
1
2   Bayesian network learned via Hybrid methods
3
4   model:
5         [10327_t_0][10327_t_1][10327_t_2][13450_t_2][19137_t_2][26194_t_2]
6         [13450_t_1|13450_t_2][19137_t_1|19137_t_2][26194_t_1|26194_t_2]
7         [10328_t_2|10327_t_2][13451_t_2|13450_t_2][19138_t_2|19137_t_2]
8         [615_t_2|26194_t_2][19137_t_0|19137_t_1][26194_t_0|26194_t_1]
9         [10328_t_1|10327_t_1:10328_t_2][13451_t_1|13450_t_1:13450_t_2:13451_t_2]
10        [19138_t_1|19137_t_1][615_t_1|26194_t_1:615_t_2][26193_t_2|19138_t_2]
11        [10328_t_0|10327_t_0:10328_t_1:10328_t_2][19138_t_0|19137_t_0:19138_t_2]
12        [615_t_0|26194_t_0:615_t_1:615_t_2][26193_t_1|19138_t_1:19138_t_2]
13        [19_t_2|26193_t_2][26193_t_0|19138_t_0:19138_t_1][19_t_1|26193_t_1]
14        [21_t_2|19_t_2][19_t_0|26193_t_0:19_t_1][21_t_1|19_t_1]
15        [24558_t_2|10328_t_2:21_t_2][26191_t_2|19_t_2:21_t_2][21_t_0|19_t_0]
16        [24558_t_1|10328_t_1:21_t_1][26191_t_1|19_t_1:21_t_1:26191_t_2]
17        [26192_t_2|10328_t_2:13450_t_2:24558_t_2][13450_t_0|13451_t_2:26192_t_2]
18        [24558_t_0|10328_t_0:21_t_0][26191_t_0|19_t_0:21_t_0:26191_t_1]
19        [26192_t_1|10328_t_1:13450_t_1:24558_t_1:26191_t_2:26192_t_2]
20        [13451_t_0|13450_t_0:13450_t_1:13451_t_1]
21        [26192_t_0|10328_t_0:13450_t_0:24558_t_0:26192_t_1]
22
23  nodes:                                42
24  arcs:                                 68
25    undirected arcs:                    0
26    directed arcs:                      68
27  average markov blanket size:          5.00
28  average neighbourhood size:           3.24
29  average branching factor:             1.62
30
31  learning algorithm:                   Two-Phase Restricted Maximization
32  constraint-based method:              Semi-Interleaved HITON-PC
33  conditional independence test:        Pearson's Correlation
34  score-based method:                   Hill-Climbing
35  score:                                BIC (Gauss.)
36  alpha threshold:                      0.05
37  penalization coefficient:             2.336414
38  tests used in the learning procedure: 1720
39  optimized:                            TRUE
```

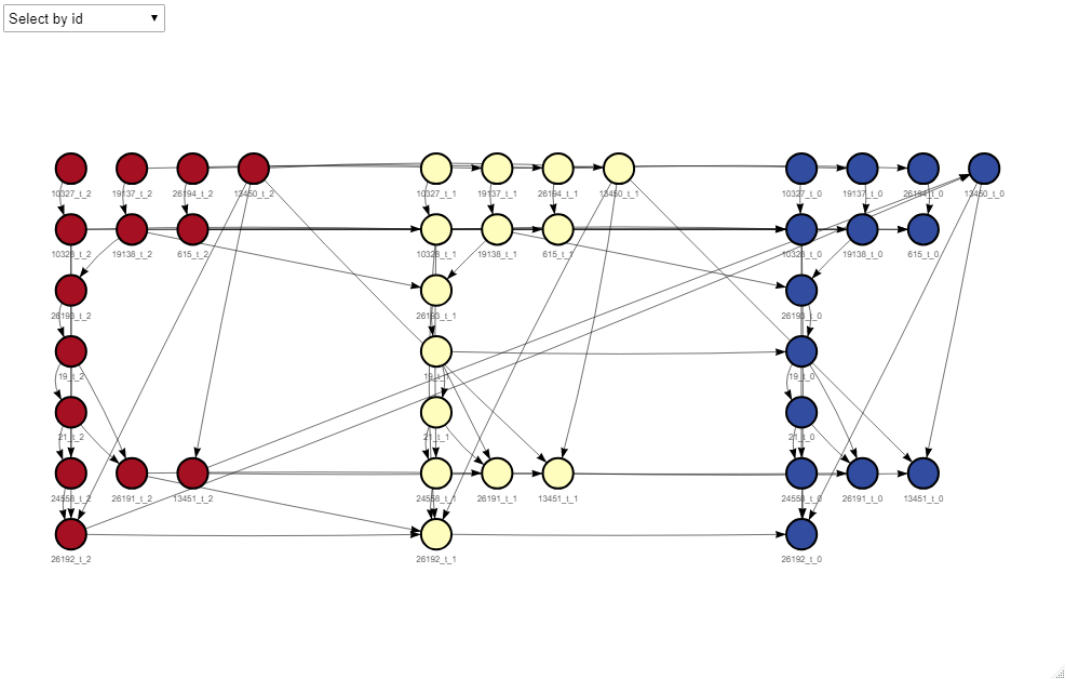


Figure 5.30: Graphical representation of the dynamic Bayesian network for articles in key category cluster 1 - The blue nodes indicate articles at the current time-point t_0, the yellow nodes indicate the same articles at one previous time point t_1 and the red nodes indicate the articles at two previous time points t_2

