DOI 10.1007/s00362-014-0605-7

# Clustering of time series via non-parametric tail dependence estimation

**Fabrizio Durante · Roberta Pappadà · Nicola Torelli**

**Abstract** We present a procedure for clustering time series according to their tail dependence behaviour as measured via a suitable copula-based tail coefficient, estimated in a non-parametric way. Simulation results about the proposed methodology together with an application to financial data are presented showing the usefulness of the proposed approach.

**Keywords** Cluster analysis · Copula · Extreme-value theory · Risk management · Tail dependence

**Mathematics Subject Classification** 62H30 · 62H20 · 62M10

## 1 Introduction

Clustering of time series represents an important tool in finance and economics, since practitioners are often interested in identifying similarities in financial assets for portfolio optimization and/or risk management purposes. As such, several methods have

F. Durante (✉) · R. Pappadà
School of Economics and Management, Free University of Bozen-Bolzano, 39100 Bolzano, Italy
e-mail: fabrizio.durante@unibz.it

R. Pappadà
Department of Statistical Sciences, University of Padua, Padua, Italy
e-mail: pappada@stat.unipd.it; roberta.pappada@unibz.it

N. Torelli
Department of Economics, Business, Mathematics and Statistics "Bruno De Finetti",
University of Trieste, Trieste, Italy
e-mail: nicola.torelli@econ.units.it

been developed in the literature (see, for instance, Piccolo 1990; Pattarin et al. 2004; Otranto 2008; Tola et al. 2008; Brida and Adrián-Risso 2010; Bastos and Caiado 2013) and the references therein.

In general, procedures for clustering financial time series are based on the choice of a relevant dissimilarity measure. Traditionally, clustering techniques have been used to find groups in a portfolio of financial assets from its correlation coefficient matrix: see, for instance, the book by Kaufman and Rousseeuw (1990) and the works by Mantegna (1999) and Bonanno et al. (2004). Such studies try to identify sub-groups of time series by using the linear association (measured by Pearson's correlation) or the comovement degree (measured by Spearman's correlation) of the time series.

Another recent approach consists of finding groups that are similar in the sense that all time series in a group tend to co-move when they are experiencing large losses. Such a different viewpoint was proposed, for instance, by De Luca and Zuccolotto (2011), where a dissimilarity matrix is constructed starting with a parametric estimation of pairwise tail dependence coefficients of the time series. We recall that a tail dependence coefficient (Joe 1997) is a measure (taking values in [0, 1]) that expresses the link between two random variables in the upper (lower) tail of their joint distribution. Similarly, Durante et al. (2013b) have provided a way for grouping time series according to the values of the conditional Spearman's correlation, starting with the seminal ideas by Longin and Solnik (2001) based on exceedance correlations.

It is worth noting that clustering methods that are based on information about the tail dependence are very useful in the analysis of the risk of a portfolio of assets. In fact, classical correlation-based clustering procedures should not be used when there is some contagion effect among the markets under consideration, namely when the positive association among the markets increases in crisis period with respect to tranquil periods. In such a situation, diversification may fail to work exactly when it is needed most (see, for instance, De Luca et al. 2010; Durante and Jaworski 2010; Durante and Foscolo 2013; Durante et al. 2013a).

Here, developing the ideas presented in De Luca and Zuccolotto (2011) we present a method for clustering time series according to tail dependence coefficients. The main novelty of our approach is that it avoids to specify any parametric model assumption on the pairwise dependence structure of the involved time series. In fact, it is only based on the rank statistics derived from the observations. The simulation results show that its overall performance is quite promising even for large portfolios (dimension equal to 128), since it allows to group together different variables according to their tail dependence. In particular, a larger cluster separation in terms of tail dependence implies better performances of the proposed approach. Moreover, while De Luca and Zuccolotto (2011) suggested a multidimensional scaling as a preliminary step to the clustering procedures, our approach may avoid this additional treatment without deteriorating the overall results.

The paper is organized as follows. Section 2 describes the proposed cluster algorithm, whose performance is checked via a simulation study in Sect. 3. An application to the analysis of MSCI Developed Market indices is given in Sect. 4 allowing a direct comparison with the results by De Luca and Zuccolotto (2011). Section 5 concludes.

## 2 The methodology

In this section we present our clustering procedure for grouping time series according to their tail behaviour. Along the manuscript, we consider a matrix of $d$ financial time series $(x_t^i)_{t=1,\dots,T}$ $(i = 1, 2, \dots, d)$ representing the returns of different financial assets collected in a given period of time ($T$ observations). The proposed procedure consists of the following steps, which will be described in detail in the subsequent sections.

1. Choose a suitable copula-based time series model in order to focus the attention to the link between the variables of interest without any influence of the marginal behaviour of each variable.
2. Estimate the pairwise lower tail dependence coefficients $\lambda_{ij}^L$ among the different time series by a non-parametric approach.
3. Define a dissimilarity matrix by using the information contained in the tail dependence coefficients and, hence, apply a clustering algorithm.

### 2.1 Select a suitable copula-based time series model

In order to provide a convenient representation of the dependence among financial time series, a copula-approach can be used (see, e.g., Jaworski et al. 2010, 2013; Patton 2012 and the references therein). Specifically, a suitable stochastic model may be built in two steps: first, the marginals are fitted (means, variances, and distribution of the standardized residuals), then the standardized residuals of the univariate models are coupled via a suitable copula model either parametrically or non-parametrically. This procedure is quite common since it allows to avoid the influence of marginal behaviour on the dependence structure (i.e. the copula) among the variables under consideration. Following these ideas, in order to analyse the time series $(x_t^1, \dots, x_t^d)_{t=1,\dots,T}$ we proceed in the following way:

1. We fit a suitable ARMA–GARCH model to each univariate time series and we check for the adequacy of the fit via standard tests of uncorrelatedness and heteroscedasticity (e.g., Ljung-Box tests, ARCH tests).
2. We rescale the obtained standardized residuals (Patton 2012 [section 2]) from each time series to the interval [0, 1] (by means of the univariate empirical cumulative distribution function) by obtaining the time series $(z_t^1, \dots, z_t^d)_{t=1,\dots,T}$ that take values in $[0, 1]^d$. Actually, these time series represent an empirical version of the link (i.e. the copula) among the time series under consideration and will be used for the further steps.

If the marginal model is correctly specified, the so-called *pseudo–observations* $(z_t^1, \dots, z_t^d)_{t=1,\dots,T}$ (also known as estimated probability integral transform variables) approximately constitute a random sample generated by the copula $C$ linking the variables of interest. Moreover, as stressed by Remillard (2010) (see also Patton 2012, 2013), the estimated parameters from the conditional mean and variance do not affect the asymptotic distribution of estimated dependence measures.

### 2.2 Estimate tail dependence coefficients

Once we have obtained the pseudo–observations $(z_t^1, \ldots, z_t^d)_{t=1,\ldots,T}$ from the original time series, in order to quantify the degree of dependence in the tail of the joint distribution function of a random pair $(X, Y)$, we adopt the concept of tail dependence coefficient (Joe 1997; Durante et al. 2014). We recall that, if $(X, Y)$ is a continuous bivariate random vector with copula $C$, then the lower and upper tail dependence coefficients (shortly, TDCs) only depend on $C$ and are defined, respectively, by

$$\lambda_L(C) = \lim_{t \to 0^+} \frac{C(t,t)}{t} \quad \text{and} \quad \lambda_U(C) = \lim_{t \to 1^-} \frac{1 - 2t + C(t,t)}{1-t}. \tag{2.1}$$

For what follows, it is important to notice that

$$\lambda_L(C) = \lambda_U(\widehat{C}), \tag{2.2}$$

where $\widehat{C}$ is the survival copula associated with $C$ and given by

$$\widehat{C}(u,v) = u + v - 1 + C(1-u, 1-v).$$

Estimators of tail dependence coefficients have been considered several times in the literature (see, e.g., Frahm et al. 2005). In particular, they are popular in the class of extreme value copulas (see, for instance, Beirlant et al. 2004; Gudendorf and Segers 2010; Salvadori et al. 2007).

