University of Cologne

Malte Bonart

Testing for Structural Breaks in Factor Copula Models - Implementation and Application in Social Media Topic Analysis

Master thesis
Supervisor: Prof.Dr. Dominik Wied

Submitted for the Master Examination in Economics at the Faculty of Management, Economics and Social Sciences of the University of Cologne in June 2018.

This is my abstract.

Contents

Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

API	Application programming interface
btw17	Bundestag election 2017
cdf	Cumulative distribution function
CV	Critical value
DGP	Data generating process
EDF	Empirical distribution function
GMM	Generalized Methods of Moments
iid	Independent and identically distributed
ML	Maximum Likelihood
SMM	Simulated methods of moments

1. Introduction

Many multivariate statistical models are based on the assumption that the variables follow a multivariate Gaussian distribution. Here, the natural measure of dependece is the covariance or correlation matrix between the variables (Joe 2015, p. 25). But using simple correlations as the only dependence measure can be misleading. In many cases real world data does not follow a Gaussian distribution and shows other distributional feautures, such as skeweness or heavy tails. Correlation requires that the variance of the marginal distribution is finite. This requirement can be problematic when dealing with heavy tailed distributions. Further more, correlation is only a measure of linear dependence and it is not invariant under monotonic transformations of the variables (Embrechts, Mcneil, and Straumann 2002, pp. 6-8). This means, that by applying a strictly increasing function (for example the log function) to each variable, the correlation matrix changes.

Due to this shortcomings of the ordinary linear correlation coefficient, other and more flexible methods for describing the dependence structure between non Gaussian variables have been developed. In this area of research, models based on so called copula functions became increasingly popular since the 1990th (Nelsen 1999, p. 1). Up to the year 1999, only 38 publications which include the topics *copula* and *dependence* have been published in peer-reviewed journals. From 2000 - 2008 the number of copula related publications was 334 and for the years 2009 - 2017 2048 publications are listed.[1]

At first, any multivariate distribution function for which its marginal distributions have a uniform distribution on the interval $[0, 1]$ can be called a copula function (Nelsen 1999, p. 1). But its popularity is based on the results of a thereom by Sklar (1959): It implies that multivariate distributions can be constructed by separately specifying the marginal distributions of the random variables and by defining the dependence structure among the variables. The dependence structure is described with a copula function. Thus, a copula links the multivariate distribution function to its marginal distributions. In case of continous multivariate distributions the copula function is uniquely defined.

Due to this results, copulas are mainly used in two ways: First, to model the dependence structure of multivariate distributions independent of their under-

---

[1]The numbers are taken from the *Web of Science Core Collection* for the search term `TOPIC: (copula) AND TOPIC: (dependence)`.

lying marginal distributions and second, to construct bivariate or multivariate distributions based on a copula function and marginal distributions (Sempi 2011, pp. 302-304).

This allows for a two-stage estimation process in multivariate models in which the marginal distributions and the copula function is estimated separately. By doing so, semiparametric techniques can be utilized. For example, the marginal distribution can be estimated using the empirical distribution function (EDF) while the copula function is estimated parametrically. (Patton 2009, p. 777).

For time series data copula theory can be used in two ways: First, to describe the cross sectional dependence structure by estimating the conditional copula function of the conditional joint distribution at some timepoint given past information. Second, copulas can be used to describe the dependence between observations of a univariate time series. This is related to the study of Markov processes. (Patton 2009, p. 771 ff).

This thesis summarizes and strucutres the current developement in the field of factor copula models. It contributes to the sicentific disucssion in two ways: First, a software library for a consistent specification, simulation and estimation of factor copula models is presented. With the library, the methods and the strucutral break test are made avalaible to a broader scientific audience and can be used by applied reasearchers. Second, we study novel ways of applying the discussed methods to areas outside of the finance community. The strucutral break test is applied to an aggregated dataset of social media posts from german politicians and political parties.

2. Theoretical foundation

In this chapter we present and summarize the theoretical foundation of this thesis. First, the general idea behind copula functions is introduced. Second, a special class of copula models, the so called *factor copula* model is presented. Third, we discuss copula models in the context of time series data and present a specific framework for modelling dependencies of multivariate time series. Fourth, the *simulated methods of moments* (SMM) is explained. It is the estimation method which is used throughout this work. Finally we summarize the ideas of a structural break test for possibly time varying parameters of a factor copula model.

2.1. Copula theory

The joint cumulative distribution function (cdf) $F_Y(y_1, \ldots, y_N) = P(Y_1 \leq y_1, \ldots, Y_N \leq y_N)$ for some multivariate random vector Y of dimension N has the continous marginal distributions $F_{Y_i}(y_i) = P(Y_i \leq y_i) \; \forall \; i = 1, \ldots, N$. Estimating F_Y is computationally demanding espacialy if N becomes large. Therefore, a copula function is introduced which can be used to link the marginal and the joint cdf.

2.1.1. Definition

A function of the type $C : [0,1]^N \rightarrow [0,1]$, with $N \geq 2$ is called a *copula* if it is the distribution function of a random vector U such that $C_U(u_1, \ldots, u_N) = P(U_1 \leq u_1, \ldots, U_N \leq u_N)$ and if its marginal distributions are $U_i \sim Unif(0,1)$, e.g. uniformly distributed (Joe 2015, p. 7). The thereom by Sklar (1959) states, that every d-variate distribution function $F_Y(y_1, \ldots, y_n)$ can be eypressed in terms of its marginal distributions and a copula function such that

$$F_Y(y_1, \ldots, y_N) = C_U(F_1(y_1), \ldots, F_N(y_N)). \tag{2.1}$$

To see this, consider the so called probability-integral transformation which states that the marginal random variables $U_i = F_{Y_i}(Y_i)$ are uniformly distributed (Embrechts, Mcneil, and Straumann 2002, p. 4). This is due to the fact that $F_{U_i}(u_i) = P(U_i \leq u) = P(F_{Y_i}(Y_i) \leq u_i) = P(Y_i \leq F_{Y_i}^{-1}(u_i)) = F_{Y_i}(F_{Y_i}^{-1}(u_i)) = u_i$ is the distribution function of a $Unif(0,1)$ distributed variable. Using this transformation we can write

$$\begin{aligned} F_Y(y_1, \ldots, y_N) &= P(Y_1 \leq y_1, \ldots, Y_N \leq y_N) \\ &= P(F_{Y_1}(Y_1) \leq F_{Y_1}(y_1), \ldots, F_{Y_N}(Y_N) \leq F_{Y_N}(y_N)) \\ &= P(U_1 \leq F_{Y_1}(y_1), \ldots, U_N \leq F_{Y_N}(y_N)) \\ &= C_U(F_{Y_1}(y_1), \ldots, F_{Y_N}(y_N)). \end{aligned} \tag{2.2}$$

Note that transforming the marginal distributions to a $Unif(0,1)$ distribution is somehow an arbitrary choice and other possibilities, for instance the Gaussian distribution, are also possible. (Mikosch 2006, p. 7)

If F_Y is *continuous* with marginal quantile functions $F_{Y_1}^{-1}, \ldots, F_{Y_N}^{-1}$ then the copula function $C_U(\mathbf{u})$ is uniquely determined by $C(\mathbf{u}) = F(F_{Y_1}^{-1}(u_1), \ldots, F_{Y_N}^{-1}(u_N))$. If not all marginal distribution are continous the copula still exist but in this case it is not unique anymore (Embrechts, Mcneil, and Straumann 2002, pp. 4-5).

The simplest form of a copula function is the copula of a vector of independent variables for wich we can write $F_Y(\boldsymbol{y}) = \prod_{i=1}^N F_{Y_i}(y_i)$. Using the derivation in (2.2), this results in the independence copula

$$C_{ind}(u_1, \ldots, u_N) = \prod_{i=1}^N u_i. \tag{2.3}$$

Many other possible funtional forms for a copula C_U exist. In this paper, we focus on a special class, the so called factor copulas which we present in section 2.2.

2.1.2. Copulas and dependency

Copulas are usefull for the study of dependence between a set of random variables, because the copula function which defines the dependency structure is unaffected by monotonic transformations of the marginal variables. In contrast to the linear correlation coefficient, this property allows the definition of alternative scale-invariant dependence measures (Nelsen 1999, p. 125).

To see the scale-invaraint property of the copula function, consider a random vector U with copula distribution function $C_U(\boldsymbol{u})$ and $C_U(F_1(y_1), \ldots, F_N(y_N)) = F_Y(y_1, \ldots, y_N)$. By applying some increasing function $T_i(Y_i)$ to the random variables and using (2.2) we can write

$$
\begin{aligned}
F_{T(Y)}(\boldsymbol{y}) &= P(T_1(Y_1) \le y_1, \ldots, T_N(Y_N) \le y_N) \\
&= P(T_1(F_{y_1}^{-1}(U_1)) \le y_1, \ldots, T_N(F_{y_N}^{-1}(U_N))) \\
&= C_U(F_{T_1(Y_1)}(y_1), \ldots, F_{T_N(Y_N)}(y_N)).
\end{aligned} \tag{2.4}
$$

Thus, allthough Y and $T(Y)$ have different joint distribution functions, they share the same copula function.

In the following we shortly present three common scale invariant measures of dependency between two variables Y_i and Y_j. In contrast to the linear

correlation coefficient, these measures can be expressed solely as a function of the underlying bivariate copula.[2] Spearman's and Kendall's rank correlation both measure the degree of monotonic dependence between the two variables Y_i and Y_j with joint distribution function $F_Y(\boldsymbol{y})$. As the linear correlation coefficient they are symmetric and normalised in the interval $[0, 1]$ (Embrechts, Mcneil, and Straumann 2002, p. 15).

Kendall's rank correlation is based on the definition of concoradance and discordance: Consider two pairs of independent and identically distributed (iid) random vectors (Y_i^1, Y_j^1) and (Y_i^2, Y_j^2). Both pairs share the same joint distribution F_Y. The pair is concordant if $(Y_i^1 - Y_i^2)(Y_j^1 - Y_j^2) > 0$ and discordant if $(Y_i^1 - Y_i^2)(Y_j^1 - Y_j^2) < 0$. In words, for the former case large (small) values of one pair occur with large (small) values of the other. For the latter case, large (small) values of one pair occur with small (large) values of the other (Nelsen 1999, pp. 125-126).

With this definition Kendalls's rank correlation is given as the probability of concordance minus the probability of discordance:

$$
\begin{aligned}
\tau_{i,j} &= P((Y_i^1 - Y_i^2)(Y_j^1 - Y_j^2) < 0) - P((Y_i^1 - Y_i^2)(Y_j^1 - Y_j^2) > 0) \\
&= 4 \int \int_{[0,1]^2} C(u_i, u_j) dC(u_i, u_j) - 1.
\end{aligned}
\tag{2.5}
$$

Spearman's rank correlation is given as the ordinary correlation coefficient between the probability-integral transforms $F_{Y_i}(Y_i)$ and $F_{Y_j}(Y_j)$:

$$
\rho_{i,j}^S = \rho(F_{Y_i}(Y_i), F_{Y_j}(Y_j)) = 12 \int \int_{[0,1]^2} u_i u_j dC(u_i, u_j) - 3.
\tag{2.6}
$$

Finally, to capture dependencies in the joint lower or joint upper parts of the distribution, one defines the coefficients of upper and lower tail dependency:

$$
\begin{aligned}
\lambda_{i,j}^U &= \lim_{q \to 1} P(Y_i > F_{Y_i}^{-1}(q) | Y_j > F_{Y_j}^{-1}(q)) = \lim_{q \to 1} \frac{1 - 2q + C(q, q)}{1 - q}, \\
\lambda_{i,j}^L &= \lim_{q \to 0} P(Y_i \leq F_{Y_i}^{-1}(q) | Y_j \leq F_{Y_j}^{-1}(q)) = \lim_{q \to 0} \frac{C(q, q)}{q}.
\end{aligned}
\tag{2.7}
$$

---

[2]For a proof of the copula representations see Embrechts, Mcneil, and Straumann (2002, pp. 16-18).

