Robust fuzzy clustering of time series based on B-splines

Pierpaolo D'Urso[a,*], Luis A. García-Escudero[c], Vincenzina Vitale[1], Livia De Giovanni[b], Agustín Mayo-Iscar[c]

[a] *Dipartimento di Scienze Sociali ed Economiche, Sapienza University of Rome, P.za Aldo Moro, 5 - 00185 Rome, Italy.*

[b] *Dipartimento di Scienze Politiche, LUISS university, Viale Romania, 32 - 00197 Rome, Italy.*

[c] *Departamento de Estadística e Investigación Opertiva, Universidad de Valladolid, 47011 Valladolid, Spain.*

Abstract

Four different approaches to robust fuzzy clustering of time series are presented and compared with respect to other existent approaches. These approaches are useful to cluster time series when outlying values are found in these time series, which is often the rule in most real data applications. A representation of the time series by using B-splines is considered and, later, robust fuzzy clustering methods are applied on the B-splines fitted coefficients. Feasible algorithms for implementing these methodologies are presented. A simulation study shows how these methods are useful to deal with contaminating time series and also switching time series due to fuzziness. A real data analysis example on financial data is also presented.

Keywords: Clustering; Fuzzy clustering; Robustness; Trimming; Noise Cluster

1. Introduction

Cluster analysis is the art of finding groups in data. If our data set is made of several time series, it makes sense to find subgroups in it in order to better extract useful information and summarize this data set. Consequently, time series clustering problem is receiving increasing attention. [1], [2], [3] and [4] are some revisions works and [5] is also a recent book exclusively devoted to this topic. Several applications for time series clustering in many different fields can be found in these references. A time series clustering package in R, with a wide choice of methods, is also provided in [6].

*Corresponding author

Email addresses: `pierpaolo.durso@uniroma1.it` (Pierpaolo D'Urso), `lagarcia@eio.uva.es` (Luis A. García-Escudero), `vincenzina.vitale@uniroma1.it` (Vincenzina Vitale), `ldegiovanni@luiss.it` (Livia De Giovanni), `agustinm@eio.uva.es` (Agustín Mayo-Iscar)

It is known the very detrimental effect that (even few) outlying measurements may have in many widely applied clustering techniques. Therefore, it is also increasingly common to apply robust clustering procedures which are specifically designed to resist well these outlying measurements [7–10]. Obviously, this is going to be the case when analyzing data sets made of time series, where it is quite likely that wrong or anomalous values could easily appear within the considered time series. The use of robust clustering techniques is also useful because they allow to also detect outlying values in the observed data set. These outlying values can be very interesting if one is able to explain the scientific reasons for these observed data departures. Some robust procedures applied in the clustering of time series can be found in [5].

An alternative approach is to view isolated or small groups of outlying time series as "clusters on their own". This idea makes sense because the outlying time series are clearly "separated" from other (main) clusters. However, this alternative approach is not always satisfactory because many times the researcher imposes in advance the total number C of clusters that he/she is interested in detecting but can be unaware of the existence of outlying time series (that can significantly alter the final clustering results). Also, a certain number of disperse anomalous time series would require a very large number C of clusters, to "accommodate" all of them properly, which can be very complex and tedious to interpret.

In this work, we propose considering a representation of each time series by a set of coefficients resulting from a least-squares regression when using the B-spline basis evaluations as the explanatory variables values. That approach, that will be detailed in Section 2, has been already applied for clustering functional data sets in [11] and [12]. Of course, other functional bases than B-splines can be analogously considered (as, for instance, Fourier or wavelets basis) if they can provide better or more informative approximations to the analyzed time series. Later, our proposal is to apply robust fuzzy clustering techniques to the set of fitted coefficients, and simply use the obtained clusters to define the final clusters for our set of time series.

Four robust fuzzy clustering techniques are considered in this work, that follow from the application of impartial trimming methodologies and the "noise cluster" approach. They are applied to both C-means and C-medoids type procedures.

The impartial trimming method was introduced in the crisp clustering framework in [13]. It allows that a fraction α of the (hopefully) most outlying observations are being trimmed or discarded. The term "impartial" means that the dataset itself tells us which units appears to be the "most outlying" ones, without the user specification of privileged trimming directions, and just taking into account the distances to some optimally determined cluster centers. On the other hand, the "noise cluster" approach creates a fictitious additional cluster that collects "noisy" observations that are equally remote to the main cluster centers.

The use of these two approaches require the specification of the number of clusters and tuning parameters that control the fraction of observations that are left unassigned (trimmed or belonging to the "noise cluster"). In this work, we will review and adapt some already existent tools with the aim of helping the user in making sensible specifications of these unknown parameters.

A simulation study is carried out by comparing these four approaches with other robust crisp clustering approaches and with other robust fuzzy clustering approaches, for instance, including Dynamic Time Warping. The simulation scenarios include both contaminating and switching time series. The simulation study shows how these robust fuzzy clustering proposals outperform other robust non-fuzzy ones in scenarios with switching time series.

A real data example is also included by analyzing the performance over time of stocks composing the FTSE-MIB index. We will see how the proposed methodologies do a good job in detecting the main underlying cluster structures and also in detecting outlying time series.

The outline of the work is as follows. Section 2 will present the proposed methodologies, together with feasible algorithms for their practical implementation. The simulation study will be presented in Section 3 and a real data application in Section 4. Finally, Section 5 will provide some concluding remarks and some further research lines.

2. Fuzzy clustering of time series based on B-splines

2.1. Time series: a B-spline representation

Let us consider a p-dimensional functional basis $\{B_s(\cdot)\}_{s=1}^p$. Given a single data time series $\{(t_i^j, x_i(t_i^j))\}_{j=1}^{Ji}$ (result of recording time series x_i at the J_i time moments $a < t_i^1 < t_i^2 < \ldots < t_i^{Ji} < b$), we can model that time series by a simple linear least-squares fit by searching some summarizing coefficients $\boldsymbol{\beta}_i = (\beta_i^1, \ldots, \beta_i^p)' \in \mathbb{R}^p$ such that

$$(\beta_i^1, \ldots, \beta_i^p)' = \arg \min_{\beta^1,\ldots,\beta^p} \sum_{j=1}^{J_i} \left(x_i(t_i^j) - \sum_{s=1}^{p} \beta^s B_s(t_i^j) \right)^2. \tag{1}$$

If $\boldsymbol{x}_i = (x_i(t_i^1), ..., x_i(t_i^{J_i}))'$ and $\boldsymbol{B}_i$ is the $J_i \times p$ matrix with terms $B_s(t_i^j)$ ($j = 1, ..., J_i$ and $s = 1, ..., p$) then $\boldsymbol{\beta}_i = (\boldsymbol{B}_i'\boldsymbol{B}_i)^{-1}\boldsymbol{B}_i'\boldsymbol{x}_i$. Thus, for n time series $\{\boldsymbol{x}_i\}_{i=1}^n$, we obtain n vectors of fitted coefficients $\boldsymbol{\beta}_i = (\beta_i^1, \ldots, \beta_i^p)'$, $i = 1, 2, \cdots, n$.

Throughout this paper, we will always consider the cubic B-spline bases for exemplifying the proposed methodology. The cubic B-spline basis approach requires considering S interior knots $a < \xi_1 < \xi_2 < \ldots < \xi_S < b$ and those interior knots remain fixed for all regressions (1) performed, $i = 1, \ldots, n$, and S is typically notably smaller than all the J_i. That representation serves to approximate each time series by a continuous function which is a cubic polynomial over (ξ_l, ξ_{l+1}) and has continuous first and second derivatives at the knots and is expected to capture smooth changes in the central tendency in the considered time series. For a fixed number of interior knots S, the cubic B-spline basis is made up of $p = S+4$ basis elements. Moreover, it allows us to easily deal with time series observed in the same period of time $[a, b]$ but at different time moments/frequencies or with missing measurements.

Other functional bases could be applied analogously, if these functional bases allows us to better grasp the underlying time series structures. Note that the computation of these $\boldsymbol{\beta}_i$ coefficients is greatly simplified when considering orthonormal functional bases since $\boldsymbol{B}_i'\boldsymbol{B}_i$ reduces to the identity matrix. For instance, Fourier bases can be used when dealing when time series of a periodic nature.

2.2. BS-based Trimmed Fuzzy C-Means Clustering method (BS-Tr-FCMC)

The BS-based Trimmed Fuzzy C-Means Clustering model (BS-Tr-FCMC model) provides a robust fuzzy clustering methodology using centroids for summarizing the features of the respective clusters. This approach is the direct extension of the fuzzy C-means method [14] when trimming a pre-specified proportion α of observations as in [13] is allowed. The C centroids will be denoted by $\hat{\boldsymbol{\beta}}_1$,...,$\hat{\boldsymbol{\beta}}_C$ with $\hat{\boldsymbol{\beta}}_c = (\hat{\beta}_c^1, ..., \hat{\beta}_c^p)$ for $c = 1, 2, ..., C$. Each centroid $\hat{\boldsymbol{\beta}}_c$ represents "synthetically" each cluster as a weighted average of a set of features observed on the objects assigned to this cluster.