We recall that a copula $C$ is called an *extreme value copula* (EVC) if $C(u^t, v^t) = C^t(u,v)$ for all $t > 0$, $u, v \in [0,1]$ (see, e.g., Gudendorf and Segers 2010). A result of Pickands (1981) states that $C$ is an EVC if and only if

$$C(u,v) = (uv)^{A\left(\frac{\log v}{\log(uv)}\right)}, \quad (u,v) \in [0,1]^2, \tag{2.3}$$

where $A : [0,1] \to [1/2, 1]$ is continuous, convex and satisfies the constraint $\max\{t, 1-t\} \leq A(t) \leq 1$ for all $t \in [0,1]$. The function $A$ is referred to as the *dependence function* associated with $C$. In particular, if $C$ is an EVC, then

$$\lambda_U(C) = 2 - 2A\left(\frac{1}{2}\right). \tag{2.4}$$

In other words, the estimation of the dependence function $A$ provides an estimation for the upper TDC. Non-parametric estimation procedures of the dependence function $A$ are quite popular and have different variants (see, for instance, Gudendorf and Segers 2010). Among various possible choices, a good choice is given by the estimator $\widehat{A}^{CFG}$ proposed by Capéraà et al. (1997), due to the fact that it seems to outperform other similar estimators (see the discussion by Genest and Segers 2009). The rank-based version of CFG estimator for $A$ is defined in Genest and Segers (2009)[section 2.2], and it is recalled here for sake of completeness.

Consider a random sample $\{(X_i, Y_i)\}_{i=1,\dots,n}$ from a pair $(X, Y)$ of continuous random variables with joint distribution function $H$ and margins $F$ and $G$. For every $i \in \{1, \dots, n\}$, define the pair $(\hat{U}_i, \hat{V}_i)$, whose coordinates are scaled ranks given by

$$\hat{U}_i = \frac{1}{n+1} \sum_{j=1}^{n} \mathbf{1}(X_j \leq X_i), \qquad \hat{V}_i = \frac{1}{n+1} \sum_{j=1}^{n} \mathbf{1}(Y_j \leq Y_i).$$

For every $i \in \{1, \dots, n\}$ and $t \in (0, 1)$ set

$$\hat{S}_i = -\log \hat{U}_i = \hat{\xi}_i(0), \quad \hat{T}_i = -\log \hat{V}_i = \hat{\xi}_i(1), \quad \hat{\xi}_i(t) = \frac{\hat{S}_i}{1-t} \wedge \frac{\hat{T}_i}{t}.$$

The non-parametric rank-based CFG estimator $\widehat{A}^{CFG}$ for $A$ is defined implicitly by

$$\log \widehat{A}^{CFG}(t) = -\gamma - \frac{1}{n} \sum_{i=1}^{n} \log \xi_i(t), \tag{2.5}$$

where $\gamma = -\int_0^\infty \log(x) e^{-x} dx \approx 0.577$ is Euler's constant. In the sequel, we will denote by $\widehat{\lambda}_U^{CFG}$ the estimator of the upper TDC obtained via formula (2.4) in terms of the estimator in (2.5).

Now, in order to use an Extreme Value Theory approach for the estimation of lower TDC of our time series we adopt the procedure described in Frahm et al. (2005)[section 3.5]. Let $C$ be the copula associated with the pseudo–observations of the considered financial returns (as described in Sect. 2.1). Let $\widehat{C}$ be the copula associated with the pseudo–observations of the corresponding losses (i.e. the opposite of the returns) given by $\widetilde{z}_t^i = 1 - z_t^i$ for every $i = 1, \dots, d$ and $t = 1, \dots, T$.

As a matter of fact, $\widehat{C}$ may not be an EVC; however, under suitable conditions, it belongs to the so-called domain of attraction of a suitable EVC $C^*$ (Gudendorf and Segers 2010). Moreover, it can be proved (Abdous et al. 1999[Lemma 1]) that $\widehat{C}$ and $C^*$ have the same upper TDC. Thus, instead of estimating directly the lower TDC from $C$ (or, equivalently, the upper TDC from $\widehat{C}$), we may estimate it by using the estimator $\widehat{\lambda}_U^{CFG}$ applied to the EVC $C^*$. Obviously, $C^*$ is unknown, but its empirical version can be obtained by extracting block maxima from the loss observations, as suggested by Frahm et al. (2005).

Specifically, the pairwise lower TDC of each pair $(z_t^i, z_t^j)_{t=1,\dots,T}, i \neq j$, of pseudo–observations is calculated via the following procedure.

1. First, set $\widetilde{z}_t^i = 1 - z_t^i$ and $\widetilde{z}_t^j = 1 - z_t^j$. Namely, we pass from the copula of the pseudo-observations to the survival copula of the pseudo–observations.
2. For $k = 1, \dots, m$ consider the coordinate-wise block maxima

$$(\widetilde{M}_k^i, \widetilde{M}_k^j) = \left( \max_{t=(k-1)l+1,\dots,kl} \widetilde{z}_t^i, \max_{t=(k-1)l+1,\dots,kl} \widetilde{z}_t^j \right).$$

That is, we extract the maxima of observations over $m$ blocks of $l = T/m$ elements.

3. Estimate the bivariate dependence function $\widehat{A}^{ij}$ from the $m$ block maxima observations $(\widetilde{M}_k^i, \widetilde{M}_k^j)$ via Eq. (2.5).
4. Then, in view of Eq. (2.4), the lower TDC of $(z_t^i, z_t^j)_{t=1,\ldots,T}$ is equal to $2 - 2\widehat{A}^{ij}(0.5)$.

Notice that, as usual in the block maxima approach (see, for instance, Embrechts et al. 1997), a trade-off necessarily takes place in determining the number and size of blocks: a larger size leads to a more accurate determination of the EVC $C^*$ in the domain of attraction; while a large number of blocks gives more data for the estimation of the dependence function $A$ of $C^*$.

### 2.3 Define a dissimilarity measure and apply a cluster algorithm

A fundamental step in cluster analysis is to obtain a suitable measure of dissimilarity (respectively, similarity) between each pair of time series. Here, as done, for instance, in De Luca and Zuccolotto (2011), we have to transform the estimated TDCs through a monotonic function in such a way that the obtained dissimilarity between two time series is small when their tail dependence is high, and monotonically increases when their tail dependence decreases. Thus, for $i, j = 1, \ldots, d$, a matrix $\Delta = \left(\Delta_{ij}\right)$ is defined whose elements are given by

$$\Delta_{ij} = -\log\left(\hat{\lambda}_{ij}^L\right), \tag{2.6}$$

where $\hat{\lambda}_{ij}^L$ is the lower tail dependence coefficient between time series $i$ and $j$ estimated non-parametrically through the procedure described in the previous section. Notice that Eq. (2.6) defines a dissimilarity measure between time series $i$ and $j$ (i.e. it satisfies the properties of non negativity, identity and symmetry) through a log–transformation multiplied with $-1$, which gives values of dissimilarities ranging from 0 (i.e. when the tail dependence is high) to infinity (i.e. when the tail dependence is extremely low). Hence, the dissimilarity measure in (2.6) is coherent with the idea of similarity we adopt, in the sense that $\Delta_{ij}$ decreases in a monotone way as time series $(x_t^i)$ and $(x_t^j)$ are more and more similar in their lower tail behaviour.

The resulting matrix can be directly used in hierarchical clustering algorithms. When, instead, partitioning methods are used, such a matrix should be further treated in order to obtain a corresponding distance matrix. In particular, we perform a cluster analysis of the $d$ time series by following two different approaches:

1. Apply an agglomerative hierarchical algorithm (e.g. single, average, complete linkage) directly to the matrix $\Delta = \left(\Delta_{ij}\right)$.
2. Perform a non–metric Multidimensional Scaling (MDS) in order to assign Euclidean coordinates to the set of $d$ time series such that the inter-point distances closely match the input dissimilarities. Then, the points configuration obtained can be used as an input for classical $k$-means algorithm. More details about this procedure will be given below.

The first clustering process is based on the hierarchical classification of the objects and generates a graphical representation of the results, the so-called dendrogram, that

shows how clusters are formed at each stage of the procedure. Single, average and complete linkage schemes refer to a different computation of the distance between two groups and can be used on data that are not restricted to Euclidean distances. Then, partitions are obtained by cutting off the dendrogram at an arbitrary height. One advantage of hierarchical clustering is that the number of clusters is not required as a parameter.