The coefficients measures the probability that extreme large (small) values occur in one variable, given extreme large (small) values in the other variable. In contrast to other common dependence measures, the coefficients of upper and lower tail dependency are defined on the interval $[0, 1]$. Since the limit only exists theoretically and not for observable data, one usually calculates upper and lower *quantile* dependency for some values of q close to 0 and 1 (Joe 2015, pp. 62-63).

2.2. Factor copulas

Factor copulas are a special class of copula models for which the copula function $C_U(u_1, \ldots, u_N)$ is based on a latent factor structure as defined in Oh and Patton (2013) and Oh and Patton (2017).

2.2.1. General set up

Consider a set of artificial variables $X_i, i = 1, \ldots, N$ which linearly depend on some latent factors $Z_k, k = 1, \ldots, K$ and some iid error e_i such that $X_i = \sum_{k=1}^{K} \beta_{ik} Z_k + \epsilon_i$. The linear coefficients β_{ik} are also called factor loadings. The latent variables Z_k and the error term ϵ_i follow some parametrized distributions with parameter vectors $\boldsymbol{\gamma}_\epsilon$ and $\boldsymbol{\gamma}_{Z_k}$ and we write: $\epsilon_i \overset{iid}{\sim} F_\epsilon(\boldsymbol{\gamma}_\epsilon)$ and $Z_k \sim F_{Z_k}(\boldsymbol{\gamma}_{Z_k})$. While the variables X_i usually dependent on each other, the latent factors are independent from each other and from the error term.

As shown in the previous section, the joint probability function $F_X(x_1, \ldots, x_N)$ of the artificial variables can be expressed in terms of its marginal distributions $F_{X_i}(x)$ and a factor copula function $C_U(u_1, \ldots, u_N)$ such that $F_X(x_1, \ldots, x_N) = C_U(F_{X_1}(x_1), \ldots, F_{X_N}(x_N); \boldsymbol{\theta})$.

The artifical variables X_i are only used to construct the factor copula function $C_U(u_1, \ldots, u_N; \boldsymbol{\theta})$. The parameters of the factor structure are chosen in such a way that the resulting copula function fits the copula of the observable data $\boldsymbol{Y}$, such that $F_Y(y_1, \ldots, y_n) = C_U(F_{Y_1}(y_1), \ldots, F_{Y_N}(y_N); \boldsymbol{\theta})$. Once the factor copula function is determined, the artificial variables and its marginal distributions $F_{X_i}(x)$ are of no interest.

The parameters of the factor model are collected in a parameter vector $\boldsymbol{\theta} = (\beta_{11}, \ldots, \beta_{i1}, \ldots \beta_{ik}, \boldsymbol{\gamma}'_{Z_1}, \ldots, \boldsymbol{\gamma}'_{Z_K}, \boldsymbol{\gamma}'_\epsilon)'$. It consists of all linear coefficients and

the distributional parameters of the error term and the latent variables. The number of latent variables K and the distribution functions $F_{Z_1}, \ldots, F_{Z_k}, F_\epsilon$ are hyper-parameters of the model which have to be chosen prior to the estimation.[3]

Using matrix notation, the model can be summarized in the following set of equations:

$$
\begin{aligned}
\boldsymbol{Y} &= (Y_1, \ldots, Y_N)' \\
\boldsymbol{X} &= (X_1, \ldots, X_N)' = \boldsymbol{\beta} \boldsymbol{Z} + \boldsymbol{\epsilon} \\
F_Y(\boldsymbol{y}) &= C_U(F_{Y_1}(y_1), \ldots, F_{Y_N}(y_N); \boldsymbol{\theta}) \\
F_X(\boldsymbol{x}) &= C_U(F_{X_1}(x_1), \ldots, F_{X_N}(x_N); \boldsymbol{\theta})
\end{aligned}
\tag{2.8}
$$

To model the joint probability $F_Y(\boldsymbol{y})$, a two-stage estimation process can be used: First, the marginal distributions $\hat{F}_{Y_i}$ are estimated parametrically or non-parametrically, e.g. by using some parametric model or the EDF. Second, the factor structure for the copula function is fitted to the data by finding the optimal $\hat{\boldsymbol{\theta}}$. In most cases, a closed form of the factor copula does not exist and approximating the likelihood functon is computationally demanding. Therefore, one has to rely on simulation based estimation methods as described in section 2.4.

2.2.2. Two model restrictions

The two-stage estimation aproach allows for a variety of different dependence structures and can be applied to high dimensional data. An upper bound for the number of model parameters $P = |\boldsymbol{\theta}|$ to be estimated is given by the size of the factor matrix and the number of additional free distributional parameters such that $P \leq (N \cdot K + |\boldsymbol{\gamma_{Z_1}}| + \cdots + |\boldsymbol{\gamma_{Z_K}}| + |\boldsymbol{\gamma_\epsilon}|)$. To reduce the number of parameters, Oh and Patton (2017, pp. 148, 150) present two restrictions on the matrix of factor loadings $\boldsymbol{\beta}$: the restrictive *equidependence* and the less restrictive *block-equidependence* model.

For the first model, it is assumed that $K = 1$ and $\boldsymbol{\beta} = (\beta, \ldots, \beta)'$. Thus, the model consists of a single latent factor and a single factor loading β which is

---

[3]Oh and Patton (2017, p. 143ff) provide a heuristic of finding the number of latent variables by analyzing so called *scree-plots*: Ordered eigenvalues from the sample rank-correlation matrix of the data.

the same for all variables. This implies equal pairwise dependencies for all observable variables.
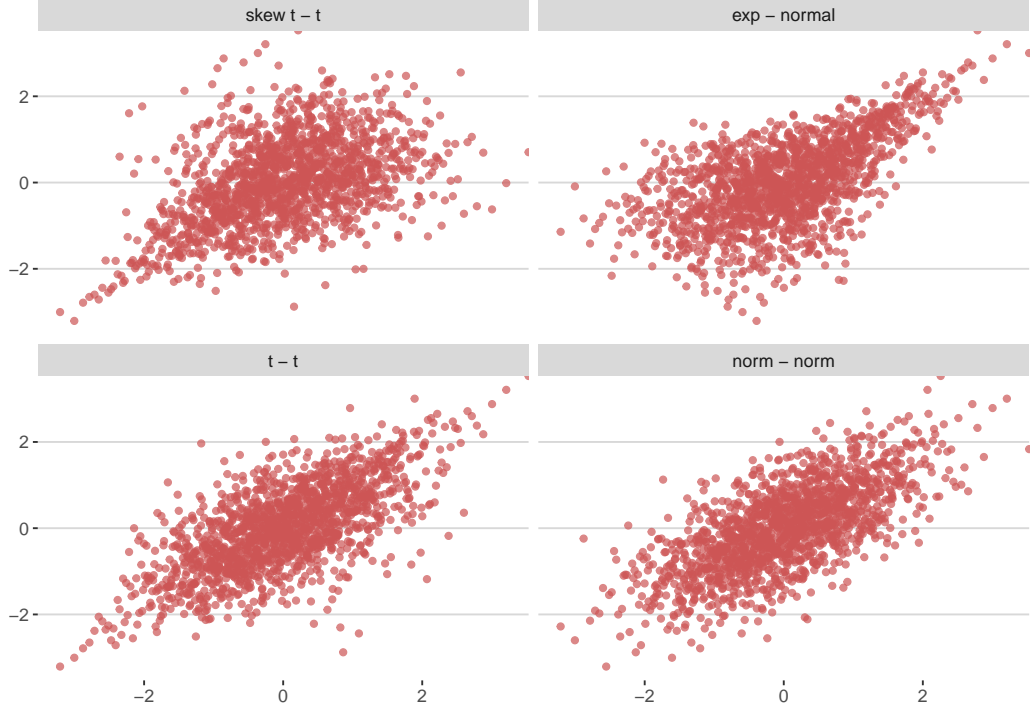
Figure 2.1: Illustration of different equidependence factor copula models with $N = 2$, $\beta = 1.5$, $Y_i \sim N(0, 1)$ and different distributions for the latent variable and the error term.

Figure 2.1 shows four different simulations from a two dimensional one factor equidependence factor copula model. The marginal distributions are standard normal, the linear coefficient is fixed at $\beta = 1.5$ but the distributions of the latent variable and the error term differ. All models produce positive dependence but the symmetry and tail dependency differs through the choice of the distributions and their parameters. The upper left panel are realizations from a factor copula model with a skew-t distributed latent factor and a t-distributed error term. The degrees of freedom are set to $df = 4$ and the skewness parameter to $\lambda = -0.8$. This produces strong assymetric tail dependencies. The factor copula in the bottom left panel has similar distributions. But here symmetric tail dependencies are produced since $\lambda = 0$. This results in an ordinary t-distribution for the latent variable. The bottom right panel shows realizations from a gaussian copula which results in a multivariate gaussian distribution with no tail dependency or assymetry. The last panel shows the combination

of an exponential latent variable with a normal distributed error term. This produces strong upper tail dependencies.

The second restriction, the block-equidependence model, is less restrictive that the equidependence model. It is suitable for higher dimensional problems and for variables which can be naturally partitioned into different groups.[4] The model assumes a common factor for all groups and a group specific factor for each group. Thus, each variable is only affected by two factors. For the matrix of factor loadings, it is further assumed that all variables in the same group have the same factor loading while variables in different groups can have different loadings. This structure implies equal pairwise intra-group dependencies while the pairwise inter-group dependencies can vary between the groups.

Formally, consider a partition of $\boldsymbol{Y} = (Y_1, \ldots, Y_N)'$ into M groups. A single variable can then be written as Y_i^j, where $i = 1, \ldots, N$, $j = 1, \ldots, M$. The value k_j is the number of variables in group j and it holds $\sum_{j=1}^{M} k_j = N$. Then the factor copula model can be summarized as:

$$
\begin{aligned}
\boldsymbol{X} &= (X_1^1, \ldots, X_{k_1}^1, X_{k_1+1}^2, \ldots, X_{k_1+k_2}^2, \ldots, \ldots, X_N^M)' = \boldsymbol{\beta Z} + \boldsymbol{\epsilon} \\
\boldsymbol{Z} &= (Z_0, Z_1, \ldots, Z_M)' \\
X_i^j &= \beta_j Z_0 + \beta_{M+j} Z_j + \epsilon_i \\
\boldsymbol{\beta} &= \begin{pmatrix}
\beta^1 & \beta^{M+1} & 0 & \cdots & 0 \\
\beta^1 & \beta^{M+1} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\beta^1 & \beta^{M+1} & 0 & \cdots & 0 \\
\beta^2 & 0 & \beta^{M+2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\beta^M & 0 & 0 & \cdots & \beta^{M+M} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\beta^M & 0 & 0 & \cdots & \beta^{M+M}
\end{pmatrix},
\end{aligned}
\tag{2.9}
$$

where the matrix $\boldsymbol{\beta}$ is of size $N \times (M + 1)$ but with only $2M$ actual factor loadings.

---

[4]E.g. this could be stock market prices grouped into different industry sectors.

2.3. Copula models for multivariate time series

Up to now, it was assumed, that the data generating process (DGP) for the copula model is static and time invariant. But one can also extent the copula model to univariate or multivariate time series processes. For the former case, the copula is used to estimate the joint distribution of a one dimensional time series $(Y_t, Y_{t+1}, \ldots, Y_{t+n})'$. For the latter, the interest lies in the conditional joint distribution of the time dependent random vector $\boldsymbol{Y_t} = (Y_{1t}, \ldots, Y_{Nt})'$. The conditional cdf can be written as $F_{\boldsymbol{Y_t}|\mathcal{F}_{t-1}}(\boldsymbol{y})$, where the σ-algebra $\mathcal{F}_{t-1}$ possibly contains past information and information from other exogenous variables $\{\boldsymbol{Y}_{t-1}', \boldsymbol{Y}_{t-2}', \ldots, \boldsymbol{X}_t', \boldsymbol{X}_{t-1}', \ldots\}$ (Patton 2012, pp. 4-6).

