Research Article

Giovanni De Luca* and Paola Zuccolotto

A double clustering algorithm for financial time series based on extreme events

DOI: 10.1515/strm-2015-0026
Received September 16, 2015; revised July 13, 2016; accepted August 24, 2016

Abstract: This paper is concerned with a procedure for financial time series clustering, aimed at creating groups of time series characterized by similar behavior with regard to extreme events. The core of our proposal is a double clustering procedure: the former is based on the lower tail dependence of all the possible pairs of time series, the latter on the upper tail dependence. Tail dependence coefficients are estimated with copula functions. The final goal is to exploit the two clustering solutions in an algorithm designed to create a portfolio that maximizes the probability of joint positive extreme returns while minimizing the risk of joint negative extreme returns. In financial crisis scenarios, such a portfolio is expected to outperform portfolios generated by the traditional methods. We describe the results of a simulation study and, finally, we apply the procedure to a dataset composed of the 50 assets included in the EUROSTOXX index.

Keywords: Financial time series clustering, tail dependence, copula functions, portfolio selection

MSC 2010: 60G70, 62M10

1 Introduction

In the analysis of the relationship between financial returns, the lower tail dependence quantifies the risk of investing on assets for which extremely negative returns could occur simultaneously. This is a very important issue for portfolio selection. Financial institutions have to offer to their customers the chance of investing money in a number of assets. But these assets should be poorly associated in the negative performances, in the sense that a fall in the quotations of an asset should not affect the quotations of the other assets. In order to have a measure of the association between two assets, some classical statistical tools are inadequate. The correlation coefficient is the main measure of association for quantitative data, but it has revealed its limit in this context. The correlation coefficient summarizes the linear relationship between two variables considering the entire probability distributions. However, in presence of non-linear relationships and when the interest is focused on the extremely low values of the variables, it is appropriate to adopt some specific association measures (see [16] for a comprehensive survey). In recent years literature has given much importance to the tail dependence, that is, the dependence between extreme values (see [18]). In particular, we can consider the upper tail dependence, when the interest lies in the very high values, and the lower tail dependence, when the interest is in the very low values.

Portfolio selection techniques are heavily affected by the estimated association of the potential assets. The classical approach has been designed by Markowitz [21] and is known as mean-variance approach (the resulting portfolio is known as mean-variance portfolio). Given n assets, the rationale of the mean-variance approach is that of choosing the weights of the assets, in such a way that the resulting portfolio has a specific expected value and the lowest possible variance. Since the variance is an indicator of the variability, and then

***Corresponding author: Giovanni De Luca:** University of Naples Parthenope, Via G. Parisi 13, 80132 Naples, Italy, e-mail: giovanni.deluca@uniparthenope.it
Paola Zuccolotto: University of Brescia, C. da S. Chiara 50, 25122 Brescia, Italy, e-mail: paola.zuccolotto@unibs.it

of the risk, the solution of the Markowitz problem provides a diversified portfolio satisfying a fairly general criterion of minimization of the risk, based on linear correlation as association measure between returns. However, in the last decades a number of contributions has been presented in literature enriching the range of opportunities. Another measure of risk built on the entire distribution of the returns has been proposed by Bernardo and Ledoit [3], and later popularized by Keating and Shadwick [19] as Omega. A largely used alternative approach is the minimization of the Conditional Value-at-Risk (CVaR). The Value-at-Risk (VaR) is certainly the most popular measure of financial risk, largely used by financial institutions. It is defined as a threshold loss value, such that the loss on the portfolio over a given time horizon can exceed this value with a given (low) probability. The CVaR is the expected loss given that a loss greater than the VaR has occurred. The problem of minimizing the CVaR has been introduced by Rockafellar and Uryasev [24] and Krokhmal, Palmquist and Uryasev [20].

When the aim is the composition of a portfolio with low association in the extremely low values of the assets, the lower tail dependence coefficient has a dominant role. In [6] we have proposed to classify the assets in groups according to their association between very low returns, measured by the lower tail dependence coefficient. As a result, we have clusters composed of assets with a strong association between extremely low returns, while the assets belonging to different groups present a weak association between extremely low returns. The topic has been also faced by other authors. In [13], Durante, Pappadà and Torelli have proposed to carry out a clustering procedure based on the conditional Spearman's correlation coefficient, and in [14] they have suggested a non-parametric estimation of the tail dependence coefficients, while in [12] Durante and Pappadà have clustered time series according to the pairwise Kendall distribution. A different approach has been studied by DiTraglia and Gerlach [11] exploiting a result from Extreme Value Theory to estimate the tail dependence and use it in portfolio selection.

The problem of time series clustering has also been explored in the literature from different perspectives; see, e.g., [2, 4, 5, 15, 22, 23, 25].

In terms of portfolio selection, we have introduced in [6] the strategy of picking one asset from each of the groups, such that the resulting portfolios are composed of assets with a low probability of a joint collapse. The weights of the selected assets are estimated using one of the known techniques. This research line has been further developed in [9] pursuing the idea that the lower tail dependence coefficients are not time-invariant, but have their own dynamics. As a result, the possible portfolios one can compose also change over time.

Up to now, the strategies proposed in this framework are designed to take into account only the dependence of returns in the lower tail, as it is considered the most crucial issue to prevent severe losses from occurring in financial crisis periods. In this paper we propose to go beyond this point and take into account the association of returns both in the lower and in the upper tail, so as to compose portfolios able to protect against crisis periods, while taking the best of booms. In detail, we present a development of the basic strategy '*pick one asset from each cluster*', proposed in [6, 9]. More specifically, we propose to exploit the results of a second clusterization of the same assets based on the upper tail dependence coefficients. The idea is that of selecting assets which belong to different lower tail dependence-based clusters and, possibly, to a unique upper tail dependence-based cluster. If the last condition cannot be satisfied, we request to get as close as possible. In this case, for each possible selection derived from the lower tail dependence-based clusters, we compute the heterogeneity Gini index according to the position of the assets in the upper tail dependence-based clusters. At the end we opt for the selection which minimizes the Gini index. The rationale behind this strategy is that of selecting a well-diversified portfolio for crisis period that, at the same time, can take advantage of positive extreme events.