The second approach consists in applying $K$-means partitioning method by providing in input the representation of normalized residuals as points in $\mathbb{R}^q$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d$, as obtained from a non–metric MDS algorithm. The general MDS problem refers to the task of assigning Euclidean coordinates to a set of objects such that given a set of dissimilarities, similarities, or ordinal relations between the objects, the embedded points have to fit as closely as possible the original relations. In particular, non–metric algorithms find an embedding respecting only the relative ordering of the input dissimilarities. As done by De Luca and Zuccolotto (2011), we consider the Shepard–Kruskal formulation of non–metric MDS first introduced by Shepard (1962a, b) and further refined by Kruskal (1964), which has been applied extensively. Given the dissimilarity matrix $\Delta = \left(\Delta_{ij}\right), i, j = 1, \ldots, d$, the Shepard–Kruskal algorithm is based on minimizing the so-called *stress function*

$$s = \sqrt{\frac{\sum_{ij}(d_{ij} - \theta(\Delta_{ij}))^2}{\sum_{ij} d_{ij}^2}},$$

where $d_{ij} = ||\boldsymbol{x}_i - \boldsymbol{x}_j||$ refers to the Euclidean distance, across all dimensions, between points $i$ and $j$, and $\theta(\cdot)$ is a weakly monotonic transformation of the input data. Starting from an initial configuration of points in 2-dimensional space, the distortion as measured by $s$ can be further minimized by increasing the number of dimensions $q$, until the stress function is lower than a certain value of tolerance. This results in an Euclidean configuration given by $q \times d$ coordinates of $d$ points, which provide an optimal approximation of the original entries. The $q$-dimensional points configuration can be used as an input for $K$-means algorithm, which minimizes the sum of distances from each object to its cluster centroid, over all clusters. Cluster centroids are computed differently for each distance measure, to minimize the sum with respect to the chosen measure. The method is faster than hierarchical clustering, but the number of clusters has to be fixed in advance. The final output is a set of clusters that are as compact and well separated as possible.

In the following, a simulation study and a practical application show how these two different procedures work.

## 3 Simulation study

In this section, we do a simulation study to check the clustering performances of the proposed methodology. We simulate various dependence structures by using copula functions of different dimensions $d$ and allowing different degrees of joint tail dependencies. Specifically, the copula models we are considering are the following:

Model 1 (cluster independence): we suppose that the random variables of interest are divided into groups such that the pairwise tail dependence is non-zero between the elements within a group, while two elements in different groups are independent. In particular, the used copula models are specified as follows:

$$C(\boldsymbol{u}) = \prod_{j=1}^{J} C_j(u_{j1}, \ldots, u_{jk_j}), \tag{3.1}$$

for all $\boldsymbol{u} = (u_{11}, \ldots, u_{1k_1}, u_{21}, \ldots, u_{2k_2}, \ldots, u_{J1}, \ldots, u_{Jk_J})$, with $k_1 + \cdots + k_J = d$, where each $C_j$ comes from a Clayton (respectively, survival Gumbel) family with a fixed lower TDC.

Model 2 (hierarchical dependence): we suppose that the random variables of interest are divided into groups such that the pairwise tail dependence between the elements within a group is greater than that one between elements belonging to different groups. In particular, the used copula models are related to the general class of hierarchical (or nested) Archimedean copulas (Hering et al. 2010; Okhrin et al. 2013), which is specified as follows:

$$C(\boldsymbol{u}) = C_0(C_1(u_{11}, \ldots, u_{1k_1}), \ldots, C_J(u_{J1}, \ldots, u_{Jk_J})), \tag{3.2}$$

for all $\boldsymbol{u} = (u_{11}, \ldots, u_{1k_1}, u_{21}, \ldots, u_{2k_2}, \ldots, u_{J1}, \ldots, u_{Jk_J})$, with $k_1 + \cdots + k_J = d$, where $C_j$ comes from an Archimedean copula. Specifically, we consider two models generated by this framework: the hierarhical Clayton copula (i.e. a copula of type (3.2) where $C_0, C_1, \ldots, C_J$ belongs to the Clayton family) and the survival copula associated with the hierarchical Gumbel copula.

Model 3 (shock-based dependence): we suppose that the random variables of interest are divided into groups such that the pairwise tail dependence between the elements within a group is non-zero and the tail dependence among elements in different groups is driven by a common random factor, that could be interpreted in terms of a shock. This kind of models is quite advantageous since it allows to put a singular probability mass in the tail of the joint distribution: for more details, see Durante et al. (2010). Here, we consider the copula models belonging to this class that may be expressed via the formula

$$C(\boldsymbol{u}) = C_0(u_{11}^{1-\alpha}, \ldots, u_{Jk_J}^{1-\alpha}) \cdot (\min((u_{11}, \ldots, u_{Jk_J}))^{\alpha}, \tag{3.3}$$

where $\alpha \in [0, 1]$ is a parameter that express the between-cluster tail dependence, while $C_0$ is a copula of type (3.1) where each of its components is a Student $t$–copula with a given TDC (Durante et al. 2010[page 685]).

Notice that, recently, other copula models could be used as well to model tail dependence behaviour among groups (Czado 2010; Brechmann 2014). However, the proposed models are quite convenient in this context since they usually give a direct way to compute the degree of tail dependence within and across groups of variables.

In order to see the performances of our methods in a variety of situations, we generate $N$ data from the above models by considering the following different values:

- the sample size $N = 500$;
- the dimension of the model $d = 32, 64, 128$;
- the number of different clusters $J = 4, 8, 16$, with $J < d/2$;
- the pairwise lower TDC varying according to the different models as specified in the sequel.

For each model, we create the dissimilarity matrix according to the method described in Sect. 2 and we apply two clustering procedures (hierarchical and non-hierarchical), namely:

1. Agglomerative hierarchical algorithm with complete linkage grouping criterion on the dissimilarity matrix $\Delta_{ij} = -\log\left(\hat{\lambda}^L_{ij}\right)$;
2. $K$-means partitioning algorithm on the points in the final configuration returned by Shepard–Kruskal non–metric MDS.

As explained in the previous section, the Shepard–Kruskal algorithm requires in input the number of dimensions $q$ of the points in the final configuration. We first set $q = 2$ and fix the threshold $t = 0.025$ for the minimum value of the stress function, over which any configuration cannot be accepted. Hence we choose to limit the amount of stress (or distortion) to tolerate to 2.5%. To this end, we iterate the algorithm by increasing $q$ until the minimum stress of the corresponding optimal configuration satisfies the constraint $\min(s) < t$.

In order to compare clustering results against external criteria, a measure of agreement is needed. To this aim, we consider the Rand Index (RI) (Rand 1971) and the Adjusted Rand Index (ARI) (Hubert and Arabie 1985). Let $P_1$ and $P_2$ be two partitions of the same set of $n$ objects $S = \{o_1, \ldots, o_n\} : P_1 \equiv \{C_{1k}, k = 1, \ldots, r\}$ and $P_2 \equiv \{C_{2k}, k = 1, \ldots, s\}$. Denote with $a$ the number of pairs of objects that are in the same set in $P_1$ and in the same set in $P_2$, and with $b$ the number of pairs of objects that are in different sets in $P_1$ and in different sets in $P_2$, that is

$$a = |S^\star|,\ S^\star = \left\{(o_i, o_j)|o_i, o_j \in C_{1r_1}, o_i, o_j \in C_{2s_1}\right\},$$
$$a = |S^{\star\star}|,\ S^{\star\star} = \left\{(o_i, o_j)|o_i \in C_{1r_2}, o_j \in C_{1r_3}, o_i \in C_{2s_2}, o_j \in C_{2s_3}\right\},$$

for some $1 \le i, j \le n, i \ne j, 1 \le r_1, r_2, r_3 \le r, r_2 \ne r_3, 1 \le s_1, s_2, s_3 \le s, s_2 \ne s_3$. Then the RI is defined by

$$\mathrm{RI} = \frac{a + b}{\binom{n}{2}}.$$

The RI lies between 0 and 1, where the maximum value is taken when two partitions agree perfectly. The ARI is the corrected-for-chance version of the RI, so as to ensure that its maximum value is 1 and its expected value is zero when the partitions are selected at random:

$$\mathrm{ARI} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{s}\binom{n_{ij}}{2} - \sum_{i=1}^{r}\binom{a_i}{2}\sum_{j=1}^{s}\binom{b_j}{2}\Big/\binom{n}{2}}{\frac{1}{2}\left[\sum_{i=1}^{r}\binom{a_i}{2} + \sum_{j=1}^{s}\binom{b_j}{2}\right] - \sum_{i=1}^{r}\binom{a_i}{2}\sum_{j=1}^{s}\binom{b_j}{2}\Big/\binom{n}{2}},$$

where $n_{ij}$ denotes the number of objects belonging to both $C_{1i}$ and $C_{2j}$, for $i = 1, \ldots, r$, $j = 1, \ldots, s$; $a_i$ the number of objects belonging to $C_{1i}$, for $i = 1, \ldots, r$; $b_j$ the number of objects belonging to $C_{2j}$, for $j = 1, \ldots, s$. The ARI can yield a value between $-1$ and 1.