Sklar's theorem given by (2.1) shows how the copula function links together the marginal and the joint cdf. It can be extended to the multivariate time series case in which the conditional distribution is split into a conditional copula $C_{U_t|\mathcal{F}_{t-1}}(\boldsymbol{u})$ and conditional marginal distributions $F_{Y_{it}|\mathcal{F}_{t-1}}(y)$.

Using the definition of a conditional copula one can write

$$F_{Y_t|\mathcal{F}_{t-1}}(y_1, \ldots, y_N) = C_{U_t|\mathcal{F}_{t-1}}(F_{Y_1t|\mathcal{F}_{t-1}}(y_1), \ldots, F_{Y_Nt|\mathcal{F}_{t-1}}(y_N)). \qquad (2.10)$$

To have a valid conditional multivariate distribution, the conditioning set must be the same for the marginal distributions and the copula (Patton 2009, p. 772).

For the following sections, another approach is choosen. We built on a semi-parametric copula-based multivariate dynamic model as described in Chen and Fan (2006, pp. 126-129). First, the vector of conditional means and variances of $\boldsymbol{Y}_t|\mathcal{F}_{t-1}$ are estimated parametrically. Second, the empirical residuals are calculated by filtering out the conditional means and variances. Third, the joint distribution of the estimated standardized residuals is then modelled using non-parametric estimates of the marginal distributions and a parametric time-invariant copula.

Thus, this model describes three levels of estimation: First, the parameters of the marginal models for the conditional time varying means and variances. Second, the estimation of the marginal distribution of the standardized residuals. This is done using the EDF. Third, the parameters of the copula function for the standardized residuals.

If we denote the parametrized conditional mean of a single variable as $\mu_{it} = E(Y_{it}|\mathcal{F}_{t-1}; \boldsymbol{\phi})$ and the parametrized conditional standard deviation as $\sigma_{it} = \sqrt{V(Y_{it}|\mathcal{F}_{t-1}; \boldsymbol{\phi})}$ we can write the multivariate time series process as:

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\sigma}_t \boldsymbol{\eta}_t \tag{2.11}$$

with $\boldsymbol{\sigma}_t = \text{diag}(\sigma_{1t}, \ldots, \sigma_{Nt})$. The standardized innovations $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{Nt})'$ are assumed to be iid and independent of past information. It follows that $F_{\boldsymbol{\eta}_t} = F_{\boldsymbol{\eta}} \forall t = 1, \ldots, T$. Hence, for this model, the notion of a conditional copula as given by (2.10) is not used. Instead, using the copula approach, the joint cdf of the residuals can be written as:

$$F_{\boldsymbol{\eta}}(u_1, \ldots, u_t) = C_U(F_{\eta_1}(u_1), \ldots, F_{\eta_N}(u_N); \boldsymbol{\theta}). \tag{2.12}$$

The first level parameters $\boldsymbol{\phi}$ of the conditional means and variances are estimated using parametric models for univariate time series, such as models from the GARCH family (Teräsvirta 2009, p. 17). The marginal distributions of the standardized residuals $\hat{\boldsymbol{\eta}}_t = \boldsymbol{\sigma}_t^{-1}(\boldsymbol{Y}_t - \boldsymbol{\mu}_t)$ are estimated with the EDF:

$$\hat{F}_{\eta_i}(u) = \frac{1}{T+1} \sum_{t=1}^{T} \mathbf{1}(\hat{\eta}_{it} \leq u). \tag{2.13}$$

The parametric copula for the innovations is estimated using a factor copula model and the SMM estimation technique which is presented in the following chapter.

2.4. Simulated method of moments estimation for factor copulas

In the following section, based on the time series model as shown in the previous section 2.3 and the factor copula model as defined in section 2.2, an estimation procedure known as Simulated Method of Moments (SMM) is presented.

In models for which the copula function is known in closed form, standard Maximum Likelihood (ML) estimation can be performed. Given the time series model as described by (2.11) multistage ML can be applied for estimating the first level conditional means and variances, the marginal distribution of the

residuals and the copula. If parametrized models for the marginal distributions are used, it is crucial that all models are correctly specified, otherwise the estimated copula parameters will be biased (Oh and Patton 2013, p. 694). Chen and Fan (2006, p. 127) showed that, if the EDF is used for the estimaton of the marginal distribution of the residuals, the first level estimates do not affect the asympotic distribution of the semiparametric ML estimator of the copula parameters.

Since the factor copula implied by (2.8) is not given in closed form approximating the Likelihood function is difficult and computationally demanding (Oh and Patton 2013, p. 694). On the other hand, it is easy and computationally cheap to simulate many values from the factor model. Therefore, an alternative approach is used. It is based on the Generalized Methods of Moments (GMM), for which simulated values are used in the moment conditions. As shown by Chen and Fan (2006) for the semiparametric ML estimator, the SMM estimator is also not affected by the first level dynamic estimates of the time series model in (2.11) (Oh and Patton 2013, p. 692).

2.4.1. General estimation procedure

SMM and GMM estimation is based on the method of moments. For many statistical parametric models, theoretical population moment conditions as functions of the model parameters and the random variables can be derived. For estimation, the expectation in these conditions can then be replaced by the respective sample average (Hall 2005, pp. 6-7).

Formally, for SMM to work the population moment conditions take the form $E(f(\boldsymbol{\eta}_t, \boldsymbol{\theta})) = E(\hat{m}(\boldsymbol{\eta}_t)) - m_0(\boldsymbol{\theta})$ with $E(f(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0)) = 0$. We define $\hat{\boldsymbol{m}}(\boldsymbol{\eta}_t)$ to be a vector of moment functions applied to the observable residual data, $\boldsymbol{\theta}_0$ the vector of true parameters and $m_0(\boldsymbol{\theta})$ the true moments as functions of the model parameters.

To estimate the parameters the expecation is replaced by the sample average. If exactly as many moment conditions as model parameters exist, the system can be solved exactly. If more moment conditions than parameter exist no unique solution exist and one minimizes the weighted distance to 0 and the estimator is given by

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}}(\frac{1}{T}\sum_{t=1}^{T}f(\boldsymbol{\eta}_t,\boldsymbol{\theta}))'\hat{W}(\frac{1}{T}\sum_{t=1}^{T}f(\boldsymbol{\eta}_t,\boldsymbol{\theta})). \tag{2.14}$$

Under suitable assumptions and by applying a law of large numbers and the central limit theorem it can be shown that the estimator convergence in probability to the true parameters and that it is normal distributed. The weighting matrix is a positive semi-definite symmetric matrix which must converge in probability to a positive definite matrix of constants. It should be choosen in such a way, that it minimizes the asymptotic variance of the GMM estimator (Hall 2005, pp. 14, 43).

For the GMM, stating the vector of true moments m_0 is often straightforward. In cases, where there exist no explicit mapping from the model paramers to the true moments but the DGP is known one can use the SMM (Hall 2005, pp. 342-347). For this, m_0 is replaced by applying the moments function to simulated data. The simulated moments which we write as $\tilde{m}(\boldsymbol{\eta}_s;\boldsymbol{\theta})$ are then compared to the observable sample moments. Here, one minimizes the weighted distance between the observed and simulated moments. The SMM estimator is given by:

$$\hat{\boldsymbol{\theta}}_{SMM} = \arg\min_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}) = \hat{g}'\hat{W}\hat{g},$$
$$\text{with } \hat{g} = \frac{1}{T}\sum_{t=1}^{T}(\hat{m}(\boldsymbol{\eta}_t)) - \frac{1}{S}\sum_{s=1}^{S}(\tilde{m}(\boldsymbol{\eta}_s;\boldsymbol{\theta})). \tag{2.15}$$

The SMM estimator can be formulated as a special case of the GMM estimator. Therefore, in most cases the SMMM estimator is also consistent and asymptotically normal distributed. However, since the estimation error from the simulations cannot be neglected, the variance of the SMM is usually larger. The innefficiency vanishes if $\frac{T}{S} \to \infty$ (Hall 2005, pp. 344-345).

2.4.2. SMM for factor copulas

The presented estimation method and its asymptotic results are only valid for well defined moment conditions based on the moments of the variables, e.g. the mean or the variance. For factor copulas, these moments cannot be formulated. The scale-invariant measures of dependency as defined in (2.5) -

2. Theoretical foundation

(2.7) are functions of the copula. Here, they are used to replace the moment conditions of the ordinary GMM approach. Therefore, the assumptions and the asymptotic results of the GMM estimator are not valid and a law of large numbers doesn't apply. However, as Oh and Patton (2013) showed, asymptotic results for the distribution of $\hat{\boldsymbol{\theta}}$ can still be derived.

For factor copulas we adapt the SMM estimator in (2.15) and replace the empirical moment functions by

$$\hat{g} = \hat{m}_T(\boldsymbol{\eta}) - \tilde{m}_S(\boldsymbol{X}(\boldsymbol{\theta})), \tag{2.16}$$

where $\boldsymbol{X}(\boldsymbol{\theta})$ are simulations from the factor model in (2.8). The function $m(\cdot)$ generates a vector of pairwise dependency statistics. For each pair of variables we choose the empirical counterparts of Spearman's Rho as given in 2.6 and lower and upper quantile dependency as given in 2.7 for $q \in (0.05, 0.1, 0.9, 0.95)$. Other invariant measures such as Kendall's Tau could also be included.

Formally, $m(\boldsymbol{Y}) = (\boldsymbol{\delta}_{i,j})'$ with $i = 1, \dots, N-1, j = i+1, \dots, N$ where $\boldsymbol{\delta}_{i,j} = (\hat{\rho}_{i,j}^S, \lambda_{i,j}^L(0.05), \lambda_{i,j}^L(0.1), \lambda_{i,j}^U(0.90), \lambda_{i,j}^U(0.95))'$ is a vector of dependencies of length 5.

For an unrestrictive factor model there exist $5(0.5 \cdot N \cdot (N-1))$ dependency statistics, five for each unique pair of variables. For the simpler equidependence model, which assumes the same factor loading for all variables, one can average over all vectors such that we can write:

$$m(\cdot) = \frac{2}{N * (N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \boldsymbol{\delta}_{i,j}. \tag{2.17}$$

This results in a single vector of dependencies.

For the bloc-equidependence model the final number of dependencies is $5M$ since one can average over all intra- and intergroup dependencies. For each group $s = 1, \dots, M$, we can write

$$\bar{\boldsymbol{\delta}}_s = \frac{1}{M}(\underbrace{\sum_{r=1, r \neq s}^{M} \frac{1}{k_s k_r} \sum_{i=1}^{k_s} \sum_{j=1}^{k_r} \boldsymbol{\delta}_{is,jr}}_{\text{intergroup dependencies}} + \underbrace{\frac{2}{k_s(k_s-1)} \sum_{i=k_s}^{k_s-1} \sum_{j=i+1}^{k_s} \boldsymbol{\delta}_{is,js}}_{\text{intragroup dependencies}}), \tag{2.18}$$

14

where $\boldsymbol{\delta}_{is,jr}$ is the vector of dependencies for the ith variable in group s and the jth variable in group r. Finally for the bloc-equidependence model, the moment condition can be written as: $m(\boldsymbol{Y}) = (\bar{\boldsymbol{\delta}}_1, \dots, \bar{\boldsymbol{\delta}}_S)'$.

Oh and Patton (2013, p. 691ff) showed that under some regulatority assumptions the SMM estimator is weakly consistent and asymptotically normal distributed. The assumptions ensure that for both the iid and the time series case the sample dependency measures converge in probability to their theoretical population values. It is important to note, that the population version of the moment conditions is assumed to be differentiable at the true parameter $\boldsymbol{\theta}_0$ while this is not true for the finite sample version.