The paper is organized as follows. Section 2 describes the clustering procedure and presents the methods for finding the weights to compose a portfolio. In Section 3 the results of an extensive simulation study are illustrated. An application to a real dataset is discussed in Section 4 and, finally, Section 5 concludes.

2 Time series clustering on tail dependence

In this paper we refer to the clustering procedure proposed in [6], where time series of financial returns are clustered using a dissimilarity measure based on tail dependence coefficients, focusing either on the lower or on the upper tail. The dissimilarity measure is defined as

$$\hat{\delta}(\{r_{it}\}, \{r_{jt}\}) = \hat{\delta}_{ij} = -\log(\hat{\lambda}_{ij}),$$

where $\{r_{it}\}_{t=1,\ldots,T}$ and $\{r_{jt}\}_{t=1,\ldots,T}$ denote the time series of returns of two assets i and j, and $\hat{\lambda}_{ij}$ is their estimated (lower or upper) tail dependence coefficient. For details about the estimation of tail dependence coefficients by means of copula functions, see [16, 17].

Given the time series of the returns of p assets, in the following we will denote by $\hat{\Lambda}$ the $p \times p$ symmetric matrix containing the (upper or lower) tail dependence coefficients between all the possible pairs of returns. The diagonal of $\hat{\Lambda}$ is composed of ones. Later, when we will need to distinguish between lower and upper tail dependence, the matrix will be denoted as $\hat{\Lambda}_L$ and $\hat{\Lambda}_U$, respectively.

Given $\hat{\Lambda}$, the clustering procedure is composed of two steps. In step 1, starting from the dissimilarity matrix $\hat{\Delta} = (\hat{\delta}_{ij})_{i,j=1,\ldots,p} = -\log(\hat{\Lambda})$, an *optimal* representation of the p time series $\{r_{1t}\}, \ldots, \{r_{pt}\}$ as p points $\mathbf{y}_1, \ldots, \mathbf{y}_p$ in $\mathbb{R}^q$ is found by means of multidimensional scaling (MDS). The term *optimal* means that the Euclidean distance matrix $D = (d_{ij})_{i,j=1,\ldots,p}$, with $d_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\|$, of the p points $\mathbf{y}_1, \ldots, \mathbf{y}_p$ in $\mathbb{R}^q$ has to fit as closely as possible the dissimilarity matrix $\hat{\Delta}$. The extent to which the interpoint distances d_{ij} "match" the dissimilarities $\hat{\delta}_{ij}$ is measured by an index called *stress*, which should be as low as possible. The algorithm of MDS works for a given value of the dimension q, which has to be given in input. So, it is proposed to start with the dimension $q = 2$ and then to repeat the analysis by increasing q until the minimum stress of the corresponding optimal configuration is lower than a given threshold $\bar{s}$. In step 2, the p points $\mathbf{y}_1, \ldots, \mathbf{y}_p$ in $\mathbb{R}^q$ are clustered using a k-means algorithm. In [6] we have shown simulation studies where this two-step procedure outperforms the application of hierarchical clustering directly on the dissimilarity matrix $\hat{\Delta}$. However, this is still an open issue. For example, Durante, Pappadà and Torelli [13, 14] propose a similar approach where the use of hierarchical clustering instead of the k-means algorithm allows them to avoid the MDS step.

The clusters obtained with this procedure are composed of assets characterized by high tail dependence in the (lower or upper) tail. We point out (see [6–9]) that the clustering solution may be exploited for a preliminary decision about which stocks should be included in a portfolio. In other words, the idea is to use the clustering solution to select a small number of stocks to invest on, from among the p stocks available. Then, a portfolio is constructed by estimating proper weights using the common portfolio selection techniques (see Section 2.1).

When the clustering solution based on the lower tail is used, the selection is made by including in the portfolio assets belonging to different *lower tail-based* clusters, that allows to counterbalance possible extreme losses. On the other hand, the clustering solution based on the upper tail may be used, as well. In this case the opposite strategy should be followed: investing on assets belonging to the same *upper tail-based* cluster allows to take advantage of simultaneous extreme profits. In [6–9] we have shown examples where portfolios including the stocks selected by means of *lower tail-based* clusters outperform the classical portfolio selection strategies. Instead, the performance of portfolios built relying on the *upper tail-based* clusters has not been explored so far, because the latest years financial situation suggested to protect from bears rather than take advantage of bulls.

In this paper we propose to select the stocks to include in the portfolio relying both on the *lower tail-based* and *upper tail-based* clusters. Again, a defensive approach is adopted, in the sense that the *upper tail-based* clusters are used as a second-best criterion to be exploited once the requirements based on the *lower tail-based* clusters have been fulfilled.

The proposed procedure is described in the next subsection.

2.1 Portfolio selection based on upper and lower tail dependence

Here we describe the procedure proposed to select the stocks to include in the portfolio, by exploiting both the *lower tail-based* and the *upper tail-based* clustering solution. Let $GL_1, GL_2, \dots, GL_{K_L}$ and $GU_1, GU_2, \dots, GU_{K_U}$ be, respectively, the *lower tail-based* and the *upper tail-based* clustering solution, obtained by carrying out the procedure described in Section 2 on p time series of returns $\{r_{it}\}$, $t = 1, \dots, T$, $i = 1, \dots, p$. The generic GL_{k_L} and GU_{k_U}, $k_L = 1, 2, \dots, K_L$ and $k_U = 1, 2, \dots, K_U$, denote the set of stocks included, respectively, in the k_Lth cluster of the *lower tail-based* solution and in the k_Uth cluster of the *upper tail-based* solution.

The selection proceeds through three steps:

(i) First selection, based on $GL_1, GL_2, \dots, GL_{K_L}$: the selection criterion based on the *lower tail-based* clustering solution requires to include in the portfolio stocks belonging to different clusters, thus avoiding to invest on assets that could be characterized by simultaneous extreme losses. We denote by n_{k_L} the number of stocks belonging to cluster k_L, i.e. the cardinality of the set G_{k_L}. Then the criterion allows us to define, on the whole, $S^{(0)} = \prod_{k_L=1}^{K_L} n_{k_L}$ possible portfolios composed of K_L stocks. We denote by $\mathbf{C}^{(0)} = \{C_1^{(0)}, C_2^{(0)}, \dots, C_{S^{(0)}}^{(0)}\}$ the set of all possible candidate portfolios, that will simply be called *candidates* in the following.