Supposed that the number of clusters to be selected is fixed and equal to $J$ for each simulation, we calculate the RI and ARI between the obtained cluster structure (from the sampled data) and the expected cluster structure (as derived from the chosen model). The calculations are repeated 250 times, and the average index is considered.

Firstly, we consider the simulation results related to Model 1, as reported in Tables 1, 2 and 3. Here the number $k$ of clusters is not estimated, but supposed to be equal to the true value $J$. The following considerations could be drawn:

- As the dimension $d$ increases, the performance seems to decrease, even if the changes are not so evident.
- For an increasing number of clusters the ARI seems to decrease (notice that it make no sense to use in this case the RI since it is highly dependent upon the number of clusters).
- The different dependence structure (in terms of TDC) matters; in fact, a stronger cluster separation (as obtained by a larger TDC) increases the performances. Moreover, different copula families used for getting the dependence within each cluster also seems to influence the results. This is due to a different tail behaviour that cannot only be captured by the TDC (for more considerations about the tail of a copula see Jaworski 2010).
- Complete linkage clustering procedure outperforms $K$-means clustering procedure, which requires an additional step (multidimensional scaling) to convert the dissimilarity matrix into a distance matrix.

**Table 1** Simulation results from Model 1 for $N = 500$, $d = 32$

| $J$ | Copula family | $\lambda$ | Complete linkage | | MDS + $K$-means | | Stress |
|---|---|---|---|---|---|---|---|
| | | | RI | ARI | RI | ARI | |
| 4 | Clayton | 0.25 | 0.8072 | 0.4845 | 0.6422 | 0.1566 | $\leq$0.024979 |
| | | 0.50 | 0.9733 | 0.9264 | 0.6819 | 0.2636 | $\leq$0.024991 |
| | | 0.75 | 1.0000 | 1.0000 | 0.7201 | 0.3725 | $\leq$0.025000 |
| | Survival Gumbel | 0.25 | 0.7815 | 0.4216 | 0.6398 | 0.1409 | $\leq$0.024974 |
| | | 0.50 | 0.9729 | 0.9253 | 0.6844 | 0.2695 | $\leq$0.024946 |
| | | 0.75 | 0.9998 | 0.9993 | 0.7101 | 0.3665 | $\leq$0.024966 |
| 8 | Clayton | 0.25 | 0.8792 | 0.3596 | 0.8019 | 0.1115 | $\leq$0.024965 |
| | | 0.50 | 0.9690 | 0.8335 | 0.8268 | 0.2346 | $\leq$0.024967 |
| | | 0.75 | 0.9997 | 0.9983 | 0.8536 | 0.3831 | $\leq$0.024999 |
| | Survival Gumbel | 0.25 | 0.8710 | 0.3210 | 0.8040 | 0.1016 | $\leq$0.024903 |
| | | 0.50 | 0.9714 | 0.8463 | 0.8255 | 0.2291 | $\leq$0.024995 |
| | | 0.75 | 0.9999 | 0.9997 | 0.8529 | 0.3876 | $\leq$0.024987 |

**Table 2** Simulation results from Model 1 for $N = 500$, $d = 64$

| $J$ | Copula family | $\lambda$ | Complete linkage | | MDS + $K$-means | | Stress |
|---|---|---|---|---|---|---|---|
| | | | RI | ARI | RI | ARI | |
| 4 | Clayton | 0.25 | 0.7576 | 0.4263 | 0.6636 | 0.2030 | ≤0.024994 |
| | | 0.50 | 0.9710 | 0.9236 | 0.7000 | 0.3066 | ≤0.024991 |
| | | 0.75 | 0.9997 | 0.9992 | 0.7129 | 0.3804 | ≤0.024924 |
| | Survival Gumbel | 0.25 | 0.7293 | 0.3752 | 0.6605 | 0.1849 | ≤0.024993 |
| | | 0.50 | 0.9750 | 0.9348 | 0.6955 | 0.2936 | ≤0.024995 |
| | | 0.75 | 1.0000 | 1.0000 | 0.7090 | 0.3808 | ≤0.024977 |
| 8 | Clayton | 0.25 | 0.8660 | 0.3662 | 0.7978 | 0.1403 | ≤0.024997 |
| | | 0.50 | 0.9712 | 0.8617 | 0.8283 | 0.2957 | ≤0.024988 |
| | | 0.75 | 0.9993 | 0.9966 | 0.8660 | 0.4722 | ≤0.024979 |
| | Survival Gumbel | 0.25 | 0.8580 | 0.3313 | 0.7959 | 0.1294 | ≤0.024996 |
| | | 0.50 | 0.9721 | 0.8669 | 0.8307 | 0.2989 | ≤0.024997 |
| | | 0.75 | 0.9997 | 0.9984 | 0.8514 | 0.4344 | ≤0.024962 |
| 16 | Clayton | 0.25 | 0.9278 | 0.2653 | 0.9039 | 0.1054 | ≤0.024981 |
| | | 0.50 | 0.9777 | 0.7695 | 0.9203 | 0.2751 | ≤0.024989 |
| | | 0.75 | 0.9996 | 0.9963 | 0.9484 | 0.5274 | ≤0.024998 |
| | Survival Gumbel | 0.25 | 0.9236 | 0.2229 | 0.9019 | 0.0865 | ≤0.024995 |
| | | 0.50 | 0.9761 | 0.7544 | 0.9193 | 0.2653 | ≤0.024992 |
| | | 0.75 | 0.9994 | 0.9940 | 0.9466 | 0.5194 | ≤0.024980 |

Notice that we also repeat the same simulation scheme for a sample size $N = 1000$ (results are not reported here). Obviously, the results improve, but the general considerations do not change.

Moreover, for completeness, we consider the case when the number of cluster is not fixed. Specifically, for each sample, we can determine the optimal number of clusters $g$ by the silhouette index (Kaufman and Rousseeuw 1990), which reflects the within-cluster compactness and between-cluster separation of a clustering. In detail, for $g = 1, 2, \ldots$, the number of clusters is chosen such that the average silhouette width is maximized over all $g$. It follows that, if the correct number of clusters is identified, then it coincides with $J$. Otherwise, the procedure has misspecified the cluster structure, a fact that will decrease the performance of our methodology. These results are reported in Table 4 for Model 1 with dimension $d = 32$ (the other dimensions give similar results). As can be seen, setting $k$ unknown does not deteriorate the overall results.

Now, let us repeat the simulation study in the more general case when the within-cluster dependence is not set equal to zero, as provided by Model 2. The results are reported in Tables 5, 6, 7 and 8. The following considerations could be made:

- As the dimension $d$ increases, the performance seems to decrease slightly.
- For an increasing number of clusters the ARI seems to decrease.
- The different dependence structure (in terms of TDC) matters; in fact, a stronger cluster separation (as obtained by larger difference of the TDC's of the copulas $C_0$

**Table 3** Simulation results from Model 1 for $N = 500$, $d = 128$

| $J$ | Copula family | $\lambda$ | Complete linkage | | MDS + $K$-means | | Stress |
|---|---|---|---|---|---|---|---|
| | | | RI | ARI | RI | ARI | |
| 4 | Clayton | 0.25 | 0.6472 | 0.2760 | 0.6855 | 0.2482 | $\leq$0.024986 |
| | | 0.50 | 0.9694 | 0.9219 | 0.7293 | 0.3667 | $\leq$0.024995 |
| | | 0.75 | 1.0000 | 1.0000 | 0.7424 | 0.4321 | $\leq$0.024996 |
| | Survival Gumbel | 0.25 | 0.5927 | 0.2095 | 0.6927 | 0.2527 | $\leq$0.024999 |
| | | 0.50 | 0.9691 | 0.9233 | 0.7235 | 0.3563 | $\leq$0.024994 |
| | | 0.75 | 1.0000 | 0.9999 | 0.7131 | 0.3852 | $\leq$0.024984 |
| 8 | Clayton | 0.25 | 0.8026 | 0.2880 | 0.8090 | 0.2027 | $\leq$0.024997 |
| | | 0.50 | 0.9688 | 0.8583 | 0.8450 | 0.3702 | $\leq$0.024998 |
| | | 0.75 | 0.9994 | 0.9971 | 0.8621 | 0.4804 | $\leq$0.024996 |
| | Survival Gumbel | 0.25 | 0.8343 | 0.2807 | 0.7650 | 0.0971 | $\leq$0.024990 |
| | | 0.50 | 0.9721 | 0.8731 | 0.8419 | 0.3629 | $\leq$0.024996 |
| | | 0.75 | 1.0000 | 0.9999 | 0.8639 | 0.4907 | $\leq$0.024992 |
| 16 | Clayton | 0.25 | 0.9183 | 0.2669 | 0.8991 | 0.1426 | $\leq$0.024998 |
| | | 0.50 | 0.9754 | 0.7770 | 0.9236 | 0.3691 | $\leq$0.024990 |
| | | 0.75 | 0.9993 | 0.9938 | 0.9491 | 0.5930 | $\leq$0.024998 |
| | Survival Gumbel | 0.25 | 0.8930 | 0.0835 | 0.8702 | 0.0246 | $\leq$0.024875 |
| | | 0.50 | 0.9753 | 0.7780 | 0.9249 | 0.3754 | $\leq$0.024996 |
| | | 0.75 | 0.9994 | 0.9941 | 0.9505 | 0.5996 | $\leq$0.024989 |