Given a trimming size $\alpha \in [0, 1)$, the method searches for a subset $\mathcal{I}$ of indexes $\mathcal{I} \subset \{1, ..., n\}$ with $\#\mathcal{I} = \lfloor n(1-\alpha) \rfloor$ such that $\boldsymbol{\beta}_i \in \mathcal{I}$ are the coefficients corresponding to the time series that will be declared as "non-outlying" ones ($\lfloor x \rfloor$ denotes the greater integer less than or equal to x). On the other hand, coefficients $\boldsymbol{\beta}_i \notin \mathcal{I}$ will correspond to the $n - \lfloor n(1-\alpha) \rfloor$ time series that are considered the "most outlying" ones and to be trimmed or removed.

With this purpose in mind, it is proposed to solve the following minimization problem:

$$\begin{cases} f_{C,\alpha,m}(\{u_{ic}\},\mathcal{I},\{\hat{\boldsymbol{\beta}}_c\}) := \min\limits_{\{u_{ic}\}} \min\limits_{\substack{\mathcal{I}\subset\{1,\dots,n\}\\ \#\mathcal{I}=\lfloor n\cdot(1-\alpha)\rfloor}} \sum\limits_{i\in\mathcal{I}}\sum\limits_{c=1}^{C} u_{ic}^m \sum\limits_{s=1}^{p}(\beta_i^s-\hat{\beta}_c^s)^2 \\ \text{s.t.: } \sum\limits_{c=1}^{C} u_{ic}=1, \quad u_{ic}\geq 0, \end{cases} \tag{2}$$

The constant m, with $m \geq 1$, is the fuzziness parameter such that the greater the value of m the more "fuzzy" is the obtained partition.

Fixed $\mathcal{I}$, it is easy to see that the optimal parameters are:

$$u_{ic} = \frac{1}{\sum\limits_{c'=1}^{C}\left[\frac{\sum_{s=1}^{p}(\beta_i^s-\hat{\beta}_c^s)^2}{\sum_{s=1}^{p}(\beta_i^s-\hat{\beta}_{c'}^s)^2}\right]^{\frac{1}{m-1}}}. \tag{3}$$

and that

$$\hat{\boldsymbol{\beta}}_c = \frac{\sum\limits_{i\in\mathcal{I}} u_{ic}^m \boldsymbol{\beta}_i}{\sum\limits_{i\in\mathcal{I}} u_{ic}^m}. \tag{4}$$

Equations (3) and (4) would include the BS-based Fuzzy C-Means Clustering method (BS-FCMC) as a limit case when $\alpha = 0$.

On the other hand, for fixed $\{u_{ic}\}$ membership values and fixed $\{\hat{\boldsymbol{\beta}}_c\}$ centroids, the optimal "non-trimmed set" of observations can be obtained by following the steps for updating $\mathcal{I}$ described in Algorithm 1.

For a given value of α, a sensible choice for the number of clusters C can be determined by considering typical internal validity indices for fuzzy clustering, like the Xie-Beni index [15] or the Fuzzy Silhouette [16].

However, in general the choice of C and α have to be done simultaneously. Indeed, an incorrect choice of C could hamper the detection of outlying objects, and at the same time an incorrect determination of α could prevent the identification of the correct number of clusters.

In this work, we follow the approach described in [17]. Basically, for a given C we plot the value of the objective function in (4) against increasing values of α. If the rate of decrease is smooth all over the range of α values, then there are no outliers. If there is an initial fast decrease rate that stabilizes when reaching certain trimming level α_0 then this would indicate that α_0 could be a sensible trimming level for that given C. To detect the correct number of clusters we can also look at the numerical second derivative of the objective function against α. If the number of clusters is incorrect, then there will be one or more "peaks" in the numerical second derivative. This is due to the fact that a whole cluster is being deleted for given values of α so that we start trimming observations from another subpopulation, so introducing changes in the decreasing rates. On the other hand, if the number of clusters is correct we would only observe two different behaviour of the second derivative, depending on whether we are trimming outliers or not.

Notice that this approach is closely related to other robust fuzzy clustering procedures as [18, 19] based on Least Trimmed Squares (LTS) principles. This LTS approach also underlies in well-know approaches for robust regression and robust multivariate estimation [see, e.g., 20] and robust Principal Components Analysis [21]. The LTS approach to robust 'crisp" clustering was first introduced in [13] in the so-called trimmed k-means approach. Trimmed k-means was already applied in combination with B-splines in [12] as a proposal for robust functional clustering. Finally, the FCLUST method in [22] is a robust fuzzy clustering approach that also considers trimming throughout constraints,that uses the same type of membership values as in (4), but also estimating cluster variance-covariance matrices to account for non-spherical cluster structures.

The computational steps of the proposed clustering model are described in Algorithm 1.

2.3. BS-based Trimmed Fuzzy C-Medoids Clustering model (BS-Tr-FCMdC)

Given a trimming size α which ranges between 0 and 1, we solve the minimization problem:

$$\begin{cases} f_{C,\alpha,m}(\{u_{ic}\},\mathcal{I},\{i_c\}) = \min\limits_{\{u_{ic}\}} \min\limits_{\substack{\mathcal{I}\subset\{1,\ldots,n\}\\ \#\mathcal{I}=\lfloor n\cdot(1-\alpha)\rfloor}} \min\limits_{\{i_1,\ldots,i_C\}\subset\{1,\ldots,n\}} \sum\limits_{i\in\mathcal{I}} \sum\limits_{c=1}^{C} u_{ic}^m \sum\limits_{s=1}^{p} (\beta_i^s - \beta_{i_c}^s)^2 \\ \text{s.t.:} \ \sum\limits_{c=1}^{C} u_{ic} = 1, \quad u_{ic} \geq 0, \end{cases} \tag{5}$$

Note that here the centroids are now associated to C specifically chosen parameters $\boldsymbol{\beta}_{i_1}$,...,$\boldsymbol{\beta}_{i_C}$ with $\{i_1, ..., i_C\} \subset \{1, 2, ..., n\}$ (C-medoids) in the B-splines representation and not associated to synthetical parameters. The problem (5) includes the BS-based Fuzzy C-Medoids Clustering method (BS-FCMdC) as a limit case when $\alpha = 0$. Each non-trimmed time series is allocated into the cluster corresponding to its closest C-medoid time series via its fitted coefficients.

It is easy to see that the local optimal membership u_{ic} values are:

$$u_{ic} = \frac{1}{\sum\limits_{c'=1}^{C} \left[\frac{\sum\limits_{s=1}^{p}(\beta_i^s - \beta_{i_c}^s)^2}{\sum\limits_{s=1}^{p}\left(\beta_{i_c}^s - \beta_{i_{c'}}^s\right)^2} \right]^{\frac{1}{m-1}}}. \tag{6}$$

Algorithm 1 BS-based Trimmed Fuzzy C-Means Clustering model (BS-Tr-FCMC) algorithm

1: Fix C, the trimming level α, `nstart` and `max.iter`;

2: Set $iter = 0$;

3: Randomly initialize $U = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ with $\boldsymbol{u}_i = (u_{i1}, ..., u_{iC})$ and $\sum_{c=1}^{C} u_{ic} = 1$ for every i;

4: Randomly initialize $\mathcal{I} \subset \{1, ..., n\}$ with $\#\mathcal{I} = \lfloor n \cdot (1-\alpha) \rfloor$;

5: Randomly set C initial centroids: $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C$;

6: **repeat**

7: Store the current centroids $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C)_{OLD} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C)$;

8: Compute $\mathbf{u}_i$ $(i = 1, \ldots, n)$ by using (3);

9: Compute the new centroids $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C$ by using (4);

10: Compute

$$r_i = \sum_{c=1}^{C} u_{ic}^m \sum_{s=1}^{p} (\hat{\beta}_i^s - \tilde{\beta}_c^s)^2 \text{ for } i = 1, \ldots, n,$$

and sort them in $r_{(1)} \leq ... \leq r_{(n)}$;

11: Take $\mathcal{I} = \{i : r_i \leq r_{(\lfloor n \cdot (1-\alpha) \rfloor)}\}$;

12: $iter \leftarrow iter + 1$;

13: **until** $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C)_{OLD} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C)$ or $iter =$ `max.iter`

14: *Repeat initialization:* Start `nstart` times from step 2 and provide as output of the algorithm those parameters yielding the smallest value of the target function (4) out of these `nstart` random initializations.