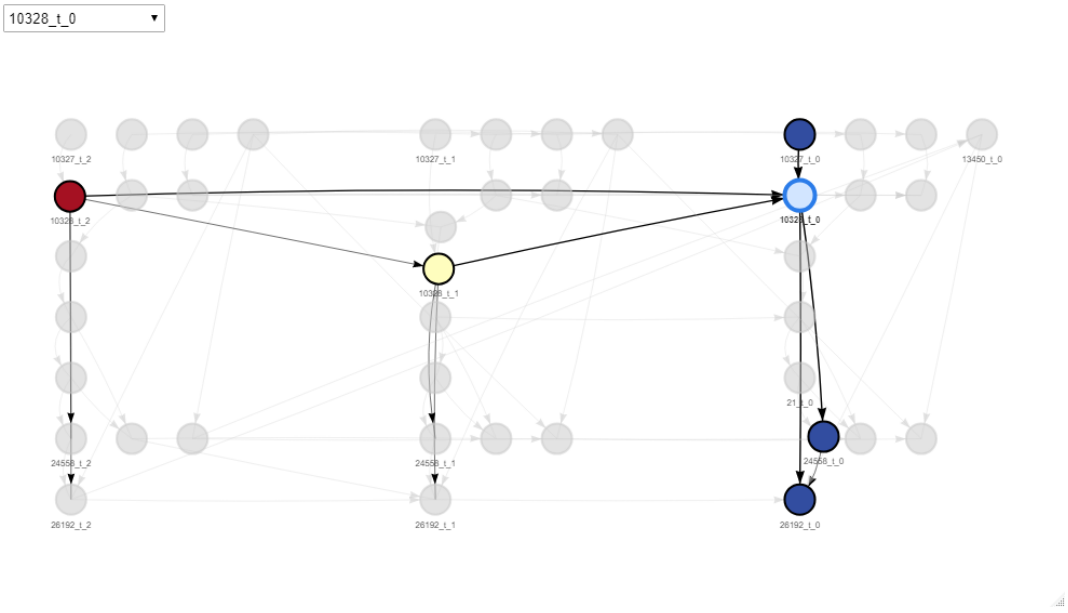


Figure 5.31: Dynamic Bayesian network graph of KCC 1 highlighting article "10328" at latest time-point t_0 and all of its adjacent nodes

With DBNs, causal effects of various sets of data can be estimated and the challenges of poor data quality or insufficient data, like discussed in Section 4.3, can be overcome.[25] Directed acyclic graphs of other data subsets can be produced to assess conditional

[25] Uusitalo [2007] gave an interesting read about this topic.

probabilities between sales of distinct articles or other objects. For instance, one might delimit the data such that only running shoes across multiple clusters (e.g. business units) can be compared, which would be arguably legitimate. Although DBNs capture causal effects, which can be interpreted in terms of product cannibalization, one drawback is that sales transferability[26] cannot be measured (which also constitutes an interesting research question). Another point of interest would be the involvement of explanatory variables, such as the associated weekly price tags or promotion intensities of the articles.

[26] i.e. The actual quantity flow of unit sales of one article towards another across successive time-points.

6 Conclusion

The objective of this thesis was to identify interdependent relationships between product sales based on adidas eCom transactions. In order to spot such dependencies, we followed versatile strategies on different data formats and concluded that different techniques resulted in different concepts of dependence, like correlation structures or causal connections.

After getting acquainted with the weekly summarized data, which have been extracted from online transaction and grouped by the article ID's, we thoroughly deep dived into the patterns appearing in the data mainly by visual and analytical exploration. Despite thousands of articles being involved, we managed to divide the data exploration into three main parts; the global demand of the brand's products, the key category clusters of interest as well as the individual articles, where a sample was drawn to emphasize some obstacles linked to restrictions in the data. The key findings of this data exploration were related to promotion activities and markdowns. Especially Black Friday periods and elevated markdowns turned out to be driving sales in spite of the excessively persisting noise.