(ii) Second selection, based on $GU_1, GU_2, \dots, GU_{K_U}$: at the second step we select, from $\mathbf{C}^{(0)}$, a subset of candidates that exhibit desirable features according to the *upper tail-based* clustering solution. Specifically, when the upper tail is considered, a reasonable criterion consists of including in the portfolio stocks belonging to the same cluster, so as to invest on assets that could be characterized by simultaneous extreme profits. So, the idea is to select from $\mathbf{C}^{(0)}$ the candidates composed of stocks as much as possible belonging to the same *upper tail-based* cluster. To do that, for each candidate $C_s^{(0)}$ we compute the heterogeneity index

$$\gamma_s = \frac{K_U}{K_U - 1}\left[1 - \sum_{k_U=1}^{K_U}\left(\frac{n_{s,k_U}}{K_L}\right)^2\right],$$

where n_{s,k_U} is the number of stocks of $C_s^{(0)}$ belonging to the k_Uth *upper tail-based* cluster. The index γ_s is the Gini heterogeneity index adapted to this context, and informs on the amount of heterogeneity of candidate $C_s^{(0)}$ with respect to the *upper tail-based* clustering solution. If $\gamma_s = 0$, the stocks composing candidate $C_s^{(0)}$ belong all to the same *upper tail-based* cluster. So, we desire γ_s to be as low as possible and we will select from $\mathbf{C}^{(0)}$ the candidates that minimize this index. Formally, we denote by $\mathbf{C} \subseteq \mathbf{C}^{(0)}$, $\mathbf{C} = \{C_1, C_2, \dots, C_S\}$, the subset of candidates selected at the second step, given by

$$\mathbf{C} = \left\{C_s^{(0)} : \gamma_s = \min_{s=1,2,\dots,S^{(0)}} (\gamma_s)\right\}.$$

(iii) Final selection: the final portfolio is selected from among the candidates in $\mathbf{C}$, by means of financial criteria. In detail, firstly the weights of all the candidate portfolios are estimated by minimizing the CVaR, then the best portfolio is selected either with the minimum CVaR or the maximum Omega index criterion.

3 Simulation study

In this section we describe the results of a simulation study investigating how the whole proposed procedure works. The main aim is to understand how effective the idea to exploit the clustering solution for portfolio selection is.

In Sections 3.1 and 3.2, we will describe the data generating process and the results obtained in the two phases, the clustering and the portfolio selection, of the proposed procedure.

As far as it concerns the data generating process, it is worth pointing out that to define a multivariate random variable able to allow different lower and upper tail dependence structures is a challenging task. Up to our knowledge, at the moment there is in the literature no procedure able to generate multivariate data with two (different) given matrices of lower and upper tail dependence coefficients. Some studies in this sense can be found in [1, 10], but they are not suited to the multivariate structure needed in this study. So, we will define a new random variable by joining two different multivariate Student-t variables, determining the behavior of the lower and the upper tail in a separate fashion.

In Section 3.2 the results of the portfolio selection are described, focusing on the out-of-sample performance of the proposed portfolios compared to two benchmark alternatives.

3.1 The data generating process

The data generating process for the simulation study has been designed so as to obtain simulated heteroskedastic financial returns with a multivariate structure allowing for given (different) lower and upper tail dependence coefficients.

Let $\{\mathbf{R}_t\} = \{R_{1,t}, R_{2,t}, \dots, R_{p,t}\}$ be a p-dimensional stochastic process with

$$R_{i,t} = \sigma_{i,t}\tau_{i,t}, \quad i = 1, 2, \dots, p,$$

and $\sigma^2_{i,t}$ following a GARCH(1, 1) specification. In order to impress to $\{\mathbf{R}_t\}$ different lower and upper tail dependence structures, the probability density function of $\tau_{i,t}$ has been defined as the ith marginal distribution of a p-variate random variable τ_t, built as a combination of two p-variate Student-t random variables T_L and T_U, with variance-covariance matrices Σ_L and Σ_U and degrees of freedom ν_L and ν_U, respectively. The random variables τ_t are assumed to be independent and identically distributed over t. According to our definition, the probability density function $f(\boldsymbol{\tau}_t)$ is given by

$$f(\boldsymbol{\tau}_t; \Sigma_L, \Sigma_U, \nu_L, \nu_U) = f_L(\boldsymbol{\tau}_t; \Sigma_L, \nu_L)I(\boldsymbol{\tau}_t) + f_U(\boldsymbol{\tau}_t; \Sigma_U, \nu_U)(1 - I(\boldsymbol{\tau}_t)),$$

where f_L and f_U are the probability density functions of T_L and T_U, respectively, and

$$I(\boldsymbol{\tau}_t) = \begin{cases} 0 & \text{if } \boldsymbol{\tau}_t'\mathbf{1} \leq 0, \\ 1 & \text{if } \boldsymbol{\tau}_t'\mathbf{1} > 0, \end{cases}$$

with $\mathbf{1}$ a properly sized vector of ones.

This definition allows for a random variable with a domain $\mathcal{D}$ ideally divided into $\mathcal{D}_L = \{\boldsymbol{\tau}_t \in \mathbb{R} : \boldsymbol{\tau}_t'\mathbf{1} \leq 0\}$ and $\mathcal{D}_U = \{\boldsymbol{\tau}_t \in \mathbb{R} : \boldsymbol{\tau}_t'\mathbf{1} > 0\}$, and different behaviors in these two portions of space, determined by the two random variables T_L and T_U. More specifically, since $\mathcal{D}_L$ contains the lower tail of the distribution and $\mathcal{D}_U$ the upper tail, the variance-covariance matrix Σ_L determines the lower tail dependence coefficients (and the implied lower tail clustering structure), while Σ_U similarly governs the upper tail.