The computational steps of the proposed clustering model are described in Algorithm 2.

Algorithm 2 BS-based Trimmed Fuzzy C-Medoids Clustering (BS-Tr-FCMdC) algorithm

1: Fix C, the trimming level α, `nstart` and `max.iter`;

2: Set $iter = 0$;

3: Randomly initialize $U = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ with $\boldsymbol{u}_i = (u_{i1}, ..., u_{iC})$;

4: Randomly initialize $\mathcal{I} \subset \{1, ..., n\}$ with $\#\mathcal{I} = \lfloor n \cdot (1 - \alpha) \rfloor$;

5: Randomly set C initial medoids: $\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C}$;

6: **repeat**

7: Store the current medoids $(\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C})_{OLD} = (\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C})$;

8: Compute $\mathbf{u}_i$ $(i = 1, \ldots, n)$ by using (6);

9: Select the new centroids $\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C}$ where

$$i_c = \arg \min_{1 \leq i' \leq n} \sum_{i \in \mathcal{I}} \sum_{c=1}^{C} u_{ic}^m \sum_{s=1}^{p} (\beta_i^s - \beta_{i'}^s)^2.$$

10: Compute

$$r_i = \sum_{c=1}^{C} u_{ic}^m \sum_{s=1}^{p} \left(\beta_i^s - \beta_{i_c}^s\right)^2 \text{ for } i = 1, \ldots, n,$$

and sort them in $r_{(1)} \leq ... \leq r_{(n)}$;

11: Take $\mathcal{I} = \{i : r_i \leq r_{(\lfloor n \cdot (1-\alpha) \rfloor)}\}$;

12: $iter \leftarrow iter + 1$;

13: **until** $(\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C})_{OLD} = (\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C})$ or $iter =$ `max.iter`

14: *Repeat initialization:* Start `nstart` times from step 2 and provide as output of the algorithm those parameters yielding the smallest value of the target function (4) out of these `nstart` random initializations.

Notice that the Algorithm 2 falls in the category of Alternating Cluster Estimation paradigm [23]. Therefore, as also happens with all the algorithms described in this section, it is not guaranteed that the global minimum is reached and, thus, considering a large number `nstart` of random starts is suggested to achieve the optimal solution.

2.4. BS-based Noise Cluster Fuzzy C-Means Clustering model (BS-NC-FCMC)

The BS-based Noise Cluster Fuzzy C-Means Clustering model (BS-NC-FCMC model) neutralizes the negative effects of noise and outliers data in deviating the centroids from their true positions introducing the so-called noise cluster, a cluster collecting units far away from the natural C clusters in the data in an "articial" $(C+1)$-th cluster. FcM-NC has been initially proposed by [24], which uses a criterion similar to [25], and later extended by [26]. It is assumed that the distance measure of unit i from centroid C is equal to δ^2 for $i = 1, \ldots, n$. We now set $\boldsymbol{u}_i = (u_{i1}, ..., u_{iC}, u_{i(C+1)})$ and consider, again, C centroids denoted by $\hat{\boldsymbol{\beta}}_1$,...,$\hat{\boldsymbol{\beta}}_C$. Formally, the BS-based Noise Cluster Fuzzy C-Means Clustering model (BS-NC-FCMC) is characterized in the following way:

$$\begin{cases} f_{C,\delta,m}(\{u_{ic}\},\{\hat{\boldsymbol{\beta}}_c\}) := \min\limits_{u_{ic}} : \sum\limits_{i=1}^{n}\sum\limits_{c=1}^{C} u_{ic}^m \sum\limits_{s=1}^{p}(\beta_i^s - \hat{\beta}_c^s)^2 + \sum_{i=1}^{n} u_{i(C+1)}^m \delta^2 \\ \text{s.t.:} \sum\limits_{c=1}^{C+1} u_{ic} = 1, \quad u_{ic} \geq 0, \end{cases} \tag{7}$$

The distance from the noise cluster depends on the average distance among units and centroids $\delta^2 = \rho(nC)^{-1}\sum_{i=1}^{n}\sum_{c=1}^{C} u_{ic}^m \sum_{s=1}^{p}(\beta_i^s - \hat{\beta}_c^s)^2$. The value of ρ may range between 0.05 and 0.5. In any case, the results do not seem very sensitive to the value of the multiplier ρ [24]. Due to the presence of δ^2, units that are close to "good" clusters are correctly classified in a good cluster while the noise units that are away from "good" clusters are classified in the "articial" $C+1$ noise cluster.

The solutions of (7) are:

$$u_{ic} = \frac{\left[\frac{1}{\sum\limits_{s=1}^{p}(\beta_i^s - \hat{\beta}_c^s)^2}\right]^{\frac{1}{m-1}}}{\sum_{c'=1}^{C}\left[\frac{1}{\sum\limits_{s=1}^{p}\left(\beta_i^s - \hat{\beta}_{c'}^s\right)^2}\right]^{\frac{1}{m-1}} + \left[\frac{1}{\delta^2}\right]^{\frac{1}{m-1}}}, \tag{8}$$

$$u_{i(C+1)} = \frac{\left[\frac{1}{\delta^2}\right]^{\frac{1}{m-1}}}{\sum_{c'=1}^{C}\left[\frac{1}{\sum\limits_{s=1}^{p}(\beta_i^s - \hat{\beta}_{c'}^s)^2}\right]^{\frac{1}{m-1}} + \left[\frac{1}{\delta^2}\right]^{\frac{1}{m-1}}} \tag{9}$$

and

$$\hat{\boldsymbol{\beta}}_c = \frac{\sum\limits_{i=1}^{n} u_{ic}^m \boldsymbol{\beta}_i}{\sum\limits_{i=1}^{n} u_{ic}^m} \text{ for } c = 1, \ldots, C. \tag{10}$$

The computational steps of the proposed clustering model are described in algorithm (3).

Algorithm 3 BS-based Noise Cluster Fuzzy C-Means Clustering model (BS-NC-FCMC) algorithm

1: Fix C, δ, `nstart` and `max.iter`;

2: Set $iter = 0$;

3: Randomly initialize $U = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ with $\boldsymbol{u}_i = (u_{i1}, ..., u_{iC})$;

4: Randomly set C initial centroids: $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C$;

5: **repeat**

6: Store the current centroids $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C)_{OLD} = \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C$;

7: Compute $\mathbf{u}_i$ $(i = 1, \ldots, n)$ by using (8) and (9);

8: Compute the new centroids $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C$ by using (10);

9: $iter \leftarrow iter + 1$;

10: **until** $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C)_{OLD} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \ldots \hat{\boldsymbol{\beta}}_C)$ or $iter =$ `max.iter`

11: *Repeat initialization:* Start `nstart` times from step 2 and provide as output of the algorithm those parameters yielding the smallest value of the target function (4) out of these `nstart` random initializations.

2.5. BS-based Noise Cluster Fuzzy C-Medoids Clustering model (BS-NC-FCMdC)

The method in Section 2.4 can be adapted by assuming that the cluster prototypes coincide with C fitted β_{i_c} B-spline coefficients representations for C time series that will act as C-medoids. This model also achieves its robustness with respect to outliers by introducing an "artificial" noise cluster represented by a noise prototype, i.e. a "fictitious" noise medoid, which is always at the same distance from all units. By following [24], let there be C good clusters and let the $(C+1)$-th cluster be the noise cluster.

Formally, the BS-based Noise Cluster Fuzzy C-Medoids Clustering model (BS-NC-FCMdC) is characterized in the following way:

$$\begin{cases} f_{C,\delta,m}(\{u_{ic}\},\{i_c\}) := \min\limits_{\{u_{ic}\}} \min\limits_{\{i_1,\ldots,i_C\}\subset\{1,\ldots,n\}} \sum\limits_{i=1}^{n}\sum\limits_{c=1}^{C} u_{ic}^m \sum\limits_{s=1}^{p} \left(\beta_i^s - \beta_{i_c}^s\right)^2 + \sum_{i=1}^{n} u_{i(C+1)}^m \delta^2 \\ \text{s.t.: } \sum\limits_{c=1}^{C+1} u_{ic} = 1, \quad u_{ic} \geq 0. \end{cases} \tag{11}$$

As in Section 2.4, it is common to set δ such that

$$\delta^2 = \rho(nC)^{-1} \sum_{i=1}^{n}\sum_{c=1}^{C} u_{ic}^m \sum_{s=1}^{p} \left(\beta_i^s - \beta_{i_c}^s\right)^2.$$

The optimal membership u_{ic} values in (11) satisfy:

$$u_{ic} \quad = \frac{\left[\frac{1}{\sum\limits_{s=1}^{p}(\beta_i^s-\beta_{i_c}^s)^2}\right]^{\frac{1}{m-1}}}{\sum_{c'=1}^{C}\left[\frac{1}{\sum\limits_{s=1}^{p}\left(\beta_i^s-\beta_{i_{c'}}^s\right)^2}\right]^{\frac{1}{m-1}}+\left[\frac{1}{\delta^2}\right]^{\frac{1}{m-1}}}, \tag{12}$$

and

$$u_{i(C+1)} = \frac{\left[\frac{1}{\delta^2}\right]^{\frac{1}{m-1}}}{\sum_{c'=1}^{C}\left[\frac{1}{\sum\limits_{s=1}^{p}\left(\beta_i^s-\beta_{i_{c'}}^s\right)^2}\right]^{\frac{1}{m-1}}+\left[\frac{1}{\delta^2}\right]^{\frac{1}{m-1}}} \tag{13}$$

The computational steps of the proposed robust clustering model are described in Algorithm 4.