**Table 4** Simulation results from Model 1 for $N = 500$, $d = 32$

| $J$ | Copula family | $\lambda$ | Complete linkage | |
|---|---|---|---|---|
| | | | RI | ARI |
| 4 | Clayton | 0.25 | 0.7648 | 0.5637 |
| | | 0.50 | 0.8699 | 0.8439 |
| | | 0.75 | 0.9406 | 0.9715 |
| | Survival Gumbel | 0.25 | 0.7484 | 0.5341 |
| | | 0.50 | 0.8757 | 0.8521 |
| | | 0.75 | 0.9272 | 0.9512 |
| 8 | Clayton | 0.25 | 0.7728 | 0.4541 |
| | | 0.50 | 0.8959 | 0.7908 |
| | | 0.75 | 0.9831 | 0.9814 |
| | Survival Gumbel | 0.25 | 0.7582 | 0.4187 |
| | | 0.50 | 0.8942 | 0.7898 |
| | | 0.75 | 0.9648 | 0.9696 |

The number of clusters $k$ is not fixed, but selected by the silhouette index

and $C_i$) increases the performances. Moreover, different copula families used for getting the dependence within each cluster also seem to influence the results.

- Complete linkage clustering procedure outperforms $K$-means clustering procedure, which requires an additional step.

**Table 5** Simulation results from Model 2 for $N = 500$, $d = 32$

| J | $C_0$ | $\lambda_{C_0}$ | $\lambda_{C_i}$ | Complete linkage | | MDS + $K$-means | | Stress |
|---|---|---|---|---|---|---|---|---|
| | | | | RI | ARI | RI | ARI | |
| 4 | Hierarchical Clayton | 0.2 | 0.4 | 0.7395 | 0.3278 | 0.5233 | 0.0933 | ≤0.024854 |
| | | 0.2 | 0.6 | 0.9571 | 0.8855 | 0.6763 | 0.3573 | ≤0.024919 |
| | | 0.4 | 0.6 | 0.8284 | 0.5499 | 0.7771 | 0.4652 | ≤0.024952 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.7813 | 0.4243 | 0.5739 | 0.1436 | ≤0.024919 |
| | | 0.2 | 0.6 | 0.9682 | 0.9148 | 0.7140 | 0.4123 | ≤0.024944 |
| | | 0.4 | 0.6 | 0.8623 | 0.6355 | 0.7979 | 0.5256 | ≤0.024992 |
| 8 | Hierarchical Clayton | 0.2 | 0.4 | 0.8447 | 0.2334 | 0.7219 | 0.0571 | ≤0.024907 |
| | | 0.2 | 0.6 | 0.9529 | 0.7548 | 0.7841 | 0.2104 | ≤0.024906 |
| | | 0.4 | 0.6 | 0.8813 | 0.4184 | 0.8508 | 0.3046 | ≤0.024965 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.8668 | 0.3202 | 0.7339 | 0.0907 | ≤0.024883 |
| | | 0.2 | 0.6 | 0.9701 | 0.8419 | 0.7922 | 0.2507 | ≤0.024708 |
| | | 0.4 | 0.6 | 0.8948 | 0.4699 | 0.8511 | 0.3139 | ≤0.024868 |

**Table 6** Simulation results from Model 2 for $N = 500$, $d = 64$

| J | $C_0$ | $\lambda_{C_0}$ | $\lambda_{C_i}$ | Complete linkage | | MDS + $K$-means | | Stress |
|---|---|---|---|---|---|---|---|---|
| | | | | RI | ARI | RI | ARI | |
| 4 | Hierarchical Clayton | 0.2 | 0.4 | 0.7254 | 0.3161 | 0.4609 | 0.0654 | ≤0.024992 |
| | | 0.2 | 0.6 | 0.9550 | 0.8850 | 0.5770 | 0.2410 | ≤0.024956 |
| | | 0.4 | 0.6 | 0.8267 | 0.5644 | 0.7430 | 0.4457 | ≤0.024937 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.7726 | 0.4201 | 0.5432 | 0.1417 | ≤0.024941 |
| | | 0.2 | 0.6 | 0.9734 | 0.9306 | 0.6197 | 0.2973 | ≤0.024988 |
| | | 0.4 | 0.6 | 0.8564 | 0.6363 | 0.7676 | 0.5039 | ≤0.024859 |
| 8 | Hierarchical Clayton | 0.2 | 0.4 | 0.8244 | 0.2247 | 0.5890 | 0.0388 | ≤0.024953 |
| | | 0.2 | 0.6 | 0.9564 | 0.7963 | 0.6687 | 0.1441 | ≤0.024938 |
| | | 0.4 | 0.6 | 0.8674 | 0.4170 | 0.7952 | 0.2594 | ≤0.024996 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.8530 | 0.3253 | 0.6598 | 0.0739 | ≤0.024997 |
| | | 0.2 | 0.6 | 0.9718 | 0.8666 | 0.6947 | 0.2006 | ≤0.024994 |
| | | 0.4 | 0.6 | 0.8893 | 0.4998 | 0.8167 | 0.3173 | ≤0.024984 |
| 16 | Hierarchical Clayton | 0.2 | 0.4 | 0.9123 | 0.1685 | 0.8114 | 0.0375 | ≤0.024982 |
| | | 0.2 | 0.6 | 0.9669 | 0.6731 | 0.8413 | 0.1285 | ≤0.024994 |
| | | 0.4 | 0.6 | 0.9266 | 0.3167 | 0.9056 | 0.1666 | ≤0.024956 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.9215 | 0.2273 | 0.8384 | 0.0507 | ≤0.024961 |
| | | 0.2 | 0.6 | 0.9773 | 0.7671 | 0.8523 | 0.1669 | ≤0.024994 |
| | | 0.4 | 0.6 | 0.9354 | 0.3794 | 0.9085 | 0.2162 | ≤0.024983 |

**Table 7** Simulation results from Model 2 for $N = 500$, $d = 128$

| J | $C_0$ | $\lambda_{C_0}$ | $\lambda_{C_i}$ | Complete linkage | | MDS + $K$-means | | Stress |
|---|---|---|---|---|---|---|---|---|
| | | | | RI | ARI | RI | ARI | |
| 4 | Hierarchical Clayton | 0.2 | 0.4 | 0.6998 | 0.2910 | 0.4305 | 0.0412 | ≤0.024961 |
| | | 0.2 | 0.6 | 0.9556 | 0.8879 | 0.5206 | 0.1704 | ≤0.024992 |
| | | 0.4 | 0.6 | 0.8461 | 0.6122 | 0.6697 | 0.3705 | ≤0.024977 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.7535 | 0.4068 | 0.4755 | 0.0798 | ≤0.024991 |
| | | 0.2 | 0.6 | 0.9743 | 0.9346 | 0.5910 | 0.2604 | ≤0.024971 |
| | | 0.4 | 0.6 | 0.8739 | 0.6785 | 0.7545 | 0.5137 | ≤0.024985 |
| 8 | Hierarchical Clayton | 0.2 | 0.4 | 0.8105 | 0.2127 | 0.5358 | 0.0293 | ≤0.024961 |
| | | 0.2 | 0.6 | 0.9563 | 0.8065 | 0.5696 | 0.1029 | ≤0.024990 |
| | | 0.4 | 0.6 | 0.8624 | 0.4294 | 0.7414 | 0.2265 | ≤0.024979 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.8409 | 0.3147 | 0.6031 | 0.0696 | ≤0.024971 |
| | | 0.2 | 0.6 | 0.9709 | 0.8692 | 0.6544 | 0.1808 | ≤0.024965 |
| | | 0.4 | 0.6 | 0.8869 | 0.5106 | 0.7820 | 0.3253 | ≤0.024915 |
| 16 | Hierarchical Clayton | 0.2 | 0.4 | 0.8981 | 0.1591 | 0.6964 | 0.0285 | ≤0.024997 |
| | | 0.2 | 0.6 | 0.9635 | 0.6838 | 0.7476 | 0.1029 | ≤0.024994 |
| | | 0.4 | 0.6 | 0.8592 | 0.2048 | 0.8111 | 0.1023 | ≤0.024994 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.9094 | 0.2209 | 0.7775 | 0.0458 | ≤0.024999 |
| | | 0.2 | 0.6 | 0.9752 | 0.7784 | 0.8040 | 0.1605 | ≤0.024986 |
| | | 0.4 | 0.6 | 0.9255 | 0.3744 | 0.8725 | 0.2084 | ≤0.024985 |