In our simulation study, we set $p = 20$, $\nu_L = (4, 4, \dots, 4)$, $\nu_U = (5, 5, \dots, 5)$. The variance-covariance matrix Σ_L for the lower tail is a block matrix with square matrices $\mathbf{w}_L^{(k_L)}$ on the diagonal ($k_L = 1, 2, 3, 4$) of sizes $n_{k_L} = 12, 2, 2, 4$ with elements all equal to 0.7 (except for the diagonal, composed of ones) and matrices $\mathbf{b}_L$ outside, properly sized with elements all equal to 0.3. The variance-covariance matrix Σ_U for the upper tail is still a block matrix with square matrices $\mathbf{w}_U^{(k_U)}$ on the diagonal ($k_U = 1, 2, 3, 4, 5$) of sizes $n_{k_U} = 6, 7, 2, 3, 2$ with elements all equal to 0.7 (except for the diagonal, composed of ones) and matrices $\mathbf{b}_U$ outside, properly sized with elements all equal to 0.3.

Since the tail dependence coefficients of two Student-t random variables are univocally determined by their linear correlation coefficient and their degrees of freedom, the implied lower and upper tail dependence structures of τ_t can be derived, separately for the lower and the upper tail, from Σ_L, Σ_U, ν_L, ν_U. So, the lower and upper tail dependence matrices Λ_L and Λ_U have the same block structure of Σ_L and Σ_U with the within and between groups lower tail dependence coefficients given, respectively, by $\lambda_{w,L} = 0.3907$ and $\lambda_{b,L} = 0.1618$, while the within and between groups upper tail dependence coefficients are, respectively, $\lambda_{w,U} = 0.3423$ and $\lambda_{b,U} = 0.1224$.

Thanks to this setting, the defined p-variate process $\{\mathbf{R}_t\} = \{R_{1,t}, R_{2,t}, \dots, R_{p,t}\}$ has the following features:

- Each process $\{R_{i,t}\}$, $i = 1, 2, \dots, p$ has a marginal GARCH structure.
- In the lower tail we recognize the presence of four groups of processes:

  $$GL_1 = \{R_{1,t}, \dots, R_{12,t}\}, \quad GL_2 = \{R_{13,t}, R_{14,t}\}, \quad GL_3 = \{R_{15,t}, R_{16,t}\}, \quad GL_4 = \{R_{17,t}, \dots, R_{20,t}\},$$

  characterized by a moderately high tail dependence between pairs belonging to the same group (the *within* groups lower tail dependence coefficient is $\lambda_{w,L} = 0.3907$) and a moderately low tail dependence between pair belonging to different groups (the *between* groups lower tail dependence coefficient is $\lambda_{b,L} = 0.1618$). Note that the values of $\lambda_{w,L}$ and $\lambda_{b,L}$ are very similar to what is usually observed in practice with financial data.
- In the upper tail we recognize the presence of five groups of processes:

  $$\begin{aligned} &GU_1 = \{R_{1,t}, \dots, R_{6,t}\}, \quad &&GU_2 = \{R_{7,t}, \dots, R_{13,t}\}, \quad &&GU_3 = \{R_{14,t}, R_{15,t}\},\\ &GU_4 = \{R_{16,t}, \dots, R_{18,t}\}, \quad &&GU_5 = \{R_{19,t}, R_{20,t}\}, \end{aligned}$$

  characterized by a moderately high tail dependence between pairs belonging to the same group (the *within* groups lower tail dependence coefficient is $\lambda_{w,U} = 0.3432$) and a moderately low tail dependence between pairs belonging to different groups (the *between* groups lower tail dependence coefficient is $\lambda_{b,U} = 0.1224$). Again, note that the values $\lambda_{w,U}$ and $\lambda_{b,U}$ are very plausible.
- According to the rule described in Section 2.1, the following twelve sets of candidates are selected from among the p processes $\{R_{1,t}\}, \dots, \{R_{p,t}\}$ for the portfolio determination:

  $$\begin{aligned} C_1 &= \{R_{7,t}, R_{13,t}, R_{16,t}, R_{17,t}\}, & C_2 &= \{R_{7,t}, R_{13,t}, R_{16,t}, R_{18,t}\},\\ C_3 &= \{R_{8,t}, R_{13,t}, R_{16,t}, R_{17,t}\}, & C_4 &= \{R_{8,t}, R_{13,t}, R_{16,t}, R_{18,t}\},\\ C_5 &= \{R_{9,t}, R_{13,t}, R_{16,t}, R_{17,t}\}, & C_6 &= \{R_{9,t}, R_{13,t}, R_{16,t}, R_{18,t}\},\\ C_7 &= \{R_{10,t}, R_{13,t}, R_{16,t}, R_{17,t}\}, & C_8 &= \{R_{10,t}, R_{13,t}, R_{16,t}, R_{18,t}\},\\ C_9 &= \{R_{11,t}, R_{13,t}, R_{16,t}, R_{17,t}\}, & C_{10} &= \{R_{11,t}, R_{13,t}, R_{16,t}, R_{18,t}\},\\ C_{11} &= \{R_{12,t}, R_{13,t}, R_{16,t}, R_{17,t}\}, & C_{12} &= \{R_{12,t}, R_{13,t}, R_{16,t}, R_{18,t}\}. \end{aligned}$$

  Note that all the twelve configurations are composed of processes belonging to four different groups in the lower tail and two different groups in the upper tail. This corresponds to a heterogeneity index $\gamma_s = 0.625$, that is the minimum value that can be reached with this configuration of the process $\{\mathbf{R}_t\}$. Also in this case, we point out that this corresponds to what usually happens with empirical data, as very low values of γ are not likely to occur in practice.

3.2 Results

The simulation study has been repeated for *Niter* = 25 iterations. At each iteration we drew a p-dimensional time series $\mathbf{r}_t = (r_{1,t}, r_{2,t}, \dots, r_{p,t})$, $p = 20$, of $T = 1000$ observations from the data generating process described in Section 3.1. For each univariate series, we removed heteroskedasticity by filtering data through a Student-t GARCH model with maximum likelihood estimated parameters. The obtained standardized residuals have then been used to derive the estimated lower and upper tail dependence matrices, $\hat{\Lambda}_L$ and $\hat{\Lambda}_U$.