Algorithm 4 BS-based Noise Cluster Fuzzy C-Medoids Clustering (BS-NC-FCMdC) algorithm

1: Fix C, δ, `nstart` and `max.iter`;

2: Set $iter = 0$;

3: Randomly initialize $U = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ with $\boldsymbol{u}_i = (u_{i1}, ..., u_{iC})$;

4: Randomly set C initial medoids: $\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C}$;

5: **repeat**

6: Store the current medoids $(\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C})_{OLD} = (\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C})$;

7: Compute $\mathbf{u}_i$ $(i = 1, \ldots, n)$ by using (12) and (13);

8: Select the new centroids $\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C}$ where

$$i_c = \arg \min_{1 \le i' \le n} \sum_{i \in \mathcal{I}} \sum_{c=1}^{C} u_{ic}^m \sum_{s=1}^{p} (\beta_i^s - \beta_{i'}^s)^2.$$

9: $iter \leftarrow iter + 1$;

10: **until** $(\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C})_{OLD} = (\boldsymbol{\beta}_{i_1}, \boldsymbol{\beta}_{i_2}, ...\boldsymbol{\beta}_{i_C})$ or $iter =$ `max.iter`

11: *Repeat initialization:* Start `nstart` times from step 2 and provide as output of the algorithm those parameters yielding the smallest value of the target function (4) out of these `nstart` random initializations.

3. Simulation study

To assess the clustering performance and accuracy of the proposed robust fuzzy clustering models BS-Tr-FCMC, BS-Tr-FCMdC, BS-NC-FCMC, BS-NC-FCMdC, in this section we show the results of a simulation study. For comparison purposes, we have considered the robust crisp versions BS-Tr-CMC, BS-Tr-CMdC, BS-NC-CMC, BS-NC-CMdC, following from the choice of the fuzzifier parameter $m = 1$, and the robust fuzzy Partitionig Around Medoids models DTW-Tr-FCMdC [27] Dynamic Time Warping-based trimmed Fuzzy clustering model, and DTW-Exp-FCMd [28] Dynamic Time Warping-based Fuzzy C-Medoids model with exponential transformation.

The dataset is made of 20 time series from two well separated clusters, 10 time series obtained from the function $f(x) = \sin(x)$ and 10 from $g(x) = \cos(x)$, evaluated in a grid of size 30 along the interval $[0, 2\pi]$. A $N(0, 0.2^2)$-distributed error is added. Two outlying time series centered at line $l(x) = 2$ were considered and two outlying time series centered at line $h(x) = -\pi + x$ were considered too. Two switching time series were considered, one centered at $f(x) = \sin(x)$ for the first half of the grid and at $g(x) = cos(x)$ for the second half, the other centered at $g(x) = \cos(x)$ for the first half of the grid and at $f(x) = \sin(x)$ for the second half. The scenarios obtained were 6, considering the 20 time series (scenario 1), the 20 time series with 2 constant outliers (scenario 2), the 20 time series with 2 linear outliers (scenario 3), the 20 time series with 2 switching outliers (scenario 4), the 20 time series with 2 switching outliers and 2 constant outliers (scenario 5), the 20 time series with 2 switching outliers and 2 linear outliers (scenario 6). The scenarios are shown in Figure 1.

The comparison between the robust fuzzy model BS-Tr-FCMC and the non-robust fuzzy model BS-FCMC (untrimmed by proper choice of the trimming parameter α equal to zero) is shown in Figure 2 for scenario 3. The untrimmed model does not recover the two cluster structures and the two main groups are joined together. The coordinates of the trimmed and untrimmed centroids, acting as coefficients in the B-spline basis, give the cluster center trajectories.

The trimming parameter is set on the basis of the number of outliers (to trim 2 outliers) and $S = 5$ interior knots were considered.

The clustering performance and accuracy of each clustering model were evaluated according to whether time series generated from the same process were grouped in the same cluster, with membership degrees equal or close to one in that cluster. Since switching time series should belong simultaneously to more than one cluster, the performance of each clustering model was also assessed

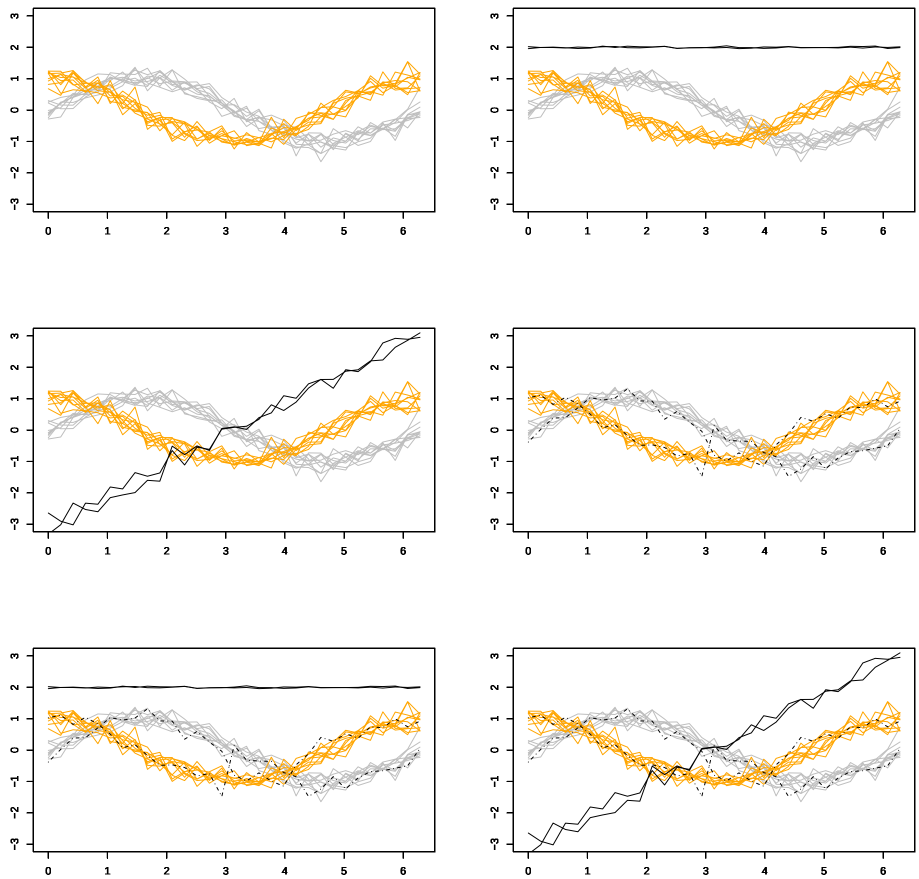


Figure 1: Simulation - 6 scenarios, numbered by row

according to the capability of the model to identify switching time series. For each scenario, 100 simulations were carried out. For each simulation, the percentage of time series correctly identified as belonging to one of the cluster, or as switching time series is computed. Then, the average percentage of correct classification over the 100 simulations for each scenario is computed. The average percentage of correct classification is a measure of clustering accuracy of the model. The higher this value, the better the classification performance of the model considered. To assign each curve to a specific cluster we have set cut-off values. In scenarios 1-3, with no switching time series, we have assigned the i-th curve to the c-th cluster if its fuzzy membership degree was $u_{ic} > 0.70$.

