Time Series Clustering on Lower Tail Dependence for Portfolio Selection

Chapter · January 2014

DOI: 10.1007/978-3-319-02499-8-12

2 authors:

Paola Zuccolotto
Università degli Studi di Brescia

Giovanni De Luca
Parthenope University of Naples

Time series clustering on lower tail dependence for portfolio selection

Giovanni De Luca and Paola Zuccolotto

Abstract In this paper we analyse a case study based on the procedure introduced by De Luca and Zuccolotto (2011), whose aim is to cluster time series of financial returns in groups being homogeneous in the sense that their joint bivariate distributions exhibit high association in the lower tail. The dissimilarity measure used for such clustering is based on tail dependence coefficients estimated using copula functions. We carry out the clustering using an algorithm requiring a preliminary transformation of the dissimilarity index into a distance metric by means of a geometric representation of the time series, obtained with Multidimensional Scaling. We show that the results of the clustering can be used for a portfolio selection purpose, when the goal is to protect investments from the effects of a financial crisis.

Key words: Time series clustering, tail dependence, copula function, portfolio selection.

1 Introduction

Several approaches to time series clustering are present in the literature. After the first studies, where dissimilarities between time series were merely derived by the comparison between observations or some simple statistics computed on the data (see for example Bohte *et al.*, 1980), more complex solutions have been proposed. An interesting review can be found in Warren Liao (2005). Piccolo (1990) and Corduas and Piccolo (2008) proposed a distance measure for time series generated by ARIMA processes, based on the comparison between the parameters of the corre-

Giovanni De Luca
University of Naples Parthenope, e-mail: giovanni.deluca@uniparthenope.it

Paola Zuccolotto
University of Brescia, e-mail: paola.zuccolotto@sis-statistica.org

sponding Wold decomposition. Otranto (2008) extended this approach to GARCH models, Galeano and Peña (2000) considered the generalized distance between the autocorrelation functions of two time series while Caiado *et al.* (2006) introduced a metric based on the normalized periodogram. Alternative methods, employing parametric and non-parametric density forecasts, was discussed in Alonso *et al.* (2006) and in Vilar *et al.* (2010), respectively. To give an idea of the great variety of approaches in this framework, it is finally worth recalling the frequency domain approach of Kakizawa *et al.* (1998) and Taniguchi and Kakizawa (2000), the use of two-dimensional singular value decomposition by Weng and Shen (2008), the procedure using a robust evolutionary algorithm of Pattarin *et al.* (2004). A comparison of sevaral parametric and non-parametric approaches can be found in Pertega Diaz and Vilar (2010). But the list of citations could be even longer. In this paper we show a case study based on the use of the procedure proposed by De Luca and Zuccolotto (2011) to cluster time series of returns of financial assets according to their association in the lower tail. Then, we show how this approach can be employed for portfolio selection, especially in a financial crisis perspective. With respect to the seminal paper of De Luca and Zuccolotto (2011), here we propose, firstly, two indexes for evaluating the quality of the clusterization from the point of view of tail dependence, and, secondly, a more specific criterion for portfolio selection, based on the Omega index proposed by Keating and Shadwick (2002). The paper is organized as follows: in Section 2 the clustering procedure is briefly recalled, while the main results of the case study are summarized in Section 3. Concluding remarks follow in Section 4.

2 Tail dependence based clustering procedure

The interest of researchers in modelling the occurring of extreme events has several empirical motivations, especially in contexts where it can be directly associated to risk measurement, such as, for example, financial markets. Recently, a great deal of attention has been devoted also to the study of association between extreme values of two or more variables. From a methodological point of view, the problem of quantifying this association has been addressed in different ways. One of the proposed approaches consists in analyzing the probability that one variable assumes an extreme value, given that an extreme value has occurred to the other variables (see Cherubini *et al.* 2004). This probability is known as lower or upper tail dependence and we will restrict its analysis to the bivariate case. Let Y_1 and Y_2 be two random variables and let $U_1 = F_1(Y_1)$ and $U_2 = F_2(Y_2)$ be their distribution functions. The lower and upper tail dependence coefficients are defined respectively as

$$\lambda_L = \lim_{v \to 0^+} P\left[U_1 \leq v | U_2 \leq v\right] \quad \text{and} \quad \lambda_U = \lim_{v \to 1^-} P\left[U_1 > v | U_2 > v\right].$$