It is worth noting that (a) the distributional assumption used to estimate the parameters of the GARCH models is not exactly the same as the random variables $\tau_{i,t}$ that have generated data, and (b) the estimation of the tail dependence coefficients has been made by means of the Joe–Clayton copula, that uses a joint probabilistic structure different from that of the multivariate random variable τ_t in the data generating process. These two points allow us to assess the robustness of the proposed procedure to misspecified distributional assumptions.

In the following, we summarize the results of the two main steps of the proposed procedure: the clustering and the portfolio determination.

i	$\pi_{i,A}$	$\pi_{i,B}$	$\bar{D}_{i,A}$	$\bar{D}_{i,B}$
C1	51.78	50.35	0.0007	0.0005
	$(< 10^{-15})$	(0.0009)	$(< 10^{-15})$	$(< 10^{-15})$
C2	53.12	53.00	0.0015	0.0013
	$(< 10^{-15})$	$(< 10^{-15})$	$(< 10^{-15})$	$(< 10^{-15})$

Table 1. Results of the simulation study, indices $\pi_{i,j}$ (in parenthesis the p-value of the test $H_0 : \pi_{i,A} = 50$) and $\bar{D}_{i,j}$ (in parenthesis the p-value of the test $H_0 : \bar{D}_{i,j} = 0$).

Clustering. The time series clustering algorithm described in Section 2 has been carried out for each p-dimensional time series $\mathbf{r}_t$ starting from the estimated matrices $\hat{\Delta}_L = -\log(\hat{\Lambda}_L)$ and $\hat{\Delta}_U = -\log(\hat{\Lambda}_U)$. In all the $Niter = 25$ cases, the procedure was able to recover the exact clustering structure impressed to the data generating process $\{\mathbf{R}_t\}$. So, in all the explored cases, the exact twelve sets of candidates $C_1, C_2, \ldots, C_{12}$ were selected for the portfolio determination step.

Portfolio. In the second step, we examined the performance of some portfolios, built according to the strategies described in Section 2.1, on out-of-sample data, also comparing them with two benchmark competitors. In detail, for each simulated series $\mathbf{r}_t = (r_{1,t}, r_{2,t}, \ldots, r_{p,t})$, we determined the optimal weights of the following portfolios:

A. (*Benchmark Portfolio 1*) Markowitz minimum variance portfolio (all the stocks).

B. (*Benchmark Portfolio 2*) Minimum CVaR portfolio (all the stocks).

C. Minimum CVaR portfolios (one portfolio for each set of candidates C_1, $C_2, \ldots, C_{12}$). In order to choose among the twelve portfolios, two possible criteria have been explored (see Section 2):

   C1. minimum CVaR,

   C2. maximum Omega.

For each option, the optimal weights have been determined using all the observations ($t = 1, 2, \ldots, T$) and the performance of the portfolio has been checked on out-of-sample data simulated from the same data generating process. In detail, for each iteration, $SC = 100$ series of $Q = 50$ observations, $\mathbf{r}_t^+ = (r_{1,t}$, $r_{2,t}, \ldots, r_{p,t})$, $t = T + 1, T + 2, \ldots, T + Q$, have been generated. In other words, each series $\mathbf{r}_t^+$ simulates a scenario for the 50 days following the last observation of $\mathbf{r}_t$ and 100 different scenarios are examined for each series $\mathbf{r}_t$. In total, for the $Niter = 25$ iterations, $Niter \cdot SC = 2500$ scenarios are analyzed, globally accounting for $Niter \cdot SC \cdot Q = 12\,500$ out-of-sample observations.

The returns of the two portfolios C1 and C2, built according to the proposed procedure, are compared to the two benchmark portfolios A and B using two indices:

- How many times (%) the return of portfolio i, $rp_{i,t}$, is higher than the return of portfolio j, $rp_{j,t}$:

$$\pi_{i,j} = \frac{\sum_{iter=1}^{Niter} \sum_{sc=1}^{SC} \sum_{q=1}^{Q} I_{sc,iter}(rp_{i,t+q}, rp_{j,t+q})}{Niter \cdot SC \cdot Q} \cdot 100,$$

  where $I_{sc,iter}(rp_{i,t+q}, rp_{j,t+q})$ denotes the indicator function assuming value 1 if, in the sc-th scenario of the iteration number $iter$, we have $rp_{i,t+q} > rp_{j,t+q}$, and 0 otherwise. When $\pi_{i,j} = 50$, portfolio i outperforms portfolio j half the times, meaning that there is no difference between i and j from this point of view. So, inference about the performance of the portfolios can be made by testing the hypothesis $H_0 : \pi_{i,j} = 50$.

- The average difference between the returns of portfolio i and j:

$$\bar{D}_{i,j} = \frac{\sum_{iter=1}^{Niter} \sum_{sc=1}^{SC} \sum_{q=1}^{Q} d_{sc,iter}(rp_{i,t+q}, rp_{j,t+q})}{Niter \cdot SC \cdot Q} \cdot 100,$$

  where $d_{sc,iter}(rp_{i,t+q}, rp_{j,t+q})$ is equal to the difference $rp_{i,t+q} - rp_{j,t+q}$ evaluated at each $iter$ and each sc. When $\bar{D}_{i,j} = 0$, there is no difference between the two portfolios i and j from the point of view of the average difference of returns. So, we have to test the hypothesis $H_0 : \bar{D}_{i,j} = 0$.

Results are reported in Table 1. The two portfolios selected according to the proposed procedure outperform the competitors. All the indices $\pi_{i,j}$ comparing portfolios C1 and C2 to A and B are significantly higher than 50%. Analogously, all the indices $\bar{D}_{i,j}$ are significantly higher than 0.