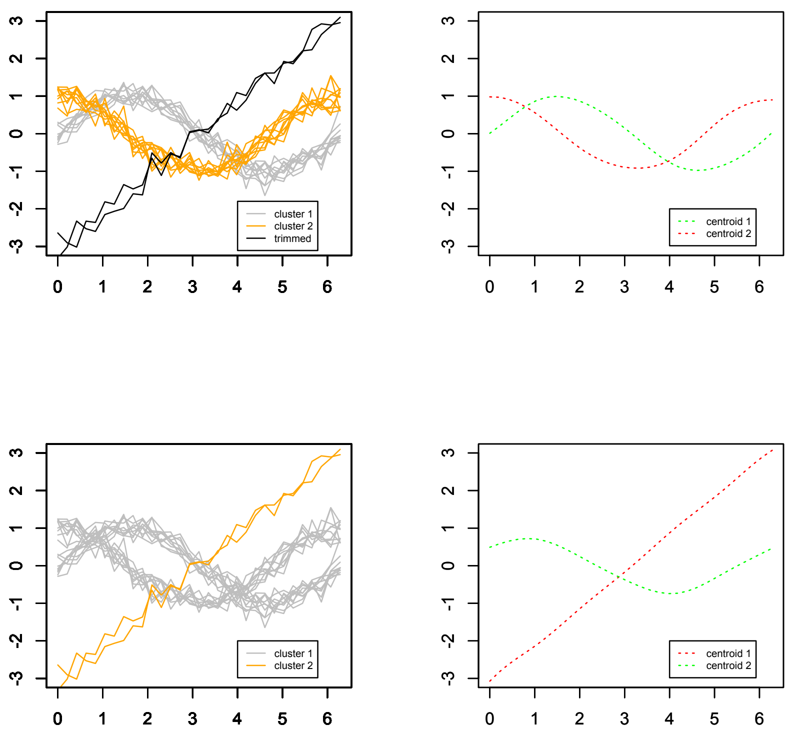


Figure 2: Robust clustering top, Non-robust clustering bottom, clusters left, centroids right

In scenarios 4-6, to identify the switching time series, we have set the membership degrees in the interval $(0.3, 0.7)$. Note that the selected cut-off values are compatible with those suggested in literature for simulation studies [29]. In all scenarios we have set the fuzziness parameter m at different values, namely $1.3, 1.5, 1.7, 2.0$ to check if results are influenced by the degree of fuzziness. For each scenario, performance values are obtained by averaging for the different values of m. The Fuzzy Silhouette (Silhouette for crisp models) [30] for the 6 scenarios was also computed and averaged over the simulations. Overall results for clustering performance and accuracy of the models are summarized in Tables 1, 2. The (averaged) membership degrees are also presented in Tables 3-6. For the computation of the performances the 20 (no swithing time series, scenarios 1-3) and 22 (2 switching time series, scenarios 4-6) time series were considered. We expect that when the dataset is contaminated only with outlier time series, the model show comparable performances due to robustness; when the dataset is contaminated with outlier and/or switching time series, the proposed fuzzy models BS-Tr-FCMC, BS-Tr-FCMdC, BS-NC-FCMC, BS-NC-FCMdC models show better performances due to fuzziness.

Scenario	**BS-Tr-FCMC**	**BS-Tr-FCMdC**	**BS-NC-FCMC**	**BS-NC-FCMdC**	**DTW-Tr-FCMdC**	**DTW-Exp-FCMd**	**BS-Tr-CMC**	**BS-Tr-CMdC**	**BS-NC-CMC**	**BS-NC-CMdC**
1 (0,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2 (2,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3 (2,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4 (0,2)	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.91	0.91	0.91
5 (2,2)	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.91	0.91	0.91
6 (2,2)	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.91	0.91	0.91

Table 1: Simulation - Percentage of correct classification (in parenthesis the number of outliers and switching time series)

Scenario	**BS-Tr-FCMC**	**BS-Tr-FCMdC**	**BS-NC-FCMC**	**BS-NC-FCMdC**	**DTW-Tr-FCMdC**	**DTW-Exp-FCMd**	**BS-Tr-CMC**	**BS-Tr-CMdC**	**BS-NC-CMC**	**BS-NC-CMdC**
1 (0,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2 (2,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3 (2,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4 (0,2)	0.93	0.93	0.93	0.93	0.93	0.93	0.86	0.85	0.84	0.87
5 (2,2)	0.91	0.92	0.91	0.92	0.91	0.93	0.87	0.89	0.86	0.85
6 (2,2)	0.90	0.91	0.90	0.91	0.89	0.90	0.83	0.84	0.84	0.85

Table 2: Simulation - Fuzzy Silhouette (in parenthesis the number of outliers and switching time series)

The results show that the robust fuzzy and non-fuzzy models present the same behavior in scenarios without switching time series. The robust fuzzy models outperform the robust non fuzzy models in scenarios with switching time series. The performances of the exponential model DTW-Exp-FCMd are comparable to the performances of the robust fuzzy models in most scenarios.

4. Real data application

In this paper, the proposed clustering approaches are applied to study the performance over time of the stocks currently composing the FTSE-MIB index (Financial Times Stock Exchange Milano Indice di Borsa) listed on the Italian Stock Exchange owned by the London Stock Exchange[1]. The FTSE-MIB index is the primary benchmark for the Italian Stock market accounting approximately for 80% of the domestic market capitalization; the stocks composing the index can vary over time,

[1]Data can be free downloaded from the finance section of the Yahoo website (http://finance.yahoo.com).

	BS-Tr-FCMC		BS-Tr-FCMdC		BS-NC-FCMC		BS-NC-FCMdC		DTW-Tr-FCMdC		DTW-Exp-FCMd		BS-Tr-CMC		BS-Tr-CMdC		BS-NC-CMC		BS-NC-CMdC	
1	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
2	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
3	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
4	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
5	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
6	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
7	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
8	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
9	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
10	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
11	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
12	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
13	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
14	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
15	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
16	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
17	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
18	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
19	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
20	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**

Table 3: Averaged membership values - Scenarios 1, 2, 3

	BS-Tr-FCMC		BS-Tr-FCMdC		BS-NC-FCMC		BS-NC-FCMdC		DTW-Tr-FCMdC		DTW-Exp-FCMd		BS-Tr-CMC		BS-Tr-CMdC		BS-NC-CMC		BS-NC-CMdC	
1	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
2	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
3	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
4	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
5	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
6	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
7	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
8	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
9	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
10	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
11	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
12	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
13	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
14	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
15	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
16	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
17	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
18	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
19	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
20	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
21	**0.61**	0.39	**0.51**	0.49	**0.59**	0.37	**0.50**	0.38	**0.57**	0.43	**0.51**	0.49	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
22	0.49	**0.51**	0.43	**0.57**	0.38	**0.58**	0.38	**0.49**	0.42	**0.58**	0.41	**0.59**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**

Table 4: Averaged membership values - Scenario 4

based on their market capitalization and liquidity; moreover any stock may never account for more than 15% of the index.

The FTSE-MIB index has been elaborated to be used for futures and options trading, as a benchmark index for Exchange Traded Funds (ETFs), and for tracking the most liquid and capitalized stocks in the Italian market. Indeed, the 40 companies belong to the major industry groups

	BS-Tr-FCMC		BS-Tr-FCMdC		BS-NC-FCMC		BS-NC-FCMdC		DTW-Tr-FCMdC		DTW-Exp-FCMd		BS-Tr-CMC		BS-Tr-CMdC		BS-NC-CMC		BS-NC-CMdC	
1	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
2	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
3	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
4	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
5	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
6	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
7	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
8	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
9	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
10	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
11	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
12	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
13	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
14	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
15	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
16	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
17	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
18	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
19	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
20	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
21	**0.63**	0.37	**0.58**	0.42	**0.61**	0.39	**0.57**	0.35	**0.61**	0.39	**0.50**	0.50	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
22	0.46	**0.54**	0.43	**0.57**	0.40	**0.60**	0.38	**0.49**	0.38	**0.62**	0.44	**0.56**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**

Table 5: Averaged membership values - Scenario 5

	BS-Tr-FCMC		BS-Tr-FCMdC		BS-NC-FCMC		BS-NC-FCMdC		DTW-Tr-FCMdC		DTW-Exp-FCMd		BS-Tr-CMC		BS-Tr-CMdC		BS-NC-CMC		BS-NC-CMdC	
1	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
2	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
3	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
4	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
5	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
6	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
7	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
8	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
9	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
10	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
11	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
12	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
13	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
14	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
15	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
16	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
17	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
18	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
19	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
20	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**
21	**0.67**	0.33	**0.66**	0.34	**0.62**	0.28	**0.57**	0.34	**0.69**	0.31	**0.58**	0.42	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00
22	0.35	**0.65**	0.41	**0.59**	0.46	**0.50**	0.38	**0.54**	0.41	**0.59**	0.41	**0.59**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	**1.00**

Table 6: Averaged membership values - Scenario 6

operating in important sectors such as computer hardware and software, telecommunications, retail/wholesale trade and biotechnology, for the US market, financial companies, utilities and energy related groups for Italy. For all these reasons, it has been object of many interesting analysis and recent contributions [31].