In practice, the tail dependence coefficients are estimated from observed data after assuming a probabilistic framework. The copula functions are commonly used for financial data with the advantage that the tail dependence estimation is both simple and flexible. A two-dimensional copula function for two random variables Y_1 and Y_2 is defined as a function $C : [0,1]^2 \to [0,1]$ such that

$$F(y_1, y_2; \theta) = C(F_1(y_1; \vartheta_1), F_2(y_2; \vartheta_2); \tau),$$

for all y_1, y_2, where $F(y_1, y_2; \theta)$ is the joint distribution function of Y_1 and Y_2 (see Nelsen 2006) and $\theta = (\vartheta_1, \vartheta_2, \tau)$. It is straightforward to show that the tail dependence coefficients can be expressed in terms of the copula function. In particular, the lower tail dependence coefficient, which will be hereafter the focus of the paper, is given by

$$\lambda_L = \lim_{v \to 0^+} \frac{C(v,v)}{v}.$$

In the analysis of the relationship between financial returns, the lower tail dependence coefficient gives an idea of the risk of investing on assets for which extremely negative returns could occur simultaneously. So, the lower tail dependence is strictly linked to the diversification of investments, especially in financial crisis periods. For this reason, De Luca and Zuccolotto (2011) proposed to cluster time series of financial returns according to a dissimilarity measure defined as

$$\delta(\{y_{it}\}, \{y_{jt}\}) = -\log(\hat{\lambda}_L),$$

where $\{y_{it}\}_{t=1,\ldots,T}$ and $\{y_{jt}\}_{t=1,\ldots,T}$ denote the time series of returns of two assets i and j, and $\hat{\lambda}_L$ is their estimated tail dependence coefficient. In this way we obtain clusters of assets characterized by high tail dependence in the lower tail. From a portfolio selection perspective, it should then be avoided portfolios containing assets belonging to the same cluster. Given p assets, the clustering procedure proposed by De Luca and Zuccolotto (2011) is composed by two steps. In the first step, starting from the dissimilarity matrix $\Delta = (\delta_{ij})_{i,j=1,\ldots,p}$, an *optimal* representation of the p time series $\{y_{1t}\}, \ldots, \{y_{pt}\}$ as p points $\mathbf{y}_1, \ldots, \mathbf{y}_p$ in $\mathbb{R}^q$ is found by means of Multidimensional Scaling (MDS). The above mentioned term *optimal* means that, with MDS, the Euclidean distance matrix $D = (d_{ij})_{i,j=1,\ldots,p}$, with $d_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\|$, of the points $\mathbf{y}_1, \ldots, \mathbf{y}_p$ in $\mathbb{R}^q$ has to fit as closely as possible the dissimilarity matrix Δ. The extent to which the interpoint distances d_{ij} "match" the dissimilarities δ_{ij} is measured by an index called *stress*, which should be as low as possible. 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 the second step, the k-means clustering algorithm is performed using the obtained geometric representation of the p time series. Among the several hierarchical and non-hierarchical clustering techniques which we may resort to, the k-means algorithm performed on the MDS geometrical representation has revealed, through simulation studies, a good performance in this context (De Luca and Zuccolotto, 2011).

3 Case study

In this case study we analyse the time series of the daily prices of the 24 stocks which have been included in FTSE MIB index during the whole period from January 3, 2006 to October 31, 2011.