**Table 8** Simulation results from Model 2 for $N = 500$, $d = 32$

| J | $C_0$ | $\lambda_{C_0}$ | $\lambda_{C_i}$ | Complete linkage | |
|---|---|---|---|---|---|
| | | | | RI | ARI |
| 4 | Hierarchical Clayton | 0.2 | 0.4 | 0.5929 | 0.5603 |
| | | 0.2 | 0.6 | 0.8274 | 0.8608 |
| | | 0.4 | 0.6 | 0.7362 | 0.6844 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.6675 | 0.6026 |
| | | 0.2 | 0.6 | 0.8881 | 0.8929 |
| | | 0.4 | 0.6 | 0.7400 | 0.7201 |
| 8 | Hierarchical Clayton | 0.2 | 0.4 | 0.5579 | 0.5128 |
| | | 0.2 | 0.6 | 0.7966 | 0.7752 |
| | | 0.4 | 0.6 | 0.6811 | 0.5685 |
| | Survival hierarchical Gumbel | 0.2 | 0.4 | 0.6707 | 0.4832 |
| | | 0.2 | 0.6 | 0.8870 | 0.8262 |
| | | 0.4 | 0.6 | 0.7421 | 0.5799 |

The number of clusters $k$ is not fixed, but selected by the silhouette index

- When the number of cluster $k$ is not specified a priori, the results maintain a reasonable good performance (Table 8).

Finally, even when using Model 3, the general considerations remain the same. The results for Model 3, when $d = 32$, are reported in Table 9.

**Table 9** Simulation results from Model 3 for $N = 500$, $d = 32$

| J | $\alpha$ | $\lambda_{C_i}$ | Complete linkage | | MDS + $K$-means | | Stress |
|---|---|---|---|---|---|---|---|
| | | | RI | ARI | RI | ARI | |
| 4 | | 0.2 | 0.7177 | 0.2651 | 0.6310 | 0.1091 | ≤0.024963 |
| | 0 | 0.4 | 0.9111 | 0.7592 | 0.6673 | 0.2145 | ≤0.024993 |
| | | 0.6 | 0.9956 | 0.9876 | 0.6831 | 0.2855 | ≤0.024981 |
| | | 0.2 | 0.7061 | 0.2326 | 0.5978 | 0.0717 | ≤0.024925 |
| | 0.2 | 0.4 | 0.9004 | 0.7293 | 0.6246 | 0.1585 | ≤0.024988 |
| | | 0.6 | 0.9916 | 0.9773 | 0.6544 | 0.2488 | ≤0.024958 |
| | | 0.2 | 0.6866 | 0.1712 | 0.5477 | 0.0447 | ≤0.025000 |
| | 0.4 | 0.4 | 0.8718 | 0.6568 | 0.5787 | 0.1246 | ≤0.024991 |
| | | 0.6 | 0.9832 | 0.9547 | 0.6098 | 0.1955 | ≤0.024998 |
| | | 0.2 | 0.6574 | 0.1005 | 0.4685 | 0.0213 | ≤0.024980 |
| | 0.6 | 0.4 | 0.8436 | 0.5814 | 0.5162 | 0.0852 | ≤0.024762 |
| | | 0.6 | 0.9763 | 0.9358 | 0.6325 | 0.2815 | ≤0.024994 |
| 8 | | 0.2 | 0.8491 | 0.2039 | 0.7989 | 0.0712 | ≤0.024993 |
| | 0 | 0.4 | 0.9285 | 0.6217 | 0.8159 | 0.1655 | ≤0.024971 |
| | | 0.6 | 0.9899 | 0.9457 | 0.8338 | 0.2817 | ≤0.024991 |
| | | 0.2 | 0.7061 | 0.2326 | 0.5978 | 0.0717 | ≤0.024925 |
| | 0.2 | 0.4 | 0.9004 | 0.7293 | 0.6246 | 0.1585 | ≤0.024988 |
| | | 0.6 | 0.9916 | 0.9773 | 0.6544 | 0.2488 | ≤0.024958 |
| | | 0.2 | 0.8324 | 0.1317 | 0.7355 | 0.0377 | ≤0.024922 |
| | 0.4 | 0.4 | 0.9084 | 0.5189 | 0.7386 | 0.0901 | ≤0.024986 |
| | | 0.6 | 0.9837 | 0.9126 | 0.7678 | 0.1750 | ≤0.024953 |
| | | 0.2 | 0.8205 | 0.0889 | 0.6751 | 0.0257 | ≤0.024984 |
| | 0.6 | 0.4 | 0.8895 | 0.4311 | 0.7004 | 0.0693 | ≤0.024967 |
| | | 0.6 | 0.9715 | 0.8496 | 0.7543 | 0.1688 | ≤0.024956 |

## 4 Application to real data

In order to illustrate our approach we analyse daily returns of time series of Morgan Stanley Capital International (MSCI) Developed Markets indices designed to measure the equity market performance of developed markets. The Dataset includes the following markets: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hong Kong, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the United Kingdom and the United States. We restrict to the time series of daily log–returns $(x_t^1, \ldots, x_t^d)$, $d = 23, t = 1, \ldots, T$, in the period from June 4, 2002 to June 10, 2010 (T=2093 observations; Source: Datastream) in order to provide a direct comparison with the results by De Luca and Zuccolotto (2011).

We preliminary fit AR-GARCH models to each series of returns with Student-t distributed errors to account for heavy tails. For all time series we then perform