The best performance of Borsa Italiana has been recorded on October 13, 2008, when the FTSE-MIB index increased of 11.49%, the worst one on March 12, 2020, when the same index recorded a slump of 16.92% due to COVID-19 pandemic.

The main objective is to find a partition of the FTSE-MIB components[2] based on the daily adjusted closing price, spanning from 02/01/2018 to 25/01/2021; it is particularly suitable to analize historical returns [32] and, properly, corrects each stock closing price to reflect that stock's value after accounting for any corporate actions that might occur before the new opening day. The time series of the daily adjusted closing price, spanning from 02/01/2018 to 25/01/2021, for the 39 stocks under consideration, are shown in Figure 3.

[2]The *Nexi* group is excluded because it is the only component with missing values for the period 02/01/2018 – 16/04/2019.

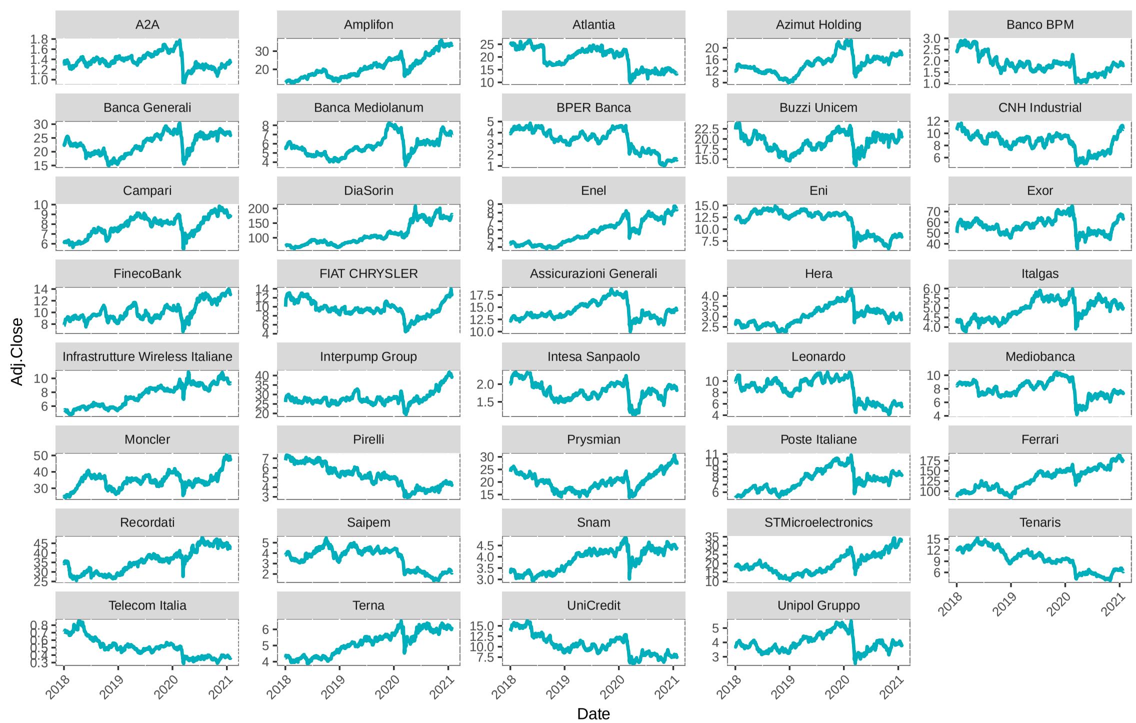


Figure 3: the time series of the daily adjusted closing price, from 02/01/2018 to 25/01/2021

The following B-spline based robust fuzzy clustering models have been applied: the BS-based Noise Cluster Fuzzy C-Means Clustering model (BS-NC-FCMC), the BS-based Noise Cluster Fuzzy C-Medoids Clustering model (BS-NC-FCMdC), the BS-based Trimmed Fuzzy C-Means Clustering method (BS-Tr-FCMC) and the BS-based Trimmed Fuzzy C-Medoids Clustering model (BS-Tr-FCMdC).

For each time series, the associated cubic B-spline basis has made up of $p = 54$ basis elements, having fixed the number of equispaced knots to $S = 50$. All algorithms have been run fixing the fuzziness parameter $m = 1.5$.

For the BS-NC-FCMC and BS-NC-FCMdC models, the number of groups has been chosen according to the Fuzzy Silhouette FS index maximization [16]. Both methods identified two clusters with three units belonging to the noise cluster. In Table 7 we reported their FS values when $C \in \{2, 3\}$ taking into account that, when $C = 3$, they rise to 0.826 and 0.823 for the BS-NC-FCMC and BS-NC-FCMdC model, respectively, if the unit 15 is excluded.

For the BS-Tr-FCMC and the BS-Tr-FCMdC models, C and α are simultaneously chosen according to the proposal in [17] and already applied in a fuzzy context in [33] and [27].

As already outlined in the introduction, the corresponding optimal value of the objective function $f_{C,\alpha,m}$ were separately plotted against increasing values of α, for each fixed C and m. By inspecting these curves , the smallest C for which the rate of decrease is smooth over all the range of α trimming levels could suggest a sensible number of clusters if there are no outliers. On the other hand, an initial fast decrease in $f_{C,\alpha,m}$ before reaching certain α_0 would suggest that outliers are being initially trimmed before reaching that α_0 trimming level.

To identify the right C, it is also useful to look at $f''_{C,\alpha,m}$ being the numerical second derivative of $f_{C,\alpha,m}$ defined as:

$$f''_{C,\alpha,m} = \frac{f_{C,\alpha-h,m} - 2f_{C,\alpha,m} + f_{C,\alpha+h,m}}{h^2},$$

where $h \in (0, 1)$ is a kind of tuning parameter that controls the roughness of the numerical second derivative, in such a way that they are more rough and data dependent when h is small. If the number of C is incorrectly identified, there will be one or more "peaks" in this numerical second derivative. If C is correctly identified, the behaviour depends on the presence of outliers: if there are no outliers, the smallest C for which the trend is smooth over all values of α is the right number of clusters, otherwise we can observe a fast decrease of the trend up to the right trimming value α, after that a smooth decay is observed.

According to [27], in this study, the value of the objective function and of its numerical second derivative were recorded for each value of $C \in \{2, ..., 6\}$ and of $\alpha \in [0, 0.5]$ (with step equal to 0.025) and plotted in Figure 4 and Figure 5 for the BS-Tr-FCMC and BS-Tr-FCMdC model, respectively.

According to the previously described procedure, for both BS-Tr-FCMC and BS-Tr-FCMdC models, regardless the value of C, the number of unit to be trimmed was equal to three, corresponding to $\alpha = 0.075$, where the objective function showed an elbow. The best number of group to take under consideration was $C = 2$ and $C = 3$. To support these evidences, their Fuzzy Silhouettes have been reported in Table 7 when fixing $\alpha = 0.075$. It is worth noting that, looking at the patterns observed in the second derivatives corresponding to the BS-Tr-FCMdC method, also the solution with $C = 4$ could have been considered. We did not focus on it since its associated Fuzzy Silhouette (FS$= 0.74$) was sensible lower if compared with those in Table 7.

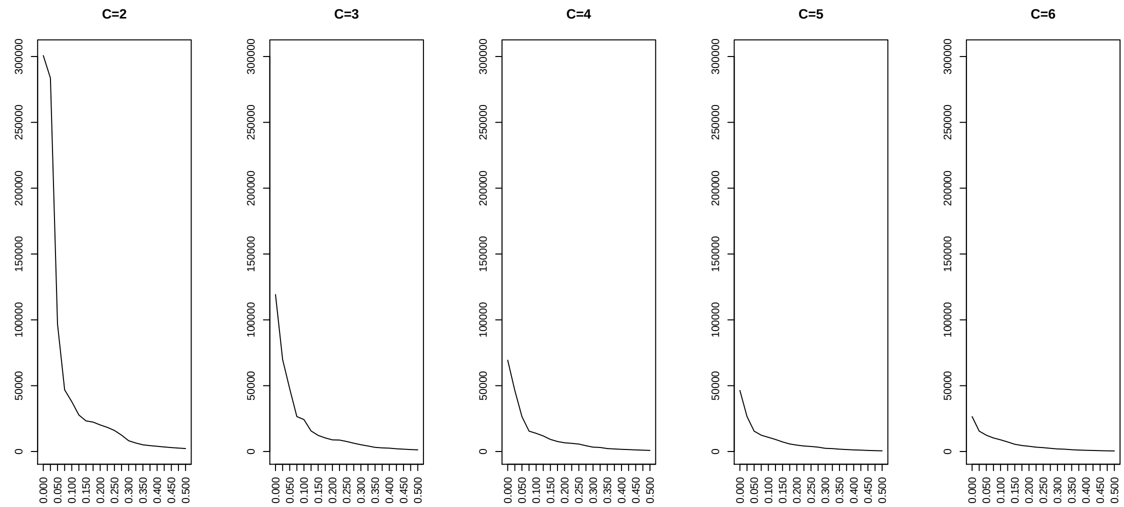
	$C = 2$	$C = 3$
BS-NC-FCMC	0.851	0.789
BS-NC-FCMdC	0.854	0.785
BS-Tr-FCMC	0.850	0.815
BS-Tr-FCMdC	0.854	0.823

Table 7: The Fuzzy silhouette index for two and three clusters solution when $\alpha = 0.075$.