3.1 Clustering

After transforming prices into log-returns, we preliminary removed autocorrelation and heteroskedasticity by means of univariate Student-*t* AR-GARCH models. For each couple of stocks we estimated a bivariate Joe-Clayton copula function (Joe, 1997),

$$C(u_1,u_2) = 1 - \{1 - [(1-(1-u_1)^{\kappa})^{-\theta} + (1-(1-u_2)^{\kappa})^{-\theta} - 1]^{-1/\theta}\}^{1/\kappa}$$

using the estimated distribution functions of the standardized residuals. After estimating the 276 lower tail dependence coefficients, which in the case of the Joe-Clayton copula are given by $\hat{\lambda}_L = 2^{-1/\hat{\theta}}$, we carried out MDS using the dissimilarity matrix $\Delta = (\delta_{ij})_{i,j=1,\dots,24}$. We set $\bar{s} = 0.005$. The minimum dimension allowing a final configuration with minimum stress lower than $\bar{s}$ resulted $q = 14$. In the second step, we performed a k-means clustering algorithm using the MDS point configuration in $\mathbb{R}^{14}$, $\mathbf{y}_1, \dots \mathbf{y}_{24}$. The graph displayed in the left panel of Figure 1 shows the pattern of the ratio of deviance between clusters over total deviance, as a function of the number of clusters k and its increments when considering the solution with k clusters, with respect to $k-1$ clusters. This graph helps the researcher in deciding the optimal number of clusters. In this case we observe that moving from $k = 3$ to $k = 4$ allows an improvement of 32.7% in the quality of the clusterization, while from $k = 5$ onward, the increments are appreciably lower and more stable. For this reason, we feel that a good choice could be $k = 4$. In addition we have computed the following indexes proposed in the literature for determining the optimal number of clusters (see Dimitriadou *et al.*, 2002 for an exhaustive review): CH (Calinski and Harabasz, 1974), H (Hartigan, 1975), RL (Ratkowsky and Lance, 1978), SS (Scott and Symons, 1971), M (Marriot, 1971), BB (Ball and Hall, 1965), *TraceCovW* (Milligan and Cooper, 1985), *TraceW* (Edwards and Cavalli-Sforza, 1965; Friedman and Rubin, 1967) $TraceW^{-1}B$ and $|T| = |W|$ (Friedman and Rubin, 1967). All the indexes except CH suggest to choose the solution with $k = 4$ clusters. So, we judge this choice the most reliably founded as it combines subjective remarks and objective criteria. The cluster composition by stocks and by economic activity sectors are displayed in Table 1 and Table 2, respectively. We observe that there are some affinities in the sectors of stocks belonging to the same cluster, so we propose to label the clusters as shown in Table 2.

In order to measure the extent to which the clusterization has been able to group stocks with high tail dependence, separating them from the others, we propose to

Table 1 Cluster composition by stocks.

Cluster 1	Cluster 2		Cluster 3	Cluster 4	
Atlantia	Autogrill	Fiat	SNAM	Banca MPS	Fondiaria
ENEL	Finmeccanica	Lottomatica	Terna	Generali	Intesa SP
ENI	Luxottica	Pirelli		Mediobanca	Mediolanum
Saipem	Stmicroelectronics	Telecom		Mediaset	Banca PM
				UBI	Unicredit

Table 2 Cluster composition by sectors (with cluster labels).

Cluster 1 **Power, energy and mobility**	**Cluster 2** **Living goods**
Oil and natural gas (2) Industrial services and products (1) Public services (1)	Travels and free time (2) Cars and components (2) Tecnology and Communications (2) House, personal utilities, fashion (1) Industrial services and products (1)
Cluster 3 **Public services**	**Cluster 4** **Banks and insurance**
Public services (2)	Banks (6) Insurance (3) Media (1)

compute two indexes, which we call average tail dependence coefficient within and between clusters, respectively. On the one hand, the average tail dependence coefficient within cluster c is given by

$$\bar{\lambda}_c^W = \frac{2}{n_c(n_c-1)} \sum_{i_c=1}^{n_c} \sum_{j_c=i_c+1}^{n_c} \hat{\lambda}_L^{i_c j_c},$$