The convergence of the estimator can be summarized as:

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta_0}, (\frac{1}{T} + \frac{1}{S})\boldsymbol{\Omega}) \quad \text{for} \ \ T, S \to \infty, \tag{2.19}$$

with covariance matrix $\boldsymbol{\Omega} = (\boldsymbol{G'WG})^{-1}\boldsymbol{G'W'\Sigma WG}(\boldsymbol{G'WG})^{-1}$, where $\boldsymbol{G} = \frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_0}$ is the Jacobian matrix of the first order derivitives of the moments function $\boldsymbol{g}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}$ is the asymptotic variance of the sample moments $\hat{\boldsymbol{m}}$.

The weight matrix $\boldsymbol{W}$ can be chosen to be the identity matrix which results in some efficiency losses. If the efficient weight matrix $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}$ is used the aymptotic variance simplifies to $\boldsymbol{\Omega} = (\boldsymbol{G'WG})^{-1}$.

The efficient weight matrix does not depend on the parameters of the factor copula model and can therefore be estimated prior to estimating $\boldsymbol{\theta}$. The asymptotic variance of the sample moments $\boldsymbol{\Sigma}$ can be estimated with an iid bootstrap procedure:

The derivative $\boldsymbol{G}$ can be estimated using a numerical approximation arround the parameter estimator $\hat{\boldsymbol{\theta}}$. For the kth column we can write:

$$\hat{\boldsymbol{G}}_{T,S,k} = \frac{\boldsymbol{g}_{T,S}(\hat{\boldsymbol{\theta}} + \boldsymbol{e}_k \epsilon_{T,S}) - \boldsymbol{g}_{T,S}(\hat{\boldsymbol{\theta}} - \boldsymbol{e}_k \epsilon_{T,S})}{2\epsilon_{T,S}},$$

where $\boldsymbol{e}$ is the kth unit vector and $\epsilon_{T,S}$ the step size which is usually set to $\epsilon_{T,S} = 0.1$.

Algorithm 1: Bootstrap procedure for the estimation of $\hat{\boldsymbol{\Sigma}}_{T,B}$.

Input: $\{\hat{\boldsymbol{\eta}}_t\}_{t=1}^T$
Output: $\hat{\boldsymbol{\Sigma}}_{T,B}$
$\hat{\boldsymbol{m}}_T \leftarrow$ compute sample moments from residuals $\{\hat{\boldsymbol{\eta}}_t\}_{t=1}^T$;
for $b \leftarrow 1$ **to** B **do**
$\quad$ $\{\hat{\boldsymbol{\eta}}_t^{(b)}\}_{t=1}^T \leftarrow$ sample T values with replacement from $\{\hat{\boldsymbol{\eta}}_t\}_{t=1}^T$;
$\quad$ $\hat{\boldsymbol{m}}_T^{(b)} \leftarrow$ compute sample moments from bootstrap sample $\{\hat{\boldsymbol{\eta}}_t^{(b)}\}_{t=1}^T$;
end

$\hat{\boldsymbol{\Sigma}}_{T,B} \leftarrow \frac{T}{B}\sum_{b=1}^B ((\hat{\boldsymbol{m}}_T^{(b)} - \hat{\boldsymbol{m}}_T)(\hat{\boldsymbol{m}}_T^{(b)} - \hat{\boldsymbol{m}}_T)')$;

return $\hat{\boldsymbol{\Sigma}}_{T,B}$;

2.5. A structural break test for factor copulas

Allthough the factor copula as presented in the last sections is intended to be used in a time series context, the underlying dependence strucuture of the residuals is assumed to be static. All dynamics are captured by the conditional mean and variance of the marginal distributions. But given the findings of time varying correlations in real datasets the assumption of a static copula seems unrealistic (Manner and Reznikova 2012, p. 655).

Different approaches exist to allow for a time varying dependence structure (Manner and Reznikova 2012, pp. 658 - 668): For instance, the parameters of the copula can be modelled as functions of the lagged variables, autoregressive terms or as an independent stochastic process. But one can also test for a strucutral break in the copula parameters at some point in time t. While these approaches assume that the functional form of the copula doesn't change other methods explicetly allow multiple alternating copula functions.

The dynamic model as presented in 2.3 allows for a wide variety of parametrization and copula functions. Here, we focus on the factor copula model and the SMM estimation procedure as presented in the previous sections. In the following section a structural break test based on Manner, Stark, and Wied (2017) for a change of the parameters of the factor copula as given by (2.11) is presented. The underlying factor copula function stays constant, while the copula's parameters are allowed to vary over time.

The test is based on recursive estimations of the copula model at each timepoint $t \in \{\epsilon T, \ldots, T\}$, with $0 < \epsilon < 1$. We write the recursive parameter as $\boldsymbol{\theta}_t$. For

the estimation, all information from the first period up to point t is used. The strictly positive trimming parameter ϵ has to be chosen such that the model parameters don't fluctutate too much due to the larger estimation error of smaller sample sizes.

The null hyopthesis states that the parameters of the factor copula model are stationary while the alternative assumes at least one significant break point in time:

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_T \qquad H_1 : \theta_t \neq \theta_{t+1} \text{ for some } t = \{1, \ldots, T\}. \quad (2.20)$$

The recursive estimator $\hat{\boldsymbol{\theta}}_t$ is compared to the the estimated parameters of the full model $\hat{\boldsymbol{\theta}}_T$. If the recursive estimator is significantly far away from the full model estimates, H_0 is rejected. Formally, the recursive statistics are calculated as the scaled squared distance between the full model and the recursive estimator. The final test statistic is then given as the maximum distance:

$$P = \max_{\epsilon T \leq t \leq T} \frac{t^2}{T} T(\hat{\boldsymbol{\theta}}_t - \hat{\boldsymbol{\theta}}_T)'(\hat{\boldsymbol{\theta}}_t - \hat{\boldsymbol{\theta}}_T). \quad (2.21)$$

Manner, Stark, and Wied (2017, p. 10) also propose a non-parametric alternative test statistic which is solely based on the moment functions as described in section 2.4.2 and does not involve the recursive estimation of a copula model:

$$M = \max_{\epsilon T \leq t \leq T} (\frac{t}{T})^2 T(\hat{\boldsymbol{m}}_t - \hat{\boldsymbol{m}}_T)'(\hat{\boldsymbol{m}}_t - \hat{\boldsymbol{m}}_T), \quad (2.22)$$

where $\hat{\boldsymbol{m}}_t$ is the recursive vector of dependencies calculated with the residuals up to time t.

For the break point detection, the test statistic is compared to a critical value (CV). If it is larger than the CV, the corresponding timepoint of the test statistic is choosen as the breakpoint. The CV is based on the $1 - \alpha$ quantile of the distribution of M or P under the null hypothesis. It can be simulated with a bootsrap procedure as described in algorithm (2).

Manner, Stark, and Wied (2017, p. 13) show, that the test statistics converges in distribution to

$$P \xrightarrow{d} \max_{\epsilon T \leq t \leq T} (\boldsymbol{A}^*(\frac{t}{T}) - \frac{t}{T}\boldsymbol{A}^*(1))'(\boldsymbol{A}^*(\frac{t}{T}) - \frac{t}{T}\boldsymbol{A}^*(1)),$$

$$M \xrightarrow{d} \max_{\epsilon T \leq t \leq T} (\boldsymbol{A}(\frac{t}{T}) - \frac{t}{T}\boldsymbol{A}(1))'(\boldsymbol{A}(\frac{t}{T}) - \frac{t}{T}\boldsymbol{A}(1)),$$

(2.23)

with $\boldsymbol{A}^*(\frac{t}{T}) = (\boldsymbol{G'WG})^{-1}\boldsymbol{G'W}(A(\frac{t}{T}) - \frac{t}{T}\sqrt{(\frac{T}{S})}A(1))$ and $T, S \to \infty$. As for the SMM estimator, $\boldsymbol{G}$ is the matrix of first order derivitives of the population moment function $\boldsymbol{g(\theta)} = \hat{m}_t - \tilde{m_S}; \boldsymbol{\theta}$, W is the weighting matrix of the SMM estimator and $A(\cdot)$ is a gaussian process.[5]

The distribution of the test statistics under the null hypothesis can be estimated using the following bootstrap procedure:

Algorithm 2: Bootstrap procedure for the estimation of the critical value for the structural break test.

Input: $\hat{\boldsymbol{\eta}}_t; \alpha$

Output: K

compute full sample moments $\hat{\boldsymbol{m}}_T$;

if *test based on P* **then**

$\quad | \quad L \leftarrow (\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'$;

else

$\quad | \quad L \leftarrow 1$;

end

for $b \leftarrow 1$ **to** B **do**

$\quad$ generate bootstrap sample $\{\hat{\boldsymbol{\eta}}_t^{(b)}\}_{t=1}^T$;

$\quad$ **for** $t \leftarrow \epsilon T$ **to** T **do**

$\quad\quad$ compute recursive bootstrap sample moments $\hat{\boldsymbol{m}}_t^{(b)}$;

$\quad\quad A_t^{(b)*} \leftarrow L \frac{t}{T}\sqrt{T}(\hat{\boldsymbol{m}}_t^{(b)} - \hat{\boldsymbol{m}}_T)$;

$\quad\quad K_t^{(b)} \leftarrow (A_t^{(b)*} - \frac{t}{T}A_T^{(b)*})'(A_t^{(b)*} - \frac{t}{T}A_T^{(b)*})$;

$\quad$ **end**

$\quad$ calculate the maximum $K^{(b)}$ over all $K_t^{(b)}$;

end

calculate the $1 - \alpha$ sample quantile K from all $K^{(b)}$;

return K;

Given the empirical standardized residuals and some significance level α the procedures mainly calcualates B bootstrap versions of the maximum scaled

---

[5]The nescessary regulatory assumptions to derive this result are similar in spirit to the assumptions used to derive the asymptotic distribtion of the parameters of the factor copula model in section 2.4.2.

quared distance between the recursive moments $\hat{\boldsymbol{m}}_t$ and the full sample estimate $\hat{\boldsymbol{m}}_T$. For the copula based test $(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}$ is estimated only once using the full sample estimates and the estimation procedures as described in section 2.4.2 and in algorithm (1).

3. *factorcopula* - an R package for simulation and estimation of factor copulas

In this chapter we focus on the implementation of the previous methods and procedures. Using the programming language R an open-source package was built, such that the methods can be easily tested, installed and distributed (R Core Team 2017).[6] First, the functions for the configuration and simulation of factor copulas are shortly presented. Second, an overview of the numerical optimization strategy is given in section 3.2). Explained code examples of the main functions can be found in the appendix section A.1. Finally, the validity of the package and the discussed procedures is tested in two simulation studies (section 3.3).

The package consists of a set of high level functions which can be used to construct, simulate and fit various factor copula models. The specification of the factor copula model is handled by the functions `config_factor`, `config_error` and `config_beta`. The two functions `fc_create` and `fc_fit` can be used to either simulate values from a factor copula or to fit a model to a dataset. For conducting the break test as described in section 2.5, the library offers the functions `fc_critval`, `fc_mstat` and `fc_pstat`. The former simulates critical values for either the moments or copula based test. The latter two calculate the recursive test statistics. A matrix of recursive $\hat{\boldsymbol{\theta}}_t$ must be provied for the calculation of the copula based test statistic. It can be obtained by recursively applying `fc_fit` to the data. Obtaining the recursive estimates is computationally costly. To speed up the estimation, the HPC cluster of the Universiy of Cologne was utlizied (Achter, Borowski, Nieroda, et al. 2018).

---

[6]See also the appendix notes on data access in section A.2.

3.1. Copula specification and simulation

With the functions `config_factor`, `config_error` and `config_beta`, the user can define the distribution of the latent variables Z and the error term ϵ as well as the matrix of factor loadings $\boldsymbol{\beta}$. For the specification of the distributions, the function name of any available random number generator can be used.