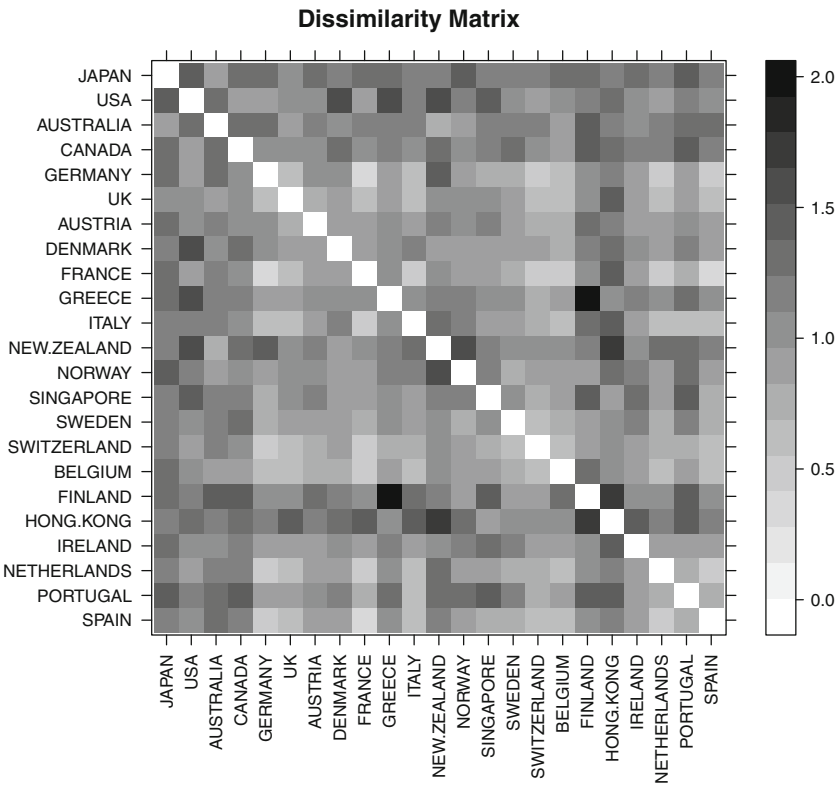


**Fig. 1** Dissimilarity matrix of the MSCI World Index Data constituents according to the described method

Box-Pierce and Ljung-Box tests at lags 1 and 5, to check for residual autocorrelation, ARCH tests at lags 1 and 5, for autoregressive conditional heteroscedasticity and Kolmogorov-Smirnov test to check for the Student hypothesis for the standardized residuals. The estimation results show a reasonable fit for all time series (available upon request). The standardized residuals from each time series are rescaled to the interval $[0, 1]$ thus obtaining the pseudo-observations $(z_t^1, \ldots, z_t^{23})$ on $[0, 1]^{23}$ which represent the empirical copula among the time series of returns. For the estimation of TDC, we consider $m = 91$ block maxima where each block contains $2093/91 = 23$ elements from each time series of residuals (i.e. we focus approximately to monthly maxima). Then, the pairwise lower TDCs $\lambda_{ij}^L$ are estimated non-parametrically by the described procedure. The total number of estimated coefficients is $d(d-1)/2 = 253$, resulting in a $23 \times 23$ symmetric matrix. Given estimates $\hat{\lambda}_{ij}^L$, the dissimilarity matrix is computed as in (2.6) (see Fig. 1). Such a matrix can be used as input for several clustering algorithms.

Among hierarchical clustering techniques the complete linkage method is chosen in order to achieve more useful hierarchies than single or average linkage from a pragmatic point of view. Moreover, it can be used on data that are not restricted to Euclidean distances. Looking at the dendrogram produced by complete linkage scheme

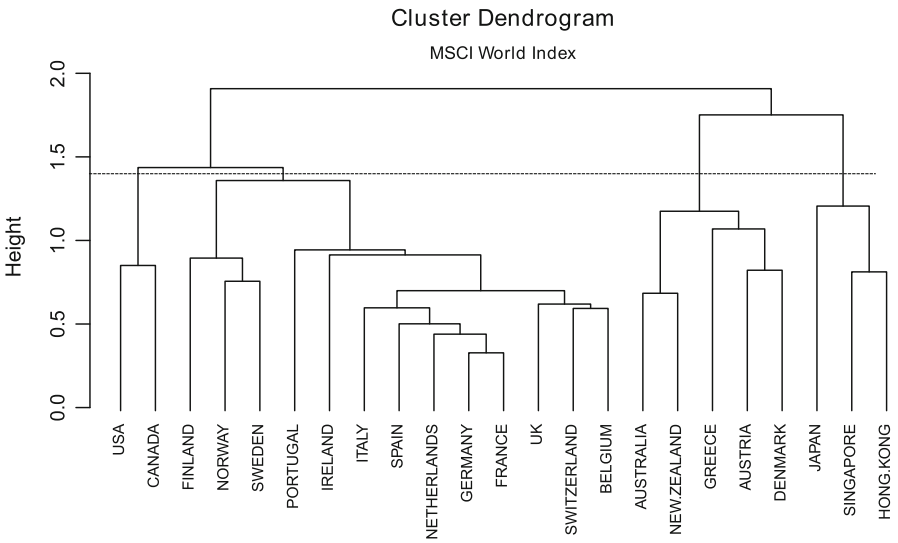


**Fig. 2** Dendrogram of the MSCI World Index Data constituents according to complete linkage. Cutting at about height 1.4 a 4-clusters solution is obtained

**Table 10** Hierarchical Clustering of MSCI World Index Data

| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|
| Belgium | Australia | Hong Kong | Canada |
| Finland | Austria | Japan | USA |
| France | Greece | Singapore | |
| Germany | Denmark | | |
| Ireland | New Zealand | | |
| Italy | | | |
| Netherlands | | | |
| Norway | | | |
| Portugal | | | |
| Spain | | | |
| Sweden | | | |
| Switzerland | | | |
| UK | | | |

(Fig. 2), we find out that $k = 4$ can be considered a good solution, where $k$ denotes the number of clusters selected. Table 10 reports the corresponding cluster composition. Notice that the results can be interpreted in terms of geographic proximity: the lower tail dependence tends to be higher within European markets, where the Scandinavian countries are grouped together as well as USA and Canada; Pacific countries tend to be divided in two separate clusters where New Zealand and Australia are joined together as well as Hong Kong, Japan and Singapore. Moreover, it seems that European markets are splitted in two groups, being some small markets inserted in a separate group (Austria and Denmark).

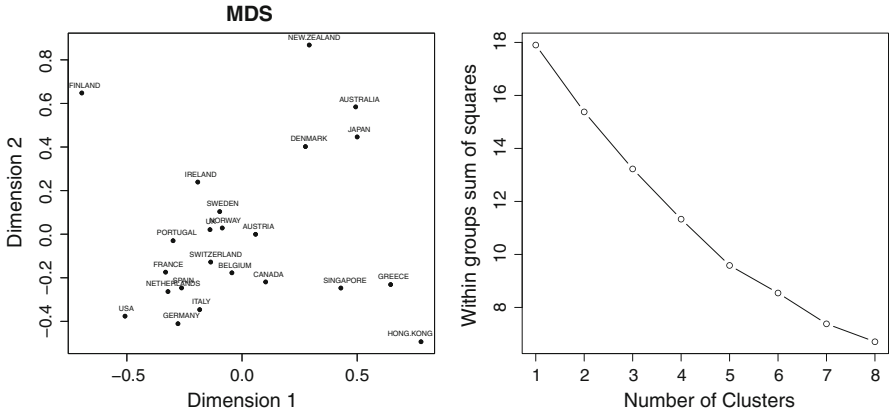


**Fig. 3** *Left* Two-dimensional MDS configuration. *Right* Within groups sum of squares versus the number of clusters in a $K$-means solution

In order to provide a further comparison, we run $K$-means algorithm on the dissimilarity matrix $\Delta = (\Delta_{ij})$. Before doing this step, as suggested in De Luca and Zuccolotto (2011), we need to convert the information coming from $\Delta$ into a set of $d$ vectors in a high-dimensional Euclidean space. As discussed in Sect. 2.3, we can adopt a non–metric MDS procedure so that each time series $(x_t^i)$ will be represented by a $q$-dimensional vector $\boldsymbol{x}_i$. Starting from an initial configuration for $q = 2$, the Shepard–Kruskal algorithm iteratively improves the accuracy of the final representation by increasing $q$ until the minimum stress of the corresponding configuration is lower than 2.5%. The final configuration results in a set of $d = 23$ points of dimension $q = 10$, corresponding to a stress value of 0.0242. Left part of Fig. 3 displays the two-dimensional MDS configuration, characterized by a stress value $\min(s) = 0.2283$. The obtained 10-dimensional points configuration can be used as input for $K$-means algorithm. As said, unlike hierarchical clustering, $K$-means clustering requires that the number of clusters to extract be specified in advance.

A plot of the within-groups sum of squares against the number of clusters extracted can help determine the appropriate number of clusters. A bend in the graph can suggest the appropriate number of clusters. From the right part of Fig. 3 we can observe that the decreasing profile in the within groups sum of squares when $k$ increases from 4 to 5 seems to be higher than the decreasing profile when $k$ increases from 5 to 6, suggesting that a 5-clusters solution can be considered an appropriate choice.

The cluster memberships are listed in Table 11 and are quite consistent with the results in Table 10. Again these results can be mainly interpreted in terms of geographic proximity, although some small European markets are separated from the rest of European markets.

The cluster solutions we carry out from the two procedures can be directly compared with the results obtained by De Luca and Zuccolotto (2011), although in our analysis we perform hierarchical clustering in addition to partitioning clustering and adopt different criteria in the choice of the number of clusters. Moreover, we would like to stress again that our procedure does not require any parametric assumption on

**Table 11** $K$-means clustering of MSCI World Index Data

| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|
| Belgium | Austria | Denmark | Australia | Greece |
| France | Canada | Finland | Japan | Hong Kong |
| Germany | Ireland | Norway | New Zealand | Singapore |
| Italy | USA | Sweden | | |
| Netherlands | | | | |
| Portugal | | | | |
| SPAIN | | | | |
| Switzerland | | | | |
| UK | | | | |

the copula linking the pairwise financial assets, which can be considered the main advantage of the proposed method.

## 5 Conclusions

We have presented a procedure for clustering time series according to their tail behaviour. The procedure follows three steps: first, a copula-based time series model is fitted to univariate time series; second, the pairwise tail dependence coefficients is computed via a non-parametric procedure; finally, a dissimilarity matrix is created from these coefficients and some standard clustering procedures are adopted.