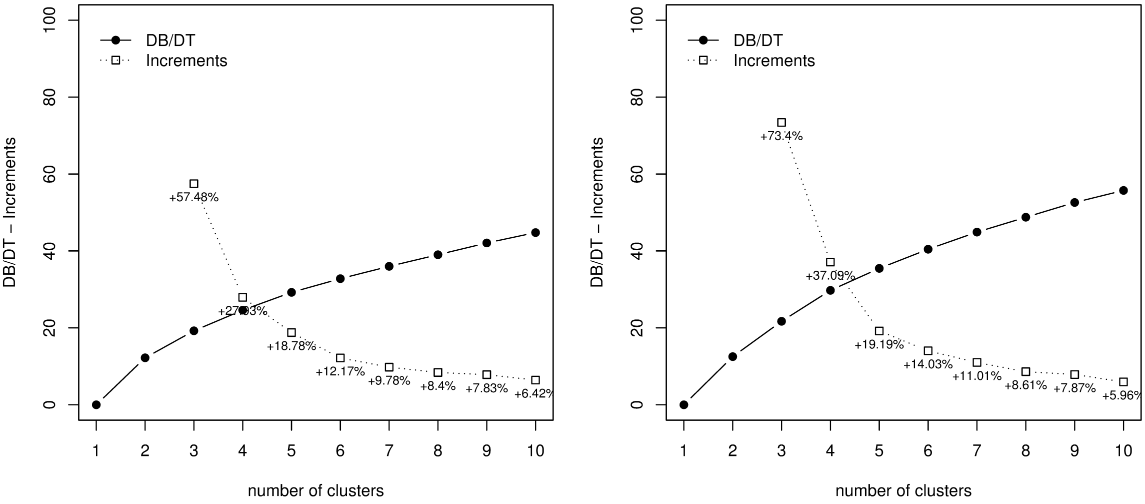

Figure 1. Ratio between deviance and total deviance: pattern and increments versus the number of clusters in the *lower tail-based* (left) and *upper tail-based* (right) clustering solutions.

We point out again that both the tail dependence coefficients $\lambda_{w,L}$, $\lambda_{w,U}$, $\lambda_{b,L}$ and $\lambda_{b,U}$ and the heterogeneity index γ have been fixed so as to resemble as close as possible what usually happens in practice with financial data. As a matter of fact, if we fixed (a) higher values of $\lambda_{w,L}$ and $\lambda_{w,U}$, (b) lower values of $\lambda_{b,L}$ and $\lambda_{b,U}$, and (c) a clustering structure allowing lower values of γ, the performance of our procedure would be even better with respect to the competitors.

4 Empirical analysis of real data

Our procedure is applied to daily log-returns of the 50 stocks included in the EUROSTOXX index observed in the period from January 2, 2008 to December 31, 2013. The total number of returns for each stock is 1540. The EUROSTOXX index is designed to reflect the performance of the largest companies in the Eurozone and so is a measure of the performance of the financial markets in Europe.

Each time series of log-returns has been filtered to remove autocorrelation and heteroskedasticity by applying a univariate Student-*t* AR-GARCH models. The order of the autoregressive component ranges from 0 to 1, while for the heteroskedastic component the GARCH(1, 1) model has provided a satisfactory fit for almost all the series (for a few time-series a GARCH(1, 2) has been estimated). Then we have computed the distribution functions of the standardized residuals. Applying the Joe–Clayton copula, we obtain the 50×50 symmetric matrices containing the estimated lower and upper tail dependence coefficients, respectively $\hat{\Lambda}_L$ and $\hat{\Lambda}_U$.

4.1 Clustering

Starting from $\hat{\Lambda}_L$ and $\hat{\Lambda}_U$, we obtained the corresponding dissimilarity matrices $\hat{\Delta}_L = -\log(\hat{\Lambda}_L)$ and $\hat{\Delta}_U = -\log(\hat{\Lambda}_U)$. Then, we followed the procedure defined in Section 2.1. In doing that, we obtained the clustering as described in Section 2 two times, using firstly Δ_L and secondly Δ_U, in order to obtain, respectively, the two (*lower tail-based* and *upper tail-based*) clustering solutions $GL_1, GL_2, \dots, GL_{K_L}$ and $GU_1, GU_2, \dots, GU_{K_U}$.

In order to select the optimal number of clusters, K_L and K_U, we inspected the graphs presented in Figure 1, displaying the pattern of the between over the total deviance and the corresponding increments, versus the number of clusters of the two clustering solutions.

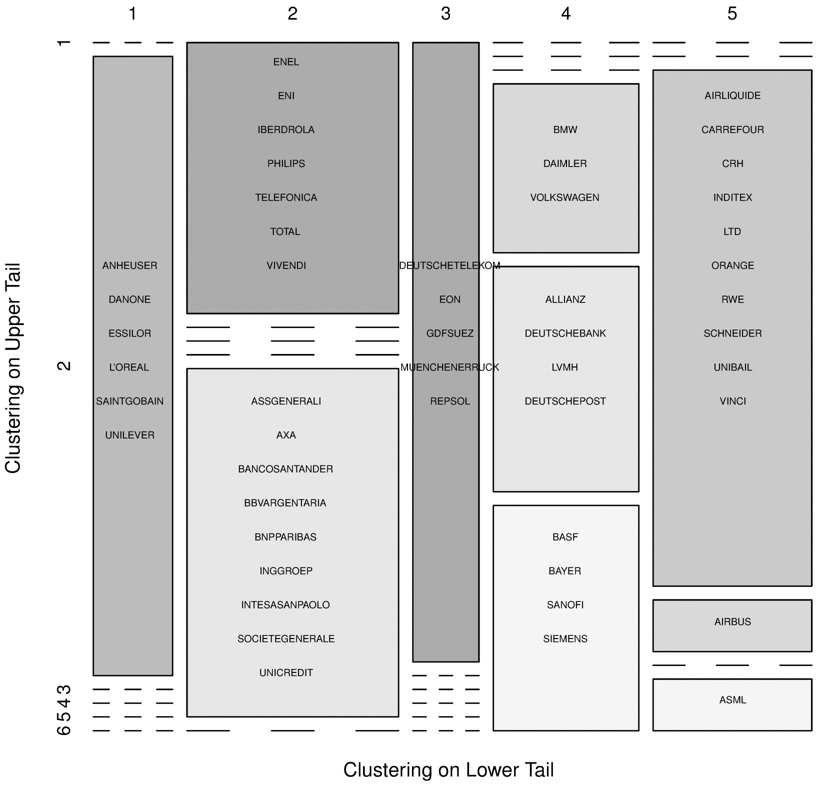

Figure 2. Joint composition of the clustering solutions $GL_1, GL_2, GL_3, GL_4, GL_5$ and $GU_1, GU_2, GU_3, GU_4, GU_5, GU_6$.