Additional arguments such as distributional parameters can be declared in a named list. Distributional parameters can either be fixed or passed as non-evaluated expressions. To distinguish free model parameters from fixed distributional parameters, an additional character vector with the name of the model parameters has to be passed to `config_factor` and `config_error`.

The random number generating functions must have an additional argument `n` which defines the number of observations to be simulated. This argument should not be set explicetly since the number of simulations is controlled by the value S in other functions of the package.

For the matrix of factor loadings the user can either manually construct a character matrix of parameters or one can use the function `config_beta`. Given a vector `k` and the number of latent variables K this functions constructs a suitable character matrix of zeros and parameter names. The vector k is of length N and defines the group for each observable variable. Therefore, an equidependence model can be specified with $k_N = (1, 1, \ldots, 1)$, an unrestrictive model with $k_N = (1, \ldots, N)$ and a bloc-equidependence model with $k_N = (1, 1, \ldots, 2, 2, \ldots, M, M, \ldots)$, where M is the number of groups.

The function `fc_create` returns itself a random number generating function. To simulate values from it, the user has to specify a vector `theta` of parameters, the number of simulations S and an optional random seed. Fixing the seed always returns the same values. The vector `theta` must be a *named* vector for which the names correspond to the model parameters specified during the configuration of the model.

The simulation of new values is a time consuming part. To avoid unnescessary calls to the underlying random number generators, the function remembers the state of the latent variables Z and the error term ϵ over repeated function calls and only updates the values if nescessary. Thus if neither the seed nor the distributional parameters in `theta` change, the function uses the same random

values from a previous call. This reduces the computation time at the cost of a higher memory consumption.

3.2. Optimization strategy

While the function `fc_create` can be used to simulate values from a factor copula model given a vector of parameters, the function `fc_fit` estimates $\boldsymbol{\theta}$ via SMM as described in section 2.4. Given a copula specification and the observed residuals, it finds the optimal $\hat{\boldsymbol{\theta}}$ which minimizes the weighted squared distance between the dependency vectors calculated with simulated data from the factor model and with observable data as presented in equation (2.15).

Optionally, the function also estimates standard errors using the bootstrap algorithm given by (1). For simplicity, the weight matrix W is set to the identity matrix. Previous studies showed no significant improvement when the theoretically efficient weight matrix $W = \hat{\boldsymbol{\Sigma}}_{T,B}^{-1}$ is used (Oh and Patton 2013, p. 694).

As backend, the method builts on top the `NLopt` optimization library which implements various algorithms for global, local or derivative-free optimization (Johnson 2018; Ypma 2014). The specific choice of the optimization algorithms and the stopping criteria can be altered by the user via the arguments `control.first.stage` and `control.second.stage`. The authors of the library recommend a two step optimization procedure for global optimization: First, a global optimizer approximates the parameter region in which the global optimum lies. Second, a local derivative free optimizer is then applied using the approximated solution from the first stage.

By default, `fc_fit` uses the *Multi-Level Single-Linkage* algorithm in the first step. The algorithm creates a sequence of optimal distributed starting values which are then passed to a local optimizer (Kucherenko and Sytsko 2005). For the local optimizer and the second stage, the *Subplex* algorithm is used, which is based on the popula Nelder-Mead Simplex procedure (Rowan 1990).

During the optimization process, many values from the factor copula model are simulated. To avoid numerical instablities, the random seed is kept fixed, such that always the same random values are drawn. As described in the previous section, the memory functionability of `fc_create` can save some computational costs, since redraws from the distributions are avoided if the distributional

parameter don't change. This can improve the overall performance of the optimizataion process. This is especially true, if only the factor loadings $\boldsymbol{\beta}$ are optimized while the distributional parameters are kept fixed. In this case, the random number generators are only called once at the beginning of the optimization routine (Gouriéroux and Monfort 1996, p. 29).

3.3. Simulation study

To illustrate the discussed methods and the validity of the package, two simulation studies are performed: First, an equidependence factor copula model with varying dimensions and sample sizes is estimated repeatedly to show the consistency propertiy of the SMM procedure. Second, both the moments and copula based structural break test are applied once to simulated data from a bloc-equidependence model as described in section 2.2.2.

3.3.1. Convergence for the equidependence model

The DGP of the first study is based on a simple time-invariant equidependence model with standard-normal marginal distributions, one skew-t distributed latent variable and a t-distributed error term. The copula model produces strong assymetric tail-dependencies. This is achieved by fixing the degrees of freedom at $df = 4$ and the skeweness at $\lambda = -0.8$. The single factor loading is set to $\beta = 1.5$. The only free parameter to be estimated is the factor loading. Thus we can write $\theta = \beta$. The model equations can be written as:

$$
\begin{aligned}
Y_i &\sim N(0,1) \ \forall \ i \in 1, \cdots, N \\
Z &\sim \text{skew-t}(df = 4, \lambda = -0.8) \\
\epsilon &\sim \text{t}(df = 4) \\
X_i &= \beta Z + \epsilon
\end{aligned}
\tag{3.1}
$$

We repeat the simulation and estimation of β over a grid of different values for the number of variables N and the sample size T. Each simulation consists of $C = 1000$ Monte-Carlo replications. The number of simulations in the SMM is set to $S = 25000$. Finally, we get a vector of estimates $\hat{\beta}_{t,n,c}$ with $t \in \{100, 1000, 10000\}, n \in \{2, 3, 10\}, c \in \{1, \ldots, C\}$.
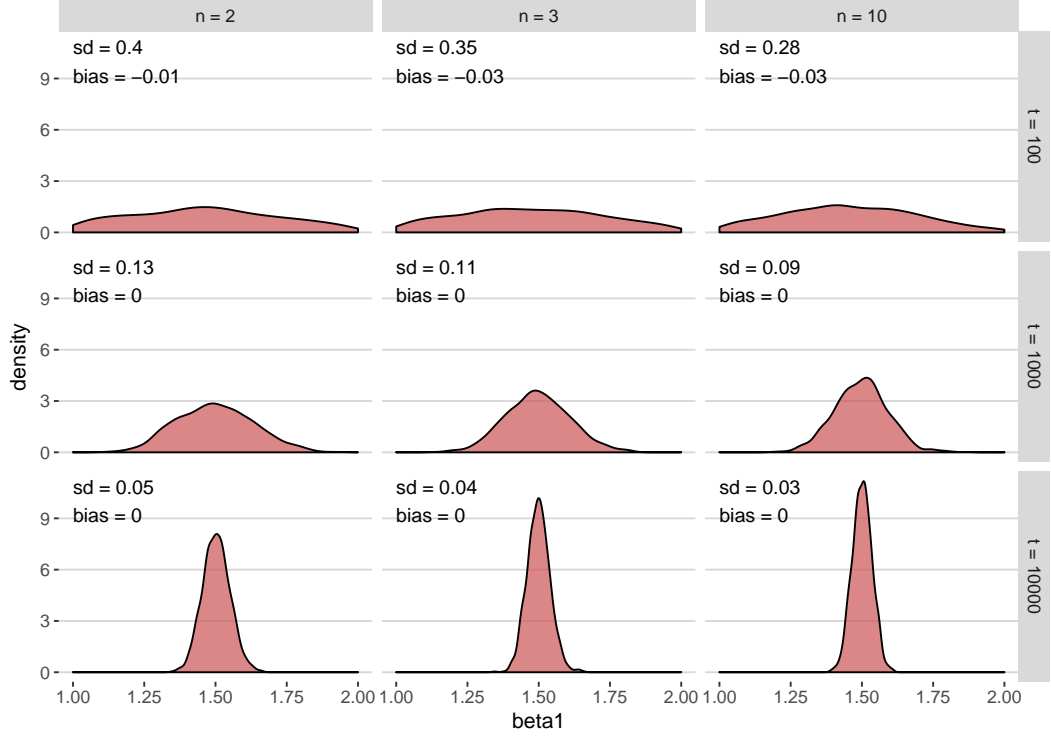
Figure 3.1: Approximated Monte-Carlo density estimators for $\hat{\beta}$ of an equidependence skew t - t factor copula model with $\beta = 1.5$, $S = 25000$, standard normal distributed marginals and different values for N and T. Each simulation is based on 1000 Monte-Carlo replications.

Figure 3.1 shows the results for the first study. For each combination of N and T the kernel density estimator over all Monte-Carlo simulations of is plotted. The bias $\hat{b}_{t,n} = \frac{1}{C} \sum_{c=1}^{C} \hat{\beta}_{t,n,c} - \beta$ and the standard deviation is printed in the top left corner.

The density mass centers arround the true value of $\beta = 1.5$. For all simulations the bias is zero or close to it. As the sample size increases, the deviation gets smaller and the SMM estimator converges to the true parameter. But small sample sizes clearly give unreliable results. Note, that more variables improves the quality of the estimator as more information is available to estimate the single parameter. This is due to the fact that the number of latent variables and parameters is constant and does not increase with the number of dimensions. As stated in (2.17) for the equidependence model, one averages over all pairwise depence vectors. Therefore the estimation error of the mean decreases if N and hence the number of pairwise combinations increases.

3.3.2. Structural break in a bloc-equidependence model

For the second study, a more sophisticated model is presented to illustrate the effectiveness of the approach even for high dimensional problems and complicated dependence structures. Analogous to the empirical examples in Manner, Stark, and Wied (2017) and Oh and Patton (2017) the DGP is based on a *bloc-equidependence* model as described in section 2.2.2.

The model consists of $N = 21$ standard normal distributed variables which are partitioned in $M = 3$ groups of equal size. Each variable is affected by a common and a group specific factor. The common latent factor is skew-t distributed with strong assymetric tail-dependency. The three group specific latent factors and the error term are t-distributed. Due to the bloc-equidependence structure, the number of factor loadings reduces from $0.5 * N * (N - 1) = 210$ to just $2M = 6$. All distributional parameters are kept fixed and only the factor loadings are estimated such that the recrusive estimator can be written as $\boldsymbol{\theta}_t = (\beta_{1,t}, \ldots, \beta_{6,t})'$.

We chose $T = 1500$, $S = 25 \times T$ and a breakpoint in the parameters of the copula at $t = 1000$. Before the break, $\boldsymbol{\theta}_{pre} = (0, 1, 1, 0, 1, 1)'$ and after the break $\boldsymbol{\theta}_{post} = (1.5, 1, 1, 1.5, 1, 1)'$. Thus, only the intra- and interdependence for the first group increases from 0 to 1.5 while the remaining factor loadings stay constant.

The test statistics and critical values are based on three different recursive calculations over a range from $t = 300$ to $T = 1500$: First, all parameters are estimated via recursive SMM. Second, only a subset of $\boldsymbol{\theta}_t$ is estimated recursively while fixing the common and group specific factor for the first group at their full sample estimates. Hence, for this calculation $\boldsymbol{\theta}_t = (\hat{\beta}_{1,T}, \beta_{2,t}, \beta_{3,t}, \hat{\beta}_{4,T}, \beta_{5,t}, \beta_{6,t})$. Finally, the non-parametric test statistics based on the moment functions are calculated.

We expect, that a breakpoint is detected arround $t = 1000$ only for the first and third recursive calculations. For the second calculation no breakpoint should be detected since the factor loadings of the second and third group are not affected by the simulated structural break.