The results are related to the work by De Luca and Zuccolotto (2011), where a similar procedure was investigated by using a parametric estimation. Moreover, contrarily to the latter work, our study suggests that the additional step of performing a MDS of the dissimilarity matrix does not provide a real advantage.

A direct application of the given procedure is mainly related to the field of portfolio optimization and selection. In fact, the proposed clustering procedure may be adopted as a graphical tool to visualize linkages among financial time series that are related to the tail of their joint distribution. In particular, it can also be exploited to perform an automatic selection procedure for financial portfolios trying to hedge against extreme risks (De Luca and Zuccolotto 2013). Alternatively, one may also use the clustering information as a constraint in a portfolio optimization problem, for instance by imposing to select exactly one asset in each sub-group (for more details about this procedure, see the work by Cesarone et al. (2013)).

In general, such clustering procedures may be applied in all fields where the (joint) tail behaviour of different time series is of interest; see, e.g., environmental science (see, for instance, Salvadori et al. 2007). However, the direct application of such techniques need a special care, since the sample size is often of limited length in this kind of data and the estimation of the tail may be not accurate.

**Acknowledgments** The first and second author acknowledge the support of Free University of Bozen-Bolzano, School of Economics and Management, via the project MODEX. The second author would like to thank Claudia Czado and Eike Brechmann (TU Munich, Germany) for useful comments and discussions.

## References

Abdous B, Ghoudi K, Khoudraji A (1999) Non-parametric estimation of the limit dependence function of multivariate extremes. Extremes 2:245–268

Bastos J, Caiado J (2013) Clustering financial time series with variance ratio statistics. Quant Financ. doi:10.1080/14697688.2012.726736

Beirlant J, Goegebeur Y, Teugels J, Segers J (2004) Statistics of extremes. Wiley series in probability and statistics. Wiley, Chichester. Theory and applications, With contributions from Daniel De Waal and Chris Ferro

Bonanno G, Caldarelli G, Lillo F, Micciché S, Vandewalle N, Mantegna R (2004) Networks of equities in financial markets. Eur Phys J B 38(2):363–371

Brechmann E (2014) Hierarchical Kendall copulas: properties and inference. Can J Stat 42(1):78–108

Brida J, Adrián-Risso W (2010) Hierarchical structure of the German stock market. Expert Syst Appl 37(5):3846–3852

Capéraà P, Fougères A-L, Genest C (1997) A nonparametric estimation procedure for bivariate extreme value copulas. Biometrika 84(3):567–577

Cesarone F, Scozzari A, Tardella F (2013) A new method for mean-variance portfolio optimization with cardinality constraints. Ann Oper Res 205(1):213–234

Czado C (2010) Pair-copula constructions of multivariate copulas. In: Jaworki P, Durante F, Härdle W, Rychlik T (eds) Copula theory and its applications, Lecture notes in statistics: proceedings, vol 198, Springer, Berlin, pp 93–109

De Luca G, Rivieccio G, Zuccolotto P (2010) Combining random forest and copula functions: a heuristic approach for selecting assets from a financial crisis perspective. Intell Syst Account financ Manag 17(2):91–109

De Luca G, Zuccolotto P (2011) A tail dependence-based dissimilarity measure for financial time series clustering. Adv Data Anal Classif 5(4):323–340

De Luca G, Zuccolotto P (2013) A conditional value-at-risk based portfolio selection with dynamic tail dependence clustering. MPRA Working Paper Series, 50129. http://mpra.ub.uni-muenchen.de/50129/

Durante F, Fernández-Sánchez J, Pappadà R (2014) Copulas, diagonals and tail dependence. Fuzzy Sets Syst. doi:10.1016/j.fss.2014.03.014

Durante F, Foscolo E (2013) An analysis of the dependence among financial markets by spatial contagion. Int J Intell Syst 28(4):319–331

Durante F, Foscolo E, Sabo M (2013) A spatial contagion test for financial markets. In: Kruse R, Berthold M, Moewes C, Gil M, Grzegorzewski P, Hryniewicz O (eds) Synergies of soft computing and statistics for intelligent data analysis, volume 190 of Advances in intelligent systems and computing. Springer, Berlin, pp 313–320

Durante F, Hofert M, Scherer M (2010) Multivariate hierarchical copulas with shocks. Methodol Comput Appl Probab 12(4):681–694

Durante F, Jaworski P (2010) Spatial contagion between financial markets: a copula-based approach. Appl Stoch Models Bus Ind 26(5):551–564

Durante F, Pappadà R, Torelli N (2013) Clustering of financial time series in risky scenarios. Adv Data Anal Classif. doi:10.1007/s11634-013-0160-4

Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling extremal events for insurance and finance, volume 33 of Applications of mathematics (New York). Springer, Berlin

Frahm G, Junker M, Schmidt R (2005) Estimating the tail-dependence coefficient: properties and pitfalls. Insur Math Econ 37(1):80–100

Genest C, Segers J (2009) Rank-based inference for bivariate extreme value copulas. Ann Stat 37(5B):2990–3022

Gudendorf G, Segers J (2010) Extreme-value copulas. In: Jaworski P, Durante F, Härdle W, Rychlik T (eds) Copula theory and its applications, Lecture notes in statistics: proceedings, vol 198, Springer, Berlin, pp 127–145

Hering C, Hofert M, Mai J-F, Scherer M (2010) Constructing hierarchical Archimedean copulas with Lévy subordinators. J Multivar Anal 101(6):1428–1433

Hubert L, Arabie P (1985) Comparing partitions. J Classif 2:193–218

Jaworski P (2010) Tail behaviour of copulas.In: Jaworski P, Durante F, Härdle W, Rychlik T (eds) Copula theory and its applications, volume 198 of Lecture notes in statistics: proceedings, Springer, Berlin, pp 161–186

Jaworski P, Durante F, Härdle WK (eds) (2013) Copulae in mathematical and quantitative finance, Lecture notes in statistics: proceedings, vol 213, Springer, Berlin

Jaworski P, Durante F, Härdle WK, Rychlik T (eds) (2010) Copula theory and its applications, Lecture notes in statistics: proceedings, vol 198, Springer, Berlin

Joe H (1997) Multivariate models and dependence concepts, volume 73 of Monographs on statistics and applied probability. Chapman & Hall, London

Kaufman L, Rousseeuw P (1990) Finding groups in data. Applied probability and statistics. Wiley, New York, Wiley Series in probability and mathematical statistics

Kruskal J (1964) Nonmetric multidimensional scaling: a numerical method. Psychometrika 29:115–129

Longin F, Solnik B (2001) Extreme correlation of international equity markets. J Financ 56(2):649–676

Mantegna R (1999) Hierarchical structure in financial markets. Eur Phys J B 11(1):193–197

Okhrin O, Okhrin Y, Schmid W (2013) On the structure and estimation of hierarchical Archimedean copulas. J Econ 173(2):189–204

Otranto E (2008) Clustering heteroskedastic time series by model-based procedures. Comput Stat Data Anal 52(10):4685–4698

Pattarin F, Paterlini S, Minerva T (2004) Clustering financial time series: an application to mutual funds style analysis. Comput Stat Data Anal 47(2):353–372

Patton A (2012) A review of copula models for economic time series. J Multivar Anal 110:4–18

Patton A (2013) Copula methods for forecasting multivariate time series. In: Elliott G, Timmermann A (eds) Handbook of economic forecasting, vol 2. Springer, Berlin

Piccolo D (1990) A distance measure for classifying ARIMA models. J Time Ser Anal 11(2):153–164

Pickands J (1981) Multivariate extreme value distributions. In: Proceedings of the 43rd session of the international statistical Institute, Vol 2 (Buenos Aires, 1981), vol 49, pp 859–878, 894–902

Rand WM (1971) Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 66(336):846–850

Remillard B (2010) Goodness-of-fit tests for copulas of multivariate time series. SSRN eLibrary

Salvadori G, De Michele C, Kottegoda NT, Rosso R (2007) Extremes in nature: an approach using copulas, volume 56 of Water science and technology Library. Springer, Dordrecht

Shepard RN (1962) The analysis of proximities: multidimensional scaling with an unknown distance function. I. Psychometrika 27(2):125–140

Shepard RN (1962) The analysis of proximities: multidimensional scaling with an unknown distance function. II. Psychometrika 27:219–246

Tola V, Lillo F, Gallegati M, Mantegna R (2008) Cluster analysis for portfolio optimization. J Econ Dyn Control 32(1):235–258