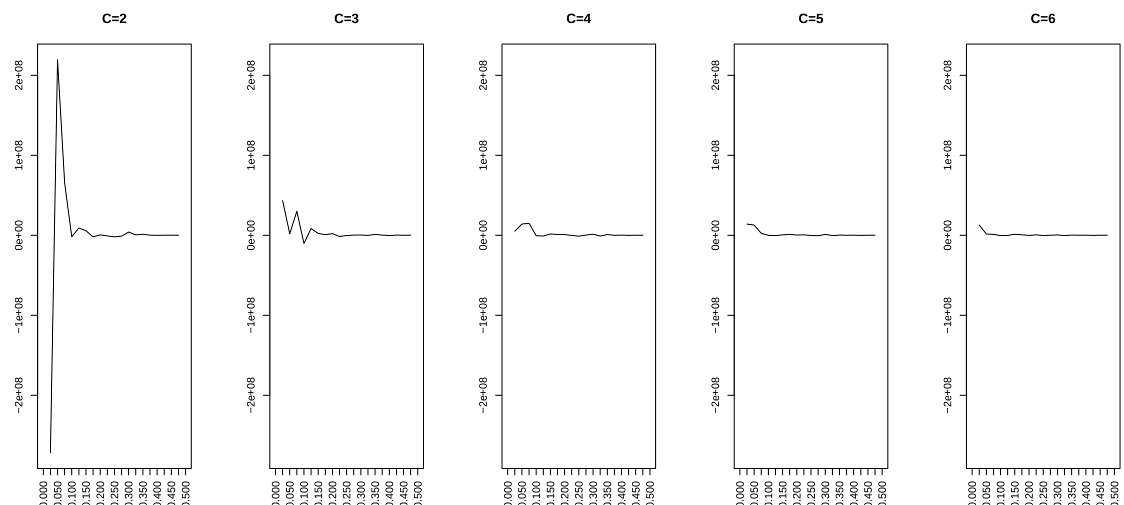
The membership degrees, for all models, have been reported in Table 8 for $C = 2$ and in Table 9 for $C = 3$, with the medoids highlighted in bold and the outliers in italic. For two groups, the "fuzzy units" have been identified with those whose membership degree was in the interval $(0.3, 0.7)$, and with those whose membership degree was in the interval $(0.3, 0.6)$ for the three groups case.

With the exception, when $C = 2$, of the fuzzy units *Azimut Holding*, for the BS-NC-FCMC and the BS-Tr-FCMC models, and *Azimut Holding* and *Generali Assicurazioni*, for both the BS-NC-FCMdC and BS-Tr-FCMdC models, all stocks have been assigned to one of the clusters with a low degree of uncertainty. For both solutions, the membership degree matrix was almost the same across models, leading to the same crisp partition with the exception, when $C = 3$, of unit 15 that has not been classified as outlier by the BS-NC-FCMC and BS-NC-FCMC models and unit 22, differently classified only by the BS-NC-FCMC model.

We focused on the solution with three clusters that, compared to the one with two clusters,

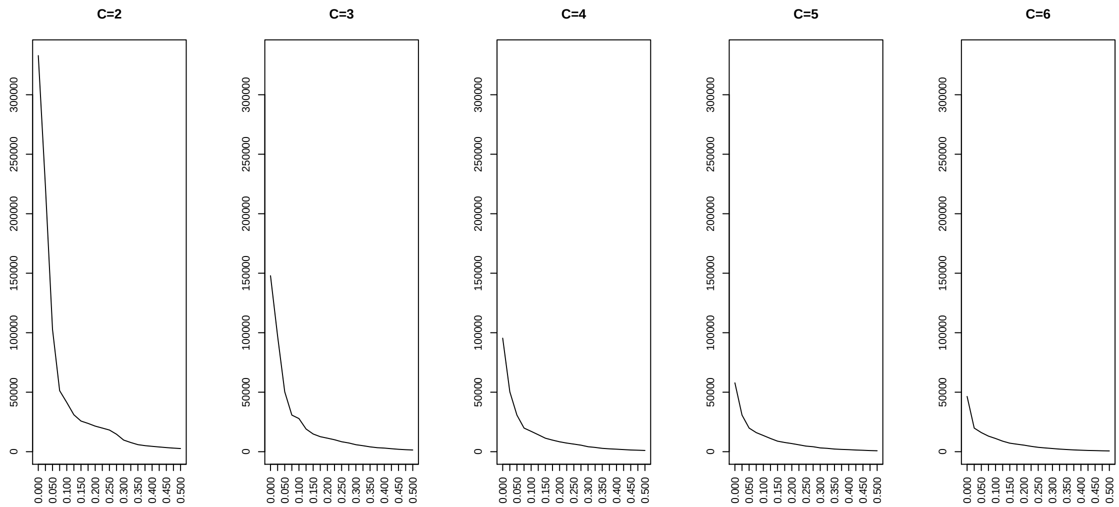


(a) Objective function

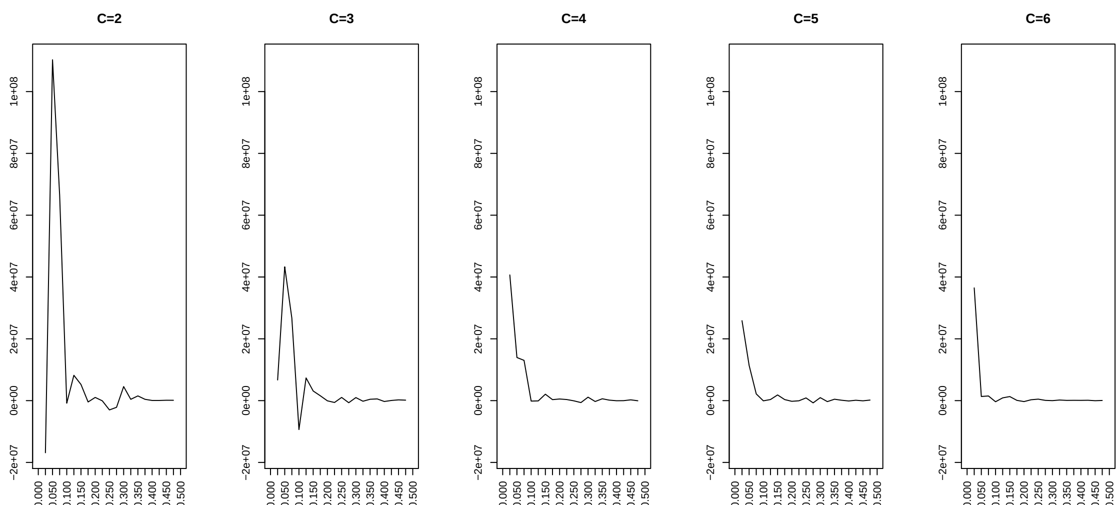


(b) Numerical second derivative

Figure 4: BS-Tr-FCMC model: objective function and numerical second derivative for different values of C and α (α values in the x-axis)



(a) Objective function



(b) Numerical second derivative

Figure 5: BS-Tr-FCMdC model: objective function and numerical second derivative for different values of C and α (α values in the x-axis)