We decided to set $K_L = 5$ and $K_U = 6$. The joint composition of the two clustering solutions is graphically summarized in Figure 2, where the five *lower tail-based* and the six *upper tail-based* clusters are represented respectively by columns and rows, while the squares contain the stocks belonging to the same cluster both in the *lower tail-based* and the *upper tail-based* solution.

Step-by-step, the procedure described in Section 2.1 produced the following results:

(i) First selection, based on $GL_1, GL_2, GL_3, GL_4, GL_5$: the cluster sizes were $n_1 = 6$, $n_2 = 16$, $n_3 = 5$, $n_4 = 11$, $n_5 = 12$, so the set $\mathbf{C}^{(0)}$ of the first-step candidates is composed of $S^{(0)} = 63\,360$ possible portfolios of five stocks belonging to different clusters.

(ii) Second selection, based on $GU_1, GU_2, GU_3, GU_4, GU_5, GU_6$: the index γ_s was computed for each first-step candidate $C_s^{(0)}$; its minimum value was $\min_{s=1,2,\ldots,63360}(\gamma_s) = 0.768$ and was reached by the 2.32% of first-step candidates. So, the set $\mathbf{C}$ of second-step candidates is composed of $S = 1470$ possible portfolios, selected from $\mathbf{C}^{(0)}$ by taking into account the *upper tail-based* clustering solution. All the 1470 candidates are composed of five stocks belonging to different *lower tail-based* clusters and to three different *upper tail-based* clusters.

(iii) Final selection: the best portfolio among the 1470 second-step candidates is then chosen with financial criteria, as described in Section 4.2.

4.2 Portfolio selection

In detail, firstly the weights of the 1470 candidate portfolios are estimated by minimizing the CVaR, then the best portfolio is selected either with the minimum CVaR (portfolio C1) or the maximum Omega index criterion (portfolio C2), as described in Section 2.1. The performance of the selected portfolios has then been evaluated in an out-of-sample period, from January 1, 2014 to January 15, 2014 and compared to the two

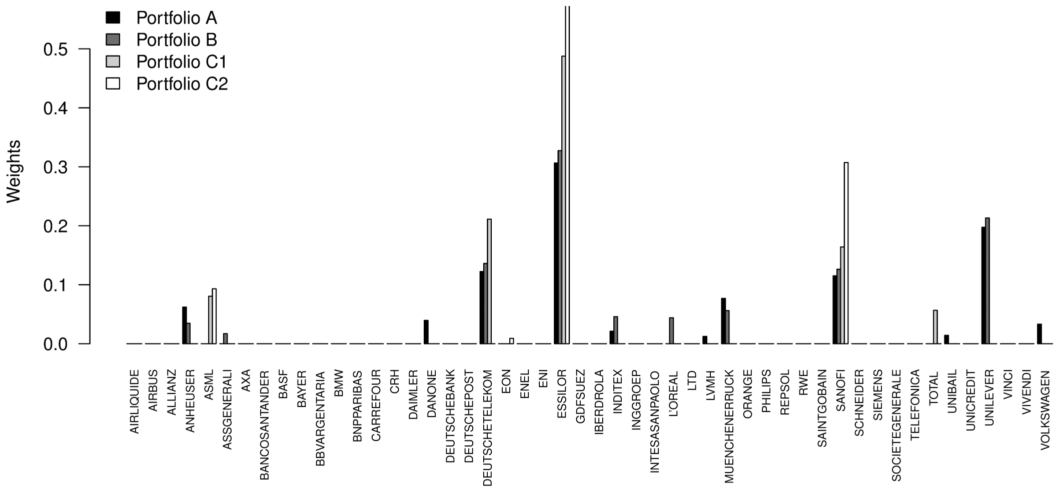

Figure 3. Weights of the stocks.

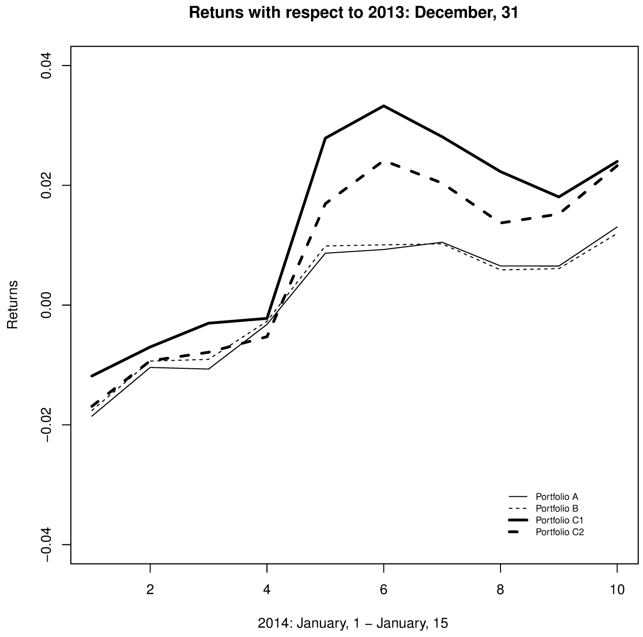

Figure 4. Returns of the portfolios.

benchmark options considered in the simulation study (A: Markowitz minimum variance portfolio built using all the stocks; B: Minimum CVaR portfolio built using all the stocks).

Figure 3 shows the composition and the weights of the four analyzed portfolios. While portfolios C1 and C2 are composed of five stocks by construction, portfolios A and B turn out to be composed of eleven and nine stocks, respectively. In the specified out-of-sample period the cumulative returns of the four portfolios (with respect to December 31, 2013) are plotted in Figure 4. Portfolios C1 and C2 largely outperform the competitors. In the first four days, all the portfolios have a loss, but the lowest loss is always recorded for portfolio C1 (0.0070) while portfolio C2 is the second best except in the fourth day when its cumulative loss is slightly higher than the competitor portfolios. However, from the fifth to the tenth day, all the portfolios have positive returns, and the superiority of portfolios built using a clustering procedure is clear. At the sixth day, the cumulative returns of portfolios C1 and C2 are, respectively, 0.033 and 0.024 against 0.009 (portfolio A) and 0.010 (portfolio B). At the end of the out-of-sample period (15 January), the cumulative returns of portfolios C1 and C2 are 0.024 and 0.023, again much higher than portfolio A (0.013) and portfolio B (0.012).