Beginning with the key category clusters, we engaged in the modelling part by taking a two-step approach. The reason being that a thorough analysis on the marginal distributions , i.e. the (log-scaled) unit sales for each of the three targeted clusters, was desired before moving forward with the dependence structures. However, a distribution family with readily interpretable parameters (including a necessary skewness parameter) is not yet available in the *GJRM* package.
Thus, the first step of this task was to implement GAMLSS models to the clusters' log-sales using ex-Gaussian families for the responses. We additionally included covariates that stood out during earlier analyses and would likely induce significant effects on the parameters. Some interesting but peculiar contrasts between markdowns and promotion activities arised, leaving data quality topics open for discussion.
The residuals from the models in the first step were resorted to act as normally distributed responses in a second step, were the temporal correlation structures between pairs of clusters were determined via the GJRM framework. By and large, adequateness of the model fits were backed by the diagnostics based on quantile residuals. Nonetheless, due

to the data quality points addressed throughout the paper, the results shall not be taken for granted but rather serve as reference indicators for business users to monitor and tackle opposite directions in sales.

As for the individual articles, other kind of questions regarding dependence structures are of more interest. They are predominantly related to cannibalization effects between product sales. Cannibalization effects are counterproductive, because a prerequisite for a healthy business is to not hurt itself by hindering their own products' sales. Thus, such effects ought to be prevented as much as possible.
While exact quantification of inflow and outflow of sales between multiple articles needs more sophisticated approaches and is typically hard to figure out, determining causal effects, for instance within the context of (dynamic) Bayesian networks, is quite possible. The paper is finalized with a superficial example of dynamic Bayesian networks applied to key category cluster 1, showcasing the advantages that DBN algorithms are able to exploit. This thesis shall set the foundations for aiming towards such potential research directions in the future.

List of Abbreviations

AIC Akaike Information Criterion

AR Auto-Regressive

BS Business Segment

CDF Cumulative Distribution Function

DAG Directed Acyclic Graph

DBN Dynamic Bayesian Network

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

MTS Multivariate Time Series

PDF Probability Density Function

RV Random Variable

STAR Structured Additive Regression Model

UTS Univariate Time Series

VAR Vector Auto-Regressive

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-group.com profile] adidas-group.com profile: *adidas-group*. – https://www.adidas-group.com/en/group/profile/, Last accessed on 2020-08-24

[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

[Grushka 1972] Grushka, Eli: Characterization of exponentially modified Gaussian peaks in chromatography. In: *Analytical Chemistry* 44 (1972), Nr. 11, S. 1733–1738

[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

[Nagarajan et al. 2013] Nagarajan, Radhakrishnan ; Scutari, Marco ; Lèbre, Sophie: Bayesian networks in r. In: *Springer* 122 (2013), S. 125–127

[Nagler et al. 2019] Nagler, T. ; Schepsmeier, U. ; Stoeber, J. ; Brechmann, E. C. ; Graeler, B. ; Erhardt, T. ; Almeida, C. ; Min, A. ; Czado, C. ; Hofmann, M. et al.: VineCopula: Statistical Inference of Vine Copulas. In: *R package version* 2 (2019), Nr. 0

[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 2001] Rigby, R. A. ; Stasinopoulos, D. M.: The GAMLSS project: a flexible approach to statistical modelling. In: *New trends in statistical modelling: Proceedings of the 16th international workshop on statistical modelling* Bd. 337 University of Southern Denmark (Veranst.), 2001, S. 345

[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 exam-*

ples. 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

[Scutari 2010] Scutari, Marco: Learning Bayesian Networks with the bnlearn R Package. In: *Journal of Statistical Software* 35 (2010), Nr. 3, S. 1–22

[Shapiro and Wilk 1965] Shapiro, Samuel S. ; Wilk, Martin B.: An analysis of variance test for normality (complete samples). In: *Biometrika* 52 (1965), Nr. 3/4, S. 591–611

[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

[Stasinopoulos et al. 2007] Stasinopoulos, D. M. ; Rigby, Robert A. et al.: Generalized additive models for location scale and shape (GAMLSS) in R. In: *Journal of Statistical Software* 23 (2007), Nr. 7, S. 1–46

[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

[Uusitalo 2007] Uusitalo, Laura: Advantages and challenges of Bayesian networks in environmental modeling. In: *Ecological Modelling* 203 (2007), 05, S. 312–318

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

Statutory Declaration

I hereby declare that I wrote this thesis paper independently, without assistance from external parties, and without use of other resources than those indicated. All information taken from other publications or sources in text or in meaning are duly acknowledged in the text. The written and electronic forms of the thesis paper are the same. I give my consent to have this thesis checked by plagiarism software.

Nuremberg, August 25, 2020

Petros Christanas