Figure 3.2 shows the test statistics of the three different break tests for a single recursive run of the simulation for $t = 300, \ldots, 1500$. The horizontal solid line indicates the critical value. Each critical value is calculated using $B = 2000$

bootstrap samples as described in algorithm (2). The vertical line indicates the theoretical breakpoint at $t = 1000$.
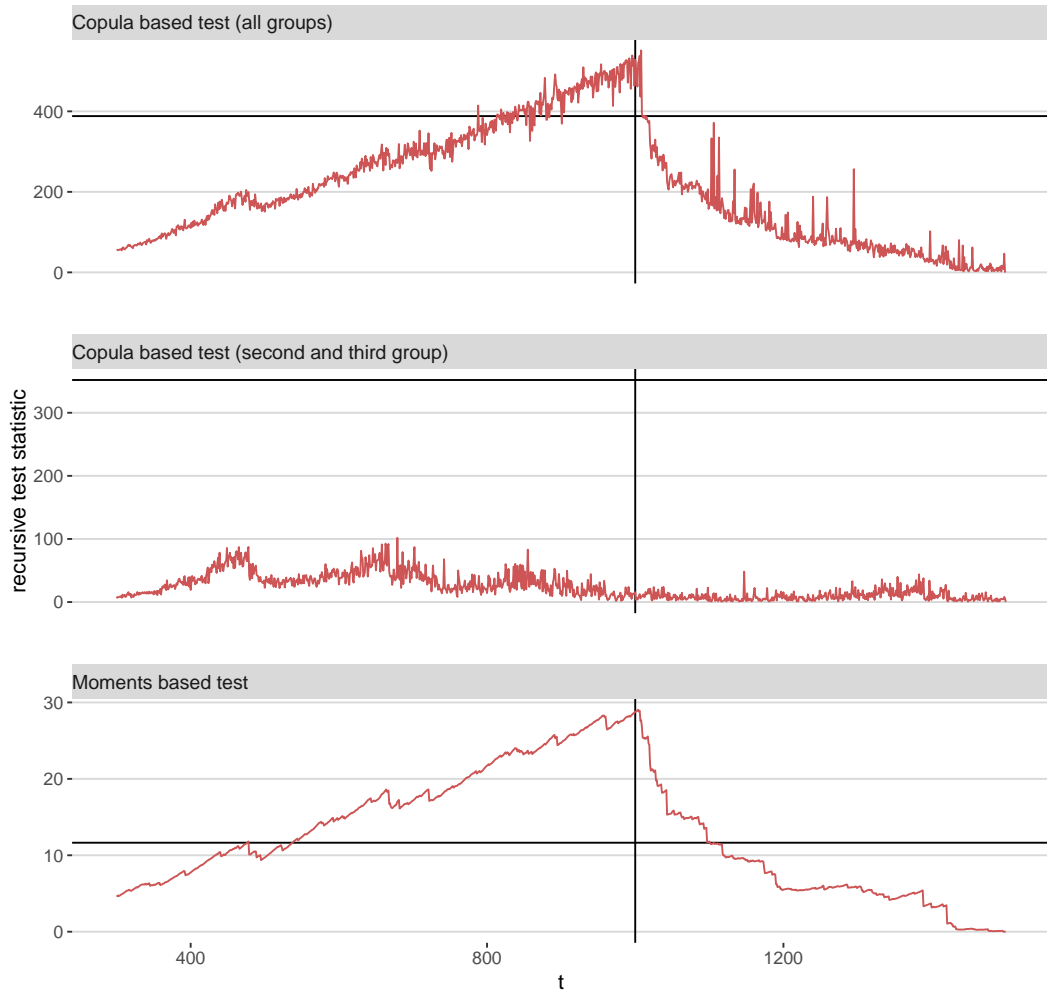

Figure 3.2: Illustration of a structural break test for a bloc-equidependence model with $N = 21$ and 3 groups of equal size. The theoretical breakpoint is at $t = 1000$ and is modelled as a change of the intra- and interdependency of the first group from 0 to 1.5. The first panel shows the copula based test based on all groups. The second panel the copula based test for only the second, and third group which are not affected by the structural break. The lower panel shows the moment based test.

As expected, both the full copula and the moments based test statistics in the top and lower panel fluctuate strongly. For both tests the maximum distance is very close to the true breakpoint and lies above the CV. The CV for the moments based test is relatively far away from the maximum while this is not true for the full copula based test. This could indicate that the moments based

test is less restrive. This observation is also confirmed in the empirical example in the following chapter. The copula based test statistics for the second and third group are clearly below their critical value. Hence, for this groups the null hypothesis of no parameter change cannot be rejected.

Often, numerical problems arise for large parameter vectors or parmater vectors which include distributional parameters, such as the degress of freedom or the skewness. Therefore, the test statistics can be substantially distorted with unusual extreme outliers above the CV. We advise to always perform a manual graphical analysis of the final test statistics. In addition, clear outliers can be detected by performing a smoothing method such as running medians or smoothing splines (see also section 5).

Finally, table 3.1 shows the results for a copula model estimated on the full sample, before and after the theoretical breakpoint. For all models we fixed the distributional parameters at their true values. The standard errors $(\frac{1}{T} + \frac{1}{S})\mathbf{\Omega}$ as given by (2.19) were estimated with $B = 2000$ bootstrap replications and are reported in paranthesis.

coefficient	pre-break ($t \leq 1000$)	post-break ($t > 1000$)	full sample
β_1	0.00 (0.10)	1.96 (0.69)	0.00 (0.13)
β_2	1.17 (0.37)	0.88 (0.29)	1.28 (0.30)
β_3	0.96 (0.27)	0.83 (0.30)	0.99 (0.27)
β_4	0.03 (0.15)	1.55 (0.64)	0.88 (0.15)
β_5	0.92 (0.18)	1.01 (0.32)	0.92 (0.18)
β_6	1.04 (0.20)	1.13 (0.40)	1.09 (0.20)
Q	0.0008	0.0026	0.0025
T	1000	500	1500
S	37500	37500	37500

Table 3.1: Estimation results for the bloc-equidependence factor copula model before and after the breakpoint and for the full sample. Standard errors in paranthesis (estimated with $B = 2000$ bootstrap samples).

Comparing the estimates and the value of the objective function at the optimal value, one can notice that the model after the breakpoint and for the full sample are less precise. For the former case, this can be due to the small sample size. For the latter case, this can be due to the misspecified copula, which assumes

no change in parameters. The estimates for the pre- and post-break sample are close to their true values. The standard errors for the post-break period are larger due to the smaller sample size. However the coefficients are stil significant.

4. Modelling topic dependencies over time with factor copulas

In this chapter, we apply the previously discussed methods to a real aggregated dataset collected from the social media platform Facebook. Typically, copula models are applied to model the dependency structures of financial data. For example, the structural break test was used to detect changing dependencies during the financial crisis (Manner, Stark, and Wied 2017, p. 18). With this analysis, we want to explore other ways of applying the discussed methods in areas outside from applications in finance.

In the following, we use the dynamic model as described by (2.11) together with the factor copula function as given by (2.8). It is the same set up as in the previous chapters. First, we give an overview over the raw dataset. Second, we give a description of the feature generation process, present the first stage models of the conditional mean and variance and give some descriptive overview. Third, the factor copula model is explaind and estimation results for various models applied to the residual data are presented. Finally, we report the results of both the moments and copula based structural break test.

4.1. The *btw17* social media dataset

The raw data consists of social media posts published by public pages on Facebook between January 2014 and December 2017. Using the official list of the candidates for the Bundestag election in 2017 (btw17), the account for each of the politicians was manually researched. Only candidates from the six factions in the *Bundestag* (CDU/CSU, SPD, Die Linke, Bündnis 90/ Die Grünen, AfD, FDP) are part of the study. Arround 84% of all 2516 candidates have an account on Facebook (Stier, Bleier, Bonart, et al. 2018, p. 16).

Due to API (application programming interface) and privacy restrictions, only information from public pages could be accessed such that arround 52% of the social media accounts could be considered for the data collection. In addition to the candiates pages, 113 official pages from the political parties, both on the federal and regional level, were included.

The data collection took place on several days between 2017-11-21 and 2018-02-06. The web-scraping software is built on top of the *restfb* Java client library and makes calls to Facebook's official Graph API (Allen and Bartels 2018). The posts are stored in a document orientated database on cloud-servers located in Germany. Besides the actual content, a post includes a timestamp, the user-id of the author and the number of likes, shares and comments it has received uppon collecting it from the API.

For this analysis, the data is restricted on cleaned textual posts only.[7]. This results in almost 664 thousand posts tagged with the party membership of their authors. The first two panels of figure 4.1 show the monthly number of active accounts and the monthly number of posts for each party. An account was defined active if it published at least one post in a month.

In early 2014, approximately 500 accounts were active. This number increased steadily to roughly 750 accounts in mid 2016. From then until the election in September 2017 the number increased rapidly to almost 1200 active accounts followed by a drop after the election. This trend has two reasons: First, we only collected accounts from politicians which were candidates in the btw17. Therefore, other politicians which were only active before the btw17 are not considered here. Second, we suspect that many politicians opened an account just for the election campaign. After the election many candidates closed their accounts due to a failure in the election or because campaigning time was over.

A similar pattern can also be observed for the monthly number of posts. Until the election year roughly 10000 posts from political parties were published per month. In the month of the election it was arround 4 times more.

The bottom panel of figure 4.1 shows the absolute number of monthly posts related to the topic "refugees". We define a post to be refugee related if it matches the regular expression `flucht|fluecht`. For example, a match occurs

---

[7]A post can also contain a foto, an album or an event with no additional text. The posts's text was cleaned by removing links and stopwords, transforming umlauts and converting it to lowercase letters.

if a post contains the words *flüchtlingskrise* (refugee crisis), *fluchtursachen* (causes of flight), *flüchten* (to flee) or *flüchtlingsheime* (refugee hostels). A sharp rise of refugee related posts can be observed in autumn and winter of 2015 and 2016. During this time many refugees entered Germany, fleeing from the war in Syria. This event is commonly labeled as the "german refugee crisis" in the media and the political discussions.
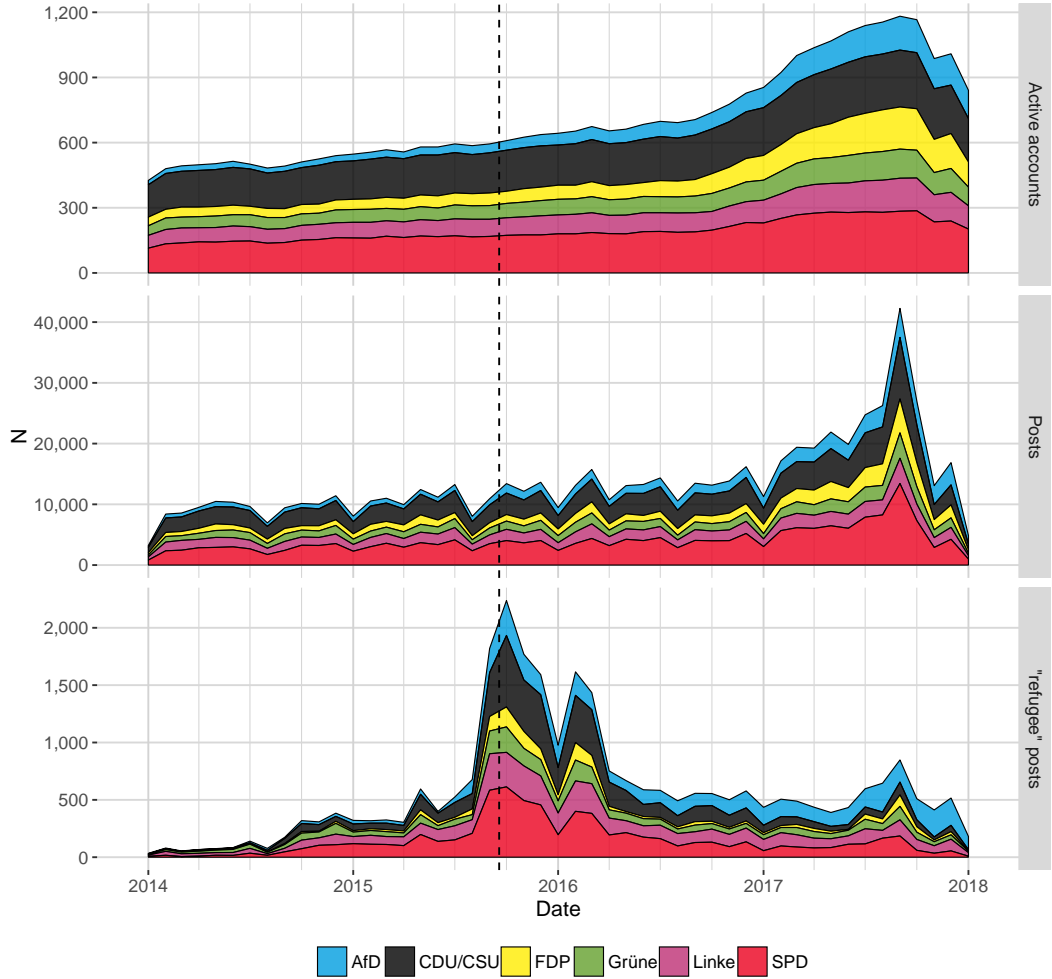

Figure 4.1: Number of active accounts and posts per party and month. The bottom panel shows the number of posts matching the regular expression `flucht|fluecht`. The vertical line indicates the breakpoint detected by the moments based test.