where n_c is the number of stocks belonging to cluster c, i_c and j_c are the ith and the jth stocks of cluster c and $\hat{\lambda}_L^{i_c j_c}$ is their estimated tail dependence coefficient. The index $\bar{\lambda}_c^W$ measures the extent to which the clusters are internally homogeneous from the point of view of tail dependence and should be as high as possible. On the other hand, the average tail dependence coefficient between cluster c and the others is given by

$$\bar{\lambda}_c^B = \frac{1}{n_c n_{\bar{c}}} \sum_{i_c=1}^{n_c} \sum_{i_{\bar{c}}=1}^{n_{\bar{c}}} \hat{\lambda}_L^{i_c i_{\bar{c}}},$$

where $n_{\bar{c}}$ is the number of stocks not belonging to cluster c, $i_{\bar{c}}$ is the ith stock outside cluster c and $\hat{\lambda}_L^{i_c i_{\bar{c}}}$ is the estimated tail dependence coefficient if stocks i_c and $i_{\bar{c}}$. This index measures the extent to which the clusters are externally separated from the point of view of tail dependence and should be as low as possible. In general, a good clusterization should have $\bar{\lambda}_c^W > \bar{\lambda}_c^B$ for all $c = 1, \cdots, k$. Results for this case study are displayed in Table 3.

Table 3 Average tail dependence within and between clusters

Cluster (c)	**1**	**2**	**3**	**4**
$\bar{\lambda}_c^W$	0.2770	0.2719	0.2975	0.4249
$\bar{\lambda}_c^B$	0.2224	0.2425	0.1211	0.2475

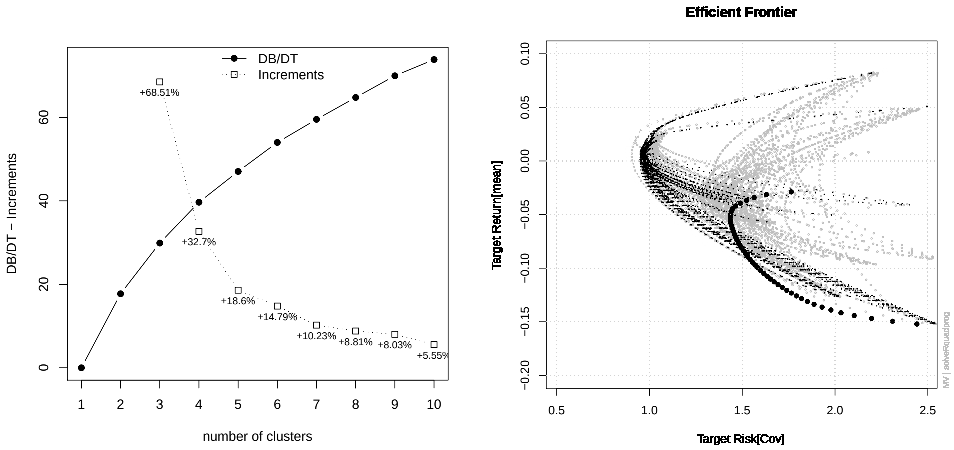

Fig. 1 *Left*: Deviance between clusters over total deviance. *Right*: Efficient frontiers of the 640 selections based on clustering (black, thin), efficient frontiers of 1000 selections built with 4 randomly selected stocks (gray) and efficient frontier of the stocks belonging to cluster 4 (black, bold).

3.2 Portfolio selection

We used the obtained clustering to construct portfolios composed by as many stocks as the number k of clusters. The stocks are selected by imposing the restriction that each stock belongs to a different cluster (De Luca and Zuccolotto, 2011); with the $k = 4$ above mentioned clusters, 640 different selections can be made according to this criterion. This strategy should protect the investments from parallel extreme losses during crisis periods, because the clustering solution is characterized by a moderate lower tail dependence between clusters. Using the popular Markowitz portfolio selection procedure, in the right panel of Figure 1, we plotted the efficient frontiers of all the possible 640 selections (black, thin), compared to those of 1000 portfolios (gray) built with 4 randomly selected stocks and to the efficient frontier of the portfolio built using all the stocks belonging to cluster 4 (black, bold), the cluster with the highest average tail dependence within cluster $\bar{\lambda}_c^W$. The efficient frontiers of the 640 selections dominate the main part of the others. After selecting the minimum variance portfolio of each frontier, we evaluated the performance of the 640 resulting portfolios in the following 36 days, thus with an out-of-sample perspective. Figure 2 displays the returns of the 640 portfolios (black) in the period