5 Concluding remarks

In this paper a portfolio selection procedure has been proposed, taking into account the behavior of stock returns in case of extreme events, both negative and positive. The innovative proposal of this paper, with respect to previous work on this theme, is the idea of considering, beyond the lower tail of the distribution, also the upper tail. Specifically, the idea is to invest on stocks exhibiting, at the same time, low and high mutual association in case of, respectively, extremely low and extremely high returns. The association in case of extreme events is measured by (lower and upper) tail dependence coefficients estimated via copula functions. The portfolio selection is based on two preliminary time series clustering procedures, aimed at grouping together stocks with high (lower and upper) tail dependence. The two clustering solutions are jointly considered in order to provide a set of candidates portfolios and the "winner" of the competition is then chosen from among these candidates, using a financial criterion such as the minimum CVaR or the maximum Omega index. The definition of the set of candidate portfolios requires to consider the heterogeneity of all the possible portfolios than can be built relying on one clusterization, with respect to the other. For this reason, the computation burden implied by the proposed procedure tends to grow rapidly as the number of considered stocks increases. Then, the method cannot realistically be applied to very large sets of stocks.

The performance of the procedure has been successfully checked on simulated data, with an experiment aimed at verifying (i) the adequateness of copula functions estimation of the tail dependence structure with a misspecified distributional assumption, (ii) the ability of the procedure in recovering the right clustering structures, and (iii) the comparison of the selected portfolios' returns to those obtained by two common portfolio selection techniques, used as benchmarks.

Finally, a case study on real data from the EUROSTOXX index shows that the portfolios selected according the proposed procedure have been able to outperform the benchmarks in a two-weeks out-of-sample period.

Funding: This research was funded by a grant from the Italian Ministry of Education, University and Research to the PRIN Project entitled "Multivariate statistical models for risks evaluation" (2010RHAHPL_005).

References

[1] K. Aas, C. Czado, A. Frigessi and H. Bakken, Pair-copula constructions of multiple dependence, *Insurance Math. Econom.* **44** (2009), 182–198.

[2] A. M. Alonso, J. R. Berrendero, A. Hernández and A. Justel, Time series clustering based on forecast densities, *Comput. Stat. Data Anal.* **51** (2006), 762–776.

[3] A. Bernardo and O. Ledoit, Gain, loss and asset pricing, *J. Political Econom.* **108** (2000), 144–172.

[4] W.-C. Chen and R. Maitra, Model-based clustering of regression time series data via APECM. An AECM algorithm sung to an even faster beat, *Stat. Anal. Data Min.* **4** (2011), 567–578.

[5] M. Corduas and D. Piccolo, Time series clustering and classification by the autoregressive metrics, *Comput. Stat. Data Anal.* **52** (2008), 1860–1872.

[6] G. De Luca and P. Zuccolotto, A tail dependence-based dissimilarity measure for financial time series clustering, *Adv. Class. Data Anal.* **5** (2011), 323–340.

[7] G. De Luca and P. Zuccolotto, Dynamic clustering of financial assets, in: *Analysis and Modeling of Complex Data in Behavioural and Social Sciences*, Springer, Berlin (2014), 103–111.

[8] G. De Luca and P. Zuccolotto, Time series clustering on lower tail dependence for portfolio selection, in: *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, Springer, Berlin (2014), 131–140.

[9] G. De Luca and P. Zuccolotto, Dynamic tail dependence clustering of financial time series, *Stat. Papers* (2015), DOI 10.1007/s00362-015-0718-7.

[10] E. Di Bernardino and D. Rullière, On tail dependence coefficients of transformed multivariate Archimedean copulas, *Fuzzy Sets and Systems* **284** (2016), 89–112.

[11] F. J. DiTraglia and J. R. Gerlach, Portfolio selection: An extreme value approach, *J. Banking Finance* **37** (2013), 305–323.

[12] F. Durante and R. Pappadà, Cluster analysis of time series via Kendall distribution, *Adv. Intell. Syst. Comput.* **315** (2015), 209–216.

[13] F. Durante, R. Pappadà and N. Torelli, Clustering of financial time series in risky scenarios, *Adv. Data Anal. Class.* **8** (2014), 359–376.

[14] F. Durante, R. Pappadà and N. Torelli, Clustering of time series via non-parametric tail dependence estimation, *Stat. Papers* **56** (2015), 701–721.

[15] P. D'Urso and E. A. Maharaj, Autocorrelation-based fuzzy clustering of time series, *Fuzzy Sets and Systems* **160** (2009), 3565–3589.

[16] J. Franke, W. Härdle and C. Hafner, *Statistics of Financial Markets: An Introduction*, 4th ed., Springer, Berlin, 2015.

[17] W. Härdle, N. Hautsch and L. Overbeck, *Applied Quantitative Finance*, 2nd ed., Springer, Berlin, 2009.

[18] H. Joe, *Multivariate Models and Dependence Concept*, Chapman & Hall, New York, 1997.

[19] C. Keating and W. Shadwick, A universal performance measure, *J. Perform. Measur.* **6** (2002), 59–84.

[20] P. Krokhmal, J. Palmquist and S. Uryasev, Portfolio optimization with conditional value-at-risk objective and constraints, *J. Risk* **4** (2002), 43–68.

[21] H. Markowitz, Portfolio selection, *J. Finance* **7** (1952), 77–91.

[22] E. Otranto, Clustering heteroskedastic time series by model-based procedures, *Comput. Stat. Data Anal.* **52** (2008), 4685–4698.

[23] F. Pattarin, S. Paterlini and T. Minerva, Clustering financial time series: An application to mutual funds style analysis, *Comput. Stat. Data Anal.* **47** (2004), 353–372.

[24] R. T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, *J. Risk* **2** (2000), 21–41.

[25] J. A. Vilar, A. M. Alonso and J. M. Vilar, Non-linear time series clustering based on non-parametric forecast densities, *Comput. Stat. Data Anal.* **54** (2010), 2850–2865.