Table 4.1 summarizes some overall aggreagated statistics for each party. It shows the overall number of posts, accounts, summed up ikes and shares for posts and the within party share of refugee related posts from 2014-01-01 to 2017-12-31. Allthough the right and left wing parties *AfD* and *Linke* have a relatively small number of accounts and posts they generate by far the greatest

number of attention in terms of likes and shares. The social democratics party *SPD* has over twice more posts than the *AfD* but roughly generates only half of the likes and only a fifth of the shares. Looking at the share of refugee related posts one can see that both the far left and far right parties talk more about refugee related topics than the average.

party	posts	accounts	likes	shares	share of posts related to
			(in Million)		"refugees" (in %)
AfD	74724	162	18.36	7.59	6.96
CDU/CSU	169115	267	14.72	1.83	3.49
FDP	71083	201	6.32	0.77	2.39
Grüne	67188	139	4.03	1.28	4.62
Linke	84723	158	16.03	4.23	6.23
SPD	196805	290	10.11	1.57	3.66
All parties	663 638	1217	69.57	17.27	4.28

Table 4.1: Overall number of posts, active accounts, likes and shares over the observation period from "2014-01-01" - "2017-12-31". The last two columns show the fraction of posts matching the regular expression `flucht|fluecht`.

4.2. Data processing and descriptive analysis

For each party we count the daily number of refugee related posts and divide it by the overall daily number of posts from this party. Thus, for each day and party, we get a relative within-party frequency of refugee related posts. We use this statistic to approximate the importance of the topic for each party. For example, a value close to 1 indicates that all political discussions are refugee related while a value close to 0 indicates that this topic is of no importance in the political discourse.

Positive dependencies between the parties indicate that refugee related topics are relevant for all parties. This can be due to external events which equally influence the discussions for all parties. Whereas no dependence indicates, that the topic is of no interest in general or that the importance is party-specific and independent from others. Tail dependencies indicate that only in the time of disruptive external events the importance of refugee related topics is equally high for alle parties.

4. Modelling topic dependencies over time with factor copulas

We first estimate the univariate models for each of the time series. The factor copula dependency analysis is then performed on the residual information. The EDF is used to model the marginal distributions of the residuals.

Standard ARIMA-GARCH models are used for modelling the univariate time series (Teräsvirta 2009). For the conditional variance we assume a GARCH(1, 1) process. For the model of the conditional mean we estimate different ARIMA models. The parts of the mean model are determined by running various models over a grid of different parameters. The model candidate with the lowest bayseian information criterion (BIC) is then chosen. Table 4.2 summarizes the final parameter values for the six time series model applied to the relative daily frequencies.

Party	AfD	CDU/CSU	FDP	Grüne	Linke	SPD
AR	4	1	2	3	2	1
I	1	1	1	1	1	1
MA	1	2	1	3	2	1

Table 4.2: ARIMA Model parameters for the ARIMA-GARCH(1,1) model used to estimate the standardized residuals.

Figure 4.2 shows the time series of relative frequencies and the residuals for each of the six parties after applying the ARIMA-GARCH(1, 1) models. Inline with the bottom panel of figure 4.1, refugee related topics start to become important in midth 2015 with peaks of up to 40% importance in late 2015. Espacially for the left and right wing parties, "Linke" and "AfD", the topic stays important even after the crisis. The ARIMA-GARCH model seem to correctly fit the time series but some extreme positive outliers are prevalent.
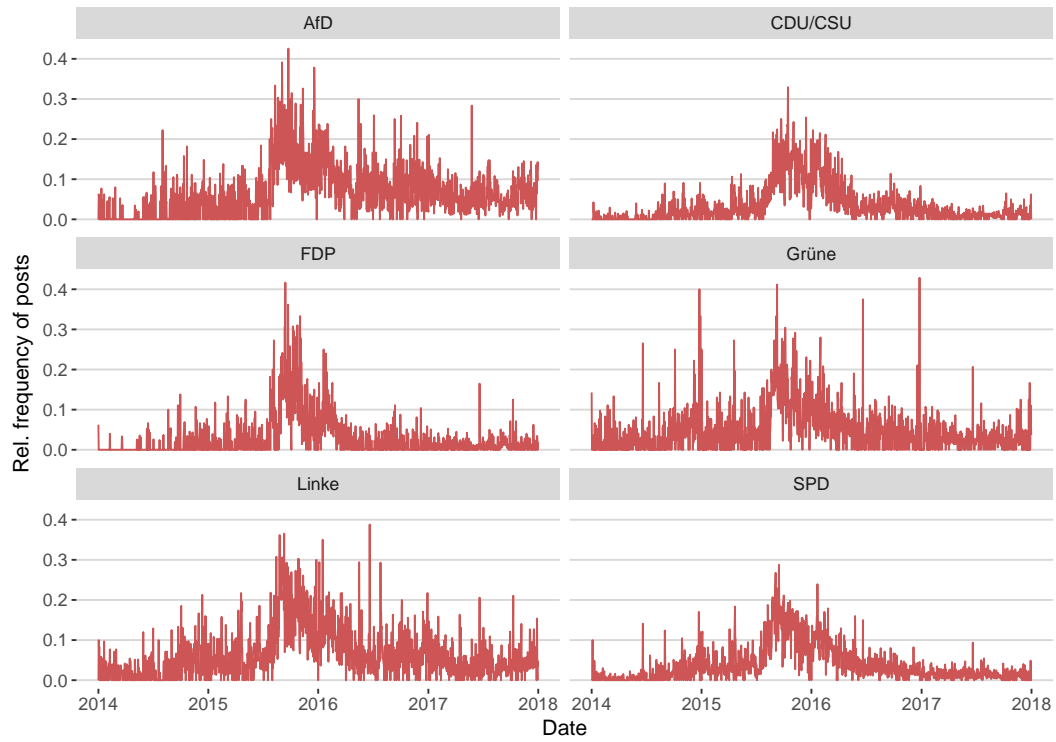
(a) Relative daily within-party frequency of refugee related posts.

(b) Residuals of ARIMA-GARCH models applied to the time series.

Figure 4.2: Univariate time series and residuals of the daily within-party frequency of refugee related topics. The residuals are estimated by applying ARIMA-GARCH models to the time series.

Table 4.3 shows the pairwise sample dependencies for the residuals after removing the estimated conditional mean and variance from the observations. Rank correlation, lower and upper quantile depence as they are also used for the SMM procedure as decribed in section 2.4.2 are presented. Looking at the rank correlation one finds only very weak to weak positive dependencies of up to 0.20 for the two largest parties CDU/CSU and SPD. The quantile dependencies are small but positive and one cannot detect some distinct asymmetry.

Pairs	Rank - correlation	Quantile-dependence			
		0.05	0.1	0.90	0.95
AfD-CDU/CSU	0.07	0.14	0.17	0.13	0.10
AfD-FDP	0.08	0.10	0.14	0.13	0.07
AfD-Grüne	0.03	0.07	0.16	0.09	0.10
AfD-Linke	0.08	0.07	0.15	0.19	0.08
AfD-SPD	0.09	0.11	0.16	0.16	0.08
CDU/CSU-FDP	0.17	0.16	0.25	0.18	0.14
CDU/CSU-Grüne	0.13	0.11	0.20	0.14	0.07
CDU/CSU-Linke	0.10	0.10	0.18	0.16	0.14
CDU/CSU-SPD	0.20	0.22	0.22	0.19	0.14
FDP-Grüne	0.10	0.08	0.19	0.15	0.10
FDP-Linke	0.10	0.10	0.16	0.17	0.10
FDP-SPD	0.12	0.19	0.23	0.18	0.07
Grüne-Linke	0.09	0.07	0.12	0.19	0.08
Grüne-SPD	0.12	0.14	0.21	0.16	0.14
Linke-SPD	0.15	0.16	0.16	0.23	0.21
Average	0.11	0.12	0.18	0.16	0.11

Table 4.3: Pairwise sample dependencies for the six parties applied to the standardized residuals.

4.3. Results

To get a first idea of the characteristics of the data various factor copula models are fitted to the complete sample. To dertermine the number of latent factors K, we apply the recommended approach by Oh and Patton (2017, p. 148) and analyze the ordered eigenvalues of the residual rank-correlation matrix.

One eigenvalue is above the threshold. Therefore, the analysis is restricted to one-factor copula models.

The best model is used to recurseviley estimate the parameters. We fix the distributional parameters at their full sample estimates to avoid numerical instabilities. Finally, we perform the moments and copula based break test and estimate factor copula models for the pre-break and post-break period.

In figure 4.4 we summarize the results of applying various one-factor copula models to the complete sample of residuals. A restrictive equidependence and an unrestrictive model is performed.

	Equidependence			Unrestrictive		
	norm-norm	t-t	skewt-t	norm-norm	t-t	skewt-t
β_1	0.43	0.42	0.42	0.23	0.23	0.25
β_2	-	-	-	0.62	0.52	0.63
β_3	-	-	-	0.42	0.44	0.43
β_4	-	-	-	0.32	0.34	0.33
β_5	-	-	-	0.37	0.41	0.35
β_6	-	-	-	0.63	0.60	0.67
df	-	99	96	-	58	38
λ	-	-	-0.29	-	-	-0.36
Q	0.0036	0.0034	0.0031	0.1081	0.0979	0.0903

Table 4.4: Estimation results for different one-factor copula specifications.

Using the one-factor skew-t equidepndence model as

	Moments based test		Copula based test
	unrestrictive	restrictive	restrictive
test statistics	6.05	116.95	1.51
95%-CV	1.98	61.90	1.67 (p-value: 0.0685)
breakpoint	2015-09-18	2015-09-18	2015-03-13

Table 4.5: Results of the break point detection tests.

coefficient	before $(t \leq 626)$	after $(t > 626)$	full model
β_1	0.39 (0.03)	0.41 (0.03)	0.42 (0.03)
λ	-0.55 (0.75)	-0.33 (0.54)	-0.29 (0.37)
Q	0.0029	0.0023	0.0031
T	626	835	1461
S	36525	36525	36525

Table 4.6: Estimation results for the btw17 dataset. A bloc-equidependence factor copula model is estimated before and after the breakpoint at $t = 626$. Standard errors in paranthesis (estimated with $B = 2000$ bootstrap samples). Degrees of freedom were fixed at the full model estimate of $df = 96$.

5. Discussion

To get better results much care has to be put into the optimiaztion algorithm and the numerical procedures. Frazier and Zhu (2017) proposes ways to use derivative based optimizataion procedures in cases of SMM where the objective is discontinuty.

Embrechts (2009) and Mikosch (2006) critizes the "hype" arround copulas.