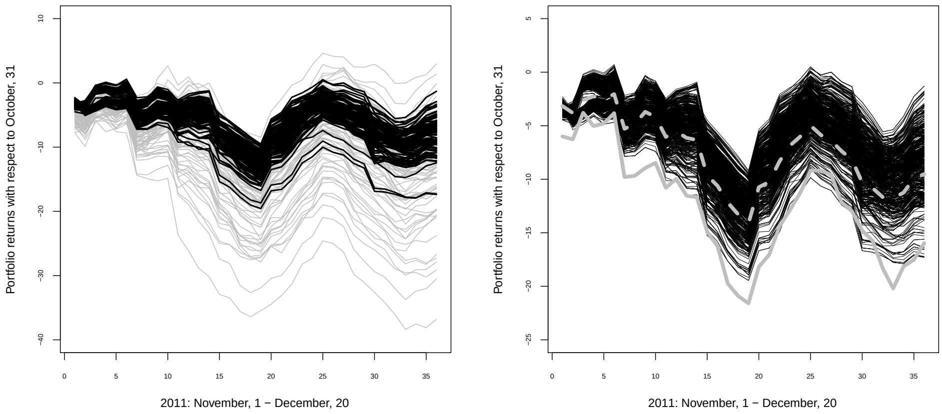

Fig. 2 *Left*: Returns of the 640 minimum variance portfolios based on clustering (black), returns of the minimum variance portfolios built using 4 randomly selected stocks (gray). *Right*: Returns of the 640 minimum variance portfolios based on clustering (black), returns of the minimum variance portfolios built using all the stocks (gray, dashed line) and using the stocks of cluster 4 (bold gray)

November, 1 - December 20, with respect to the price of October, 31. These returns are compared to (*i*) the returns of the minimum variance portfolios built using 4 randomly selected stocks (left panel, gray), (*ii*) the returns of the minimum variance portfolio built using all the stocks (right panel, bold gray, dashed line) and to (*iii*) the returns of the minimum variance portfolio built using all the stocks belonging to cluster 4 (right panel, bold gray). We observe a good performance of the 640 portfolios with respect to the selected competitors. At this step, a criterion should be given in order to select one of these 640 portfolios. To this purpose, we propose to resort to a proper index, such as for example the Sharpe Ratio (Sharpe, 1966) or the Omega index (Keating and Shadwick, 2002). In order to take into account the whole returns distribution, we decide to use the Omega index, given by

$$\Omega(L) = \frac{\displaystyle\int_L^b (1 - F(r))dr}{\displaystyle\int_a^L F(r)dr},$$

where $F(r)$ denotes the cumulative probability distributions of the portfolio returns, $(a,b) \in \mathbb{R}$ their domain, L a reference return, often set equal to 0. As a first approach, we select, among the 640 minimum variance portfolios, the one with the highest value of Ω. Alternatively, instead of restricting the choice to the minimum variance portfolios, we may explore all the portfolios lying on the 640 efficient frontiers plotted in the right panel of Figure 1.

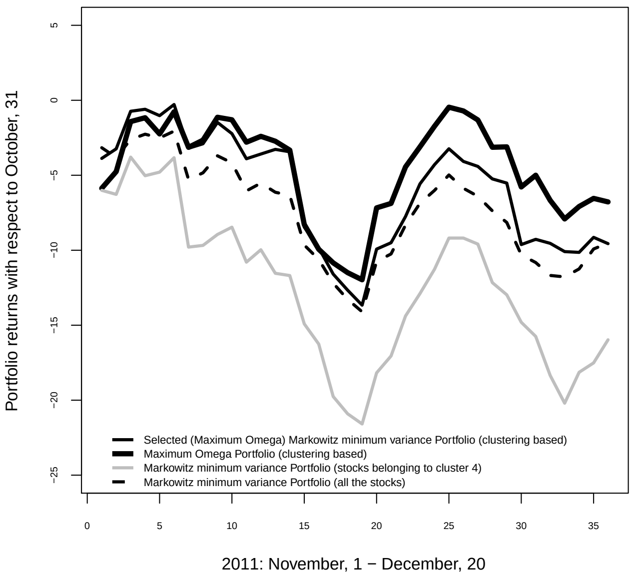

Fig. 3 Returns of the selected portfolios.