A. Appendix

A.1. Code example

In the following example, a factor copula with three latent variables is defined. The first factor is skewed-t distributed, the second and third are standard normal distributed. The error term is t distributed with 4 degrees of freedom. The distributional parameters `df` and `lambda` of the skewed-t distribution are free parameters of the model.

```r
library(factorcopula)
Z <- config_factor(rst = list(nu = df, lambda = lambda),
                    rnorm = list(),
                    rnorm = list(),
                    par = c("df", "lambda"))
eps <- config_error(rt = list(df = 4))
```

Continuing the example, a bloc-equidependence model with $K = 3$, $N = 6$, $k_1 = k_2 = 3$ and $M = 2$ is constructed. Together with the specification of the latent variables this results in a bloc-equidependence model with a skewed-t distribution for the common factor and a standard normal distribution for each of the two group specific factors.

```r
k <- c(1, 1, 1, 2, 2, 2)
beta <- config_beta(k, 3)
```

In the example, $S = 10$ random values from the copula model are simulated. The loadings for the common factor are $\beta_1 = \beta_2 = 1$ and the group specific loadings β_3 and β_4 are set to zero.

```r
copfun <- fc_create(Z, eps, beta)
theta <- c(beta1 = 1, beta2 = 1, beta3 = 0, beta4 = 0,
           df = 10, lambda = -0.8)
U <- copfun(theta, S = 10)
```

A.2. Notes on data and source code access

An online version of this thesis is publicly available as a git-repository under https://github.com/bonartm/factorcopula-thesis. The repository contains

notes on how to install all dependencies. The source code files for the analyses are located online in the `source` folder.

Due to data restrictions by Facebook it is not possible to publish the original dataset of Facebook posts. However, the residuals of the ARIMA-GARCH models which were applied to the aggregated Facebook data are located online at `data/topics_residuals.rds`.

The methods for the simulation and estimation of factor copula models and the break test are not part of this repository. Instead, they are available via the R-package `factorcopula`. The package can be installed from Github. Further notes can be found under: https://github.com/bonartm/factorcopula (See also chapter 3).

For almost all estimation procedures, the HPC cluster of the Universiy of Cologne was utlizied (Achter, Borowski, Nieroda, et al. 2018). To simplify the workflow we wrote the R-package `cheopsr` which allows the execution of job scripts from within the local R environment. The package is available online at https://github.com/bonartm/cheopsr. To run the package, a Unix-like system and accesss rights to the HPC cluster are obligatory.
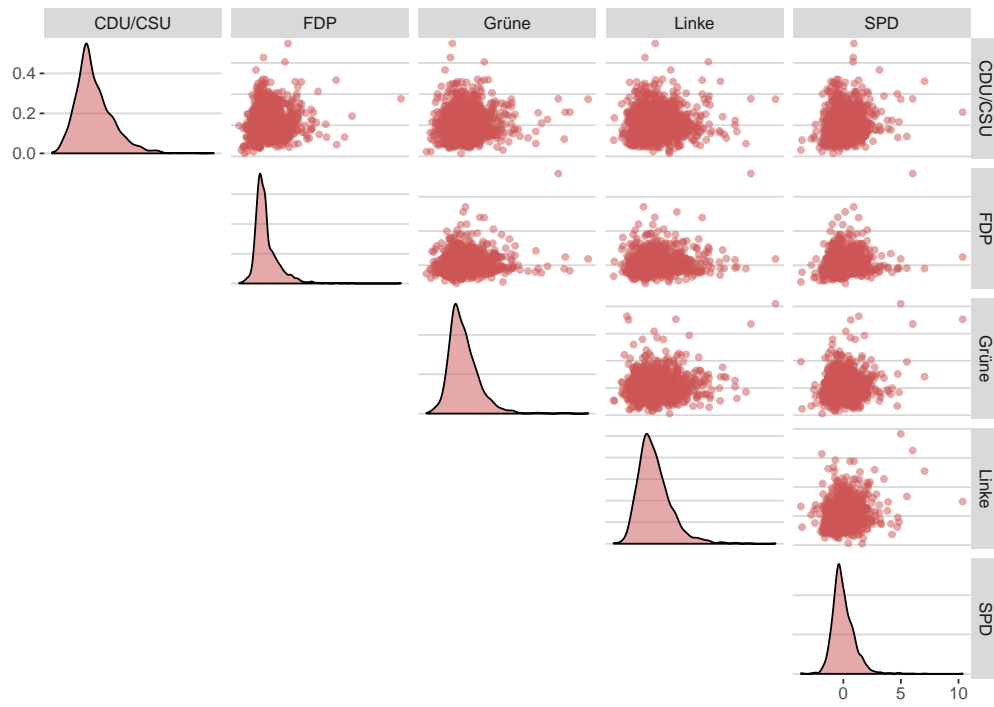
A.3. Additional figures

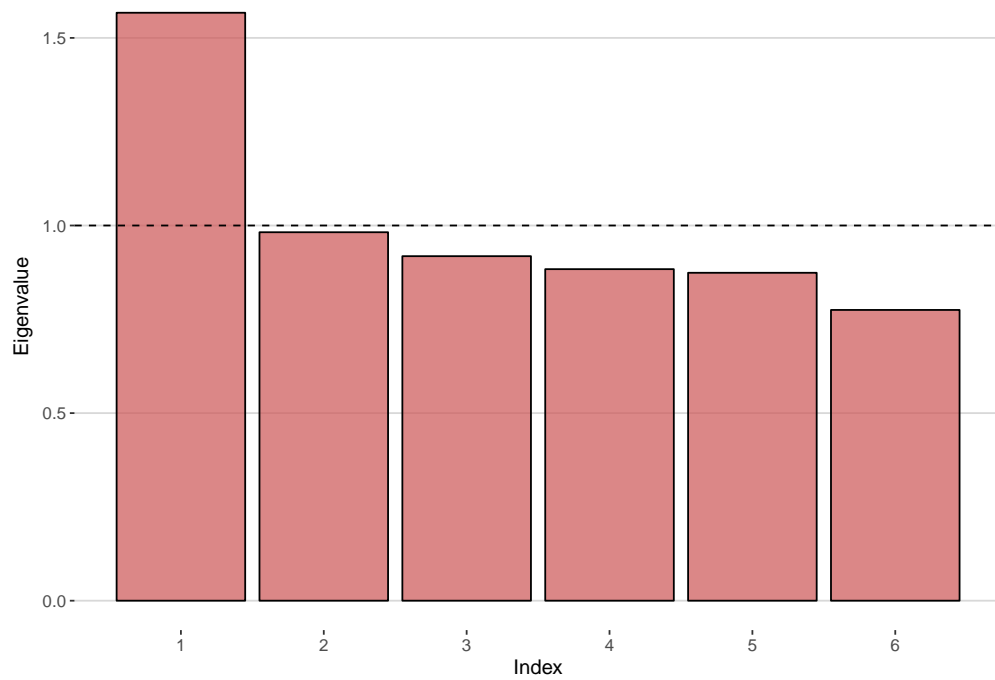Figure A.1: Pairwise scatterplot of the estimated residuals.

Figure A.2: Scree-plot of ranked eigenvalues based on the pairwise rank-correlation matrix.

B. References

Achter, Viktor, Stefan Borowski, Lech Nieroda, et al. (2018). *CHEOPS Cologne High Efficient Operating Platform for Science.* https://rrzk.uni-koeln.de/sites/rrzk/HPC_Projekte/CHEOPS_Brief_Instructions.pdf. [Online techreport; accessed 12-May-2018].

Allen, Mark and Norbert Bartels (2018). *RestFB - a pure Java Facebook Graph API client.* http://restfb.com. [Online documentation; accessed 12-May-2018].

Chen, Xiaohong and Yanqin Fan (2006). "Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification". In: *Journal of Econometrics* 135.1, pp. 125–154.

Embrechts, Paul (2009). "Copulas: A Personal View". In: *Journal of Risk and Insurance* 76.3, pp. 639–650.

Embrechts, Paul, Alexander Mcneil, and Daniel Straumann (2002). "Correlation and dependence in risk management: Properties and pitfalls". In: *RISK Management: Value at Risk and Beyond.* Cambridge University Press, pp. 176–223.

Frazier, David T. and Dan Zhu (2017). *Derivative-Based Optimization with a Non-Smooth Simulated Criterion.* https://arxiv.org/abs/1708.02365. [Online working paper; accessed 12-May-2018].

Gouriéroux, Christian and Alain Monfort (1996). *Simulation-based Econometric Methods.* CORE lectures. Oxford University Press.

Hall, Alastair R. (2005). *Generalized Method of Moments.* Advanced texts in econometrics. Oxford University Press.

Joe, Harry (2015). *Dependence Modeling with Copulas.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability 134. Taylor & Francis.

Johnson, Steven G. (2018). *The NLopt nonlinear-optimization package.* https://nlopt.readthedocs.io/en/latest/. [Online documentation; accessed 12-May-2018].

Kucherenko, Sergei and Yury Sytsko (2005). "Application of Deterministic Low-Discrepancy Sequences in Global Optimization". In: *Computational Optimization and Applications* 30.3, pp. 297–318.

Manner, Hans and Olga Reznikova (2012). "A Survey on Time-Varying Copulas: Specification, Simulations, and Application". In: *Econometric Reviews* 31.6, pp. 654–687.

Manner, Hans, Florian Stark, and Dominik Wied (2017). *Testing for Structural Breaks in Factor Copula Models.* https://www.wiwi.uni-due.de/

B. References

fileadmin/fileupload/VWL-WIPO/WiWi-Kolloquium/Manner_Stark_ Wied___2017_.pdf. [Online working paper; accessed 12-May-2018].

Mikosch, Thomas (2006). "Copulas: Tales and facts". In: *Extremes* 9.1, pp. 3–20.

Nelsen, Roger B. (1999). *An Introduction to Copulas*. Lecture notes in statistics 139. Berlin, Heidelberg: Springer.

Oh, Dong Hwan and Andrew J. Patton (2013). "Simulated Method of Moments Estimation for Copula-Based Multivariate Models". In: *Journal of the American Statistical Association* 108.502, pp. 689–700.

– (2017). "Modeling Dependence in High Dimensions With Factor Copulas". In: *Journal of Business & Economic Statistics* 35.1, pp. 139–154.

Patton, Andrew J. (2009). "Copula–Based Models for Financial Time Series". In: *Handbook of Financial Time Series*. Ed. by Thomas Mikosch, Jens-Peter Kreiß, Richard A. Davis, et al. Berlin, Heidelberg: Springer, pp. 767–785.

– (2012). "A Review of Copula Models for Economic Time Series". In: *J. Multivar. Anal.* 110, pp. 4–18.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/. [Online documentation; accessed 12-May-2018]. Vienna, Austria: R Foundation for Statistical Computing.

Rowan, Thomas Harvey (1990). "Functional Stability Analysis of Numerical Algorithms". PhD thesis. Austin, TX, USA.

Sempi, Carlo (2011). "Copulas". In: *International Encyclopedia of Statistical Science*. Ed. by Miodrag Lovric. Berlin, Heidelberg: Springer, pp. 302–305.

Sklar, Abe (1959). "Fonctions de répartition à n dimensions et leurs marges". In: *Publications de l'Institut de Statistique de L'Université de Paris* 8, pp. 229–231.

Stier, Sebastian, Arnim Bleier, Malte Bonart, et al. (2018). "Systematically Monitoring Social Media: the case of the German federal election 2017". In: *GESIS Papers* 2018/04, p. 25.

Teräsvirta, Timo (2009). "An Introduction to Univariate GARCH Models". In: *Handbook of Financial Time Series*. Ed. by Thomas Mikosch, Jens-Peter Kreiß, Richard A. Davis, et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 17–42.

Ypma, Jelmer (2014). *Introduction to nloptr: an R interface to NLopt*. https://cran.r-project.org/web/packages/nloptr/vignettes/nloptr.pdf. [Online documentation; accessed 12-May-2018].

C. Statutory Declaration

Eidesstattliche Versicherung

Hiermit versichere ich an Eides Statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichen und nicht veröffentlichen Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Köln, den 28. Mai 2018

(Malte Bonart)