Figure 3 displays the returns of the two portfolios corresponding to these two approaches, compared to the competitors described above. The composition of the portfolios in Figure 3 is given in Table 4. Both the portfolios tend to outperform the others, the one exploring the whole efficient frontier being the best one in the long period. The results obtained with the Sharpe Ratio are similar.

4 Concluding remarks

In this paper, we have clustered 24 stocks included in FTSE MIB index according to the association among extremely low returns. For each couple of stocks we have estimated a bivariate copula function and computed the lower tail dependence coefficient. Then, following a two-step clustering procedure integrating the use of Multi Dimensional Scaling and the *k*-means clustering algorithm, we have formed four groups of stocks. The obtained clustering has been used to build portfolios according to the criterion of selecting one stock from each cluster. We have shown that in an out-of-sample period characterized by financial crisis, the returns of some of these portfolios are less unfavourable with respect to the return of the minimum variance portfolio built using all the stocks or to the returns of portfolios built with

Table 4 Composition of the portfolios in Figure 3

Selected (Maximum Ω) Markowitz minimum variance Portfolio (clustering based)			
Saipem	Fiat	Terna	Intesa
0.0390	0.0290	0.9062	0.0258

Maximum Ω Portfolio (clustering based)			
Saipem	Fiat	Terna	UBI
0.5384	0	0	0.4616

Markowitz minimum variance Portfolio (stocks belonging to cluster 4)				
Banca Mps	Fondiaria	Generali	Intesa	Mediobanca
0	0	0.3222	0	0.2663
Mediolanum	Mediaset	Banca PM	UBI	Unicredit
0	0.3246	0	0.0869	0

Markowitz minimum variance Portfolio (all the stocks)					
Atlantia	Autogrill	Banca Mps	ENEL	ENI	Fiat
0.0438	0.0399	0	0	0	0
Finmeccanica	Fondiaria	Generali	Intesa	Lottomatica	Luxottica
0.0400	0	0	0	0.0374	0.0113
Mediobanca	Mediolanum	Mediaset	Pirelli	Banca PM	Saipem
0.0765	0	0	0	0	0
SNAM	Stm	Telecom	Terna	UBI	Unicredit
0.4456	0	0 0.3055	0	0	

randomly selected stocks. Moreover, portfolios composed of stocks belonging to the same cluster can exhibit a very bad performance. Finally, we have proposed a criterion for selecting one portfolio out of the hundreds of possible choices deriving from the rule of taking one stock from each cluster. The procedure has revealed effective on the data of the described case study. Future research could be developed in several directions. For example, the clustering procedure could be adapted to a time varying perspective, with the result of a dynamic clustering and portfolio selection. In addition, some effort should be devoted to refine the portfolio selection procedure in order to more specifically focus on the tails of the returns distribution, coherently with the proposed tail dependence clustering.

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

References

1. Alonso, A.M., Berrendero J.R. Hernández A., & Justel, A.: Time series clustering based on forecast densities. Computational Statistics and Data Analysis. **51**, 762—776 (2006)

2. Ball, G.H., & Hall, D.J.: ISODATA, A novel method of data analysis and pattern classification. Tech. Rep. NTIS No.AD 699616, Stanford Research Institute (1965).
3. Bohte, Z., Cepar, D., & Kosmelj, K.: Clustering of time series. In: Barritt, M.M., & Wishart, D. (eds) COMPSTAT 1980, Proceedings in Computational statistics. Wien: Physica-Verlag, 587–593 (1980)
4. Caiado, J., Crato N., & Peña, D.: A periodogram-based metric for time series classification. Computational Statistics and Data Analysis. **50**, 2668–2684 (2006)
5. Calinski, R.B., & Harabasz, J.: A dendrite method for cluster analysis. Communications in Statistics, **3**, 1–27 (1974)
6. Cherubini, U., Luciano E., & Vecchiato, W.: Copula methods in finance. New York: Wiley (2004)
7. Corduas, M., & Piccolo, D.: Time series clustering and classification by the autoregressive metrics. Computational Statistics and Data Analysis. **52**, 1860–1872 (2008)
8. De Luca, G., & Zuccolotto, P.: A tail dependence-based dissimilarity measure for financial time series clustering. Advances in Classification and Data Analysis, **5**, 323–340 (2011)
9. Dimitriadou, E., Dolničar, S., & Weingessel A.: An examination of indexes for determining the number of clusters in binary data sets. Psychometrika, **67**, 137–159 (2002)
10. Edwards, A.W.F., & Cavalli-Sforza, L.: A method for cluster analysis. Biometrics, **21**, 362–375 (1965)
11. Friedman, H.P., & Rubin, J.: On some invariant criteria for grouping data. Journal of the American Statistical Association, **62**, 1159–1178 (1967)
12. Galeano, P., & Peña, D.: Multivariate analysis in vector time series. Resenhas, **4**, 383–404 (2000)
13. Hartigan, J.A.: Clustering algorithms. New York: Willey (1975)
14. Joe, H.: Multivariate models and dependence concepts. New York: Chapman & Hall/CRC (1997)
15. Kakizawa, Y., Shumway, R.H., & Taniguchi, M.: Discrimination and clustering for multivariate time series. Journal of the American Statistical Association, **93**, 328–340 (1998)
16. Keating, C., & Shadwick, W. F.: A universal performance measure. Journal of performance measurement, **6**, 59–84 (2002).
17. Marriot, F.H.C.: Practical problems in a method of cluster analysis. Biometrics, **27**, 501–514 (1971)
18. Milligan, G.W., & Cooper, M.C.: An examination of procedures for determining the number of clusters in a data set. Psychometrika, **50**, 159–179 (1985)
19. Nelsen, R.: An introduction to copulas. New York: Springer (2006)
20. Otranto, E.: Clustering heteroskedastic time series by model-based procedures. Computational Statistics and Data Analysis, **52**, 4685–4698 (2008)
21. Pattarin, F., Paterlini, S., & Minerva, T.: Clustering financial time series: an application to mutual funds style analysis. Computational Statistics and Data Analysis, **47**, 353–372 (2004)
22. Pertega Diaz, S. & Vilar J.A.: Comparing Several Parametric and Nonparametric Approaches to time series clustering: a simulation study. Journal of Classification, **27**, 333–362 (2010)
23. Piccolo, D.: A distance measure for classifying ARMA models. Journal of Time Series analysis, **11**, 153–164 (1990)
24. Ratkowsky, D. A., & Lance, G.N.: A criterion for determining the number of groups in a classification. Australian Computer Journal, **10**, 115–117 (1978).
25. Scott, A.J. & Symons, M.J.: Clustering methods based on likelihood ratio criteria. Biometrics, **27**, 387–397 (1971)
26. Sharpe, W. F.: Mutual Fund Performance. Journal of Business, **39**, 119–138 (1966).
27. Taniguchi, M., & Kakizawa, Y.: Asymptotic theory of statistical inference for time series. New York: Springer (2000)
28. Vilar, J.A., Alonso A.M., & Vilar, J.M.: Non-linear time series clustering based on nonparametric forecast densities. Computational Statistics and Data Analysis, **54**, 2850–2865 (2010)
29. Warren Liao, T.: Clustering of time series a survey. Pattern recognition, **38**,1857–1874 (2005)
30. Weng, X., & Shen, J.: Classification of multivariate time series using twodimensional singular value decomposition. Knowledge-Based Systems, **21**, 535–539 (2008)