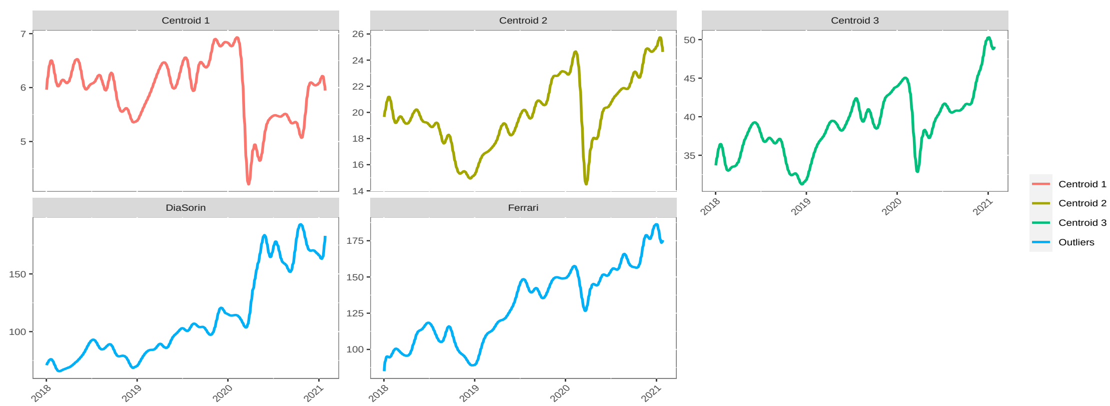
		BS-NC-FCMC			BS-NC-FCMdC			BS-Tr-FCMC		BS-Tr-FCMdC	
		Cluster 1	Cluster 2	Noise Cluster	Cluster 1	Cluster 2	Noise Cluster	Cluster 1	Cluster 2	Cluster 1	Cluster 2
1	A2A	0.995	0.002	0.003	0.998	0.002	0.001	0.998	0.002	0.998	0.002
2	Amplifon	0.009	0.988	0.003	0.008	0.991	0.001	0.009	0.991	0.008	0.992
3	Atlantia	0.098	0.889	0.013	0.047	0.949	0.003	0.098	0.902	0.047	0.953
4	Azimut Holding	0.651	0.335	0.014	0.443	0.552	0.005	0.651	0.349	0.445	0.555
5	Banco BPM	0.997	0.002	0.001	0.998	0.001	0.000	0.998	0.002	0.999	0.001
6	**Banca Generali**	0.001	0.999	0.000	0.000	1.000	0.000	0.001	0.999	0.000	1.000
7	**Banca Mediolanum**	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000
8	BPER Banca	0.999	0.000	0.001	0.999	0.000	0.000	1.000	0.000	1.000	0.000
9	Buzzi Unicem	0.036	0.959	0.005	0.012	0.987	0.001	0.035	0.965	0.012	0.988
10	CNH Industrial	0.999	0.001	0.000	0.997	0.003	0.000	0.999	0.001	0.997	0.003
11	Campari	1.000	0.000	0.000	0.999	0.000	0.000	1.000	0.000	1.000	0.000
12	*DiaSorin*	0.002	0.003	0.995	0.004	0.007	0.989	*trimmed*		*trimmed*	
13	Enel	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000
14	Eni	0.969	0.027	0.004	0.927	0.071	0.003	0.972	0.028	0.929	0.071
15	*Exor*	0.031	0.174	0.795	0.057	0.280	0.663	*trimmed*		*trimmed*	
16	FinecoBank	0.992	0.007	0.001	0.984	0.015	0.001	0.993	0.007	0.985	0.015
17	FIAT Chrysler (Stellantis N.V.)	0.996	0.004	0.000	0.989	0.011	0.001	0.996	0.004	0.989	0.011
18	Generali Assicurazioni	0.768	0.220	0.012	0.582	0.413	0.005	0.770	0.230	0.585	0.415
19	Hera	0.999	0.001	0.000	0.999	0.000	0.000	0.999	0.001	1.000	0.000
20	Italgas	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000
21	Infrastrutture Wireless Italiane	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000
22	Interpump Group	0.002	0.996	0.002	0.008	0.989	0.004	0.002	0.998	0.008	0.992
23	Intesa Sanpaolo	0.997	0.002	0.001	0.998	0.001	0.000	0.998	0.002	0.999	0.001
24	Leonardo	0.999	0.001	0.000	0.996	0.003	0.000	0.999	0.001	0.997	0.003
25	Mediobanca	1.000	0.000	0.000	0.999	0.001	0.000	1.000	0.000	0.999	0.001
26	Moncler	0.019	0.929	0.052	0.035	0.921	0.043	0.021	0.979	0.037	0.963
27	Pirelli	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000
28	Prysmian	0.010	0.988	0.002	0.006	0.994	0.001	0.009	0.991	0.006	0.994
29	Poste Italiane	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000
30	*Ferrari*	0.001	0.002	0.997	0.002	0.004	0.994	*trimmed*		*trimmed*	
31	Recordati	0.023	0.906	0.071	0.036	0.912	0.051	0.025	0.975	0.038	0.962
32	Saipem	0.999	0.000	0.001	1.000	0.000	0.000	1.000	0.000	1.000	0.000
33	Snam	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000
34	STMicroelectronics	0.017	0.980	0.003	0.011	0.988	0.001	0.016	0.984	0.011	0.989
35	Tenaris	0.994	0.005	0.001	0.983	0.016	0.001	0.995	0.005	0.984	0.016
36	Telecom Italia	0.992	0.004	0.004	0.996	0.003	0.001	0.996	0.004	0.997	0.003
37	Terna	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000
38	UniCredit	0.988	0.010	0.002	0.971	0.028	0.001	0.990	0.010	0.972	0.028
39	Unipol Gruppo	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000

Table 8: Membership degree for the BS-NC-FCMdC, BS-Tr-FCMC and BS-Tr-FCMdC models, with $C = 2$ (for the trimmed models $\alpha = 0.075$)

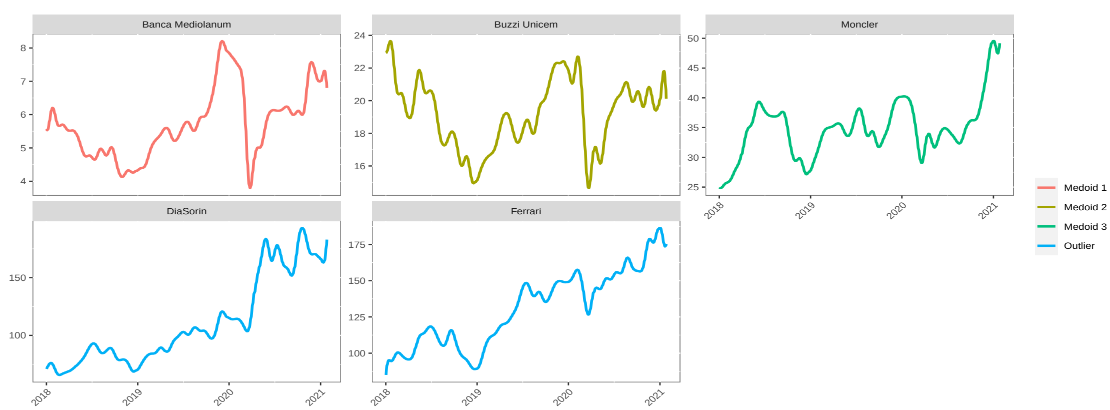
		BS-NC-FCMC				BS-NC-FCMdC				BS-Tr-FCMC			BS-Tr-FCMdC		
		Cluster 1	Cluster 2	Cluster 3	Noise Cluster	Cluster 1	Cluster 2	Cluster 3	Noise Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
1	A2A	0.994	0.004	0.000	0.002	0.995	0.004	0.000	0.001	0.995	0.004	0.000	0.996	0.004	0.000
2	Amplifon	0.006	0.987	0.005	0.002	0.022	0.935	0.040	0.003	0.009	0.961	0.031	0.022	0.937	0.040
3	Atlantia	0.022	0.970	0.005	0.003	0.008	0.987	0.004	0.001	0.019	0.968	0.013	0.008	0.988	0.004
4	Azimut Holding	0.144	0.850	0.003	0.003	0.126	0.869	0.005	0.001	0.077	0.919	0.004	0.126	0.870	0.005
5	Banco BPM	0.996	0.003	0.000	0.001	0.997	0.003	0.000	0.000	0.997	0.003	0.000	0.997	0.003	0.000
6	Banca Generali	0.001	0.996	0.002	0.001	0.005	0.980	0.015	0.001	0.003	0.975	0.022	0.005	0.981	0.015
7	**Banca Mediolanum**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
8	BPER Banca	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.999	0.001	0.000
9	**Buzzi Unicem**	0.001	0.999	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.999	0.000	0.000	1.000	0.000
10	CNH Industrial	0.995	0.004	0.000	0.001	0.990	0.009	0.000	0.000	0.993	0.006	0.000	0.991	0.009	0.000
11	Campari	0.999	0.001	0.000	0.000	0.998	0.001	0.000	0.000	0.999	0.001	0.000	0.999	0.001	0.000
12	*DiaSorin*	0.002	0.003	0.006	0.989	0.004	0.006	0.012	0.978	*trimmed*			*trimmed*		
13	Enel	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
14	Eni	0.855	0.139	0.002	0.004	0.732	0.262	0.004	0.002	0.798	0.198	0.004	0.733	0.263	0.004
15	*Exor*	0.012	0.042	0.630	0.316	0.025	0.084	0.600	0.291	*trimmed*			*trimmed*		
16	FinecoBank	0.958	0.040	0.001	0.001	0.944	0.054	0.001	0.001	0.939	0.060	0.001	0.945	0.054	0.001
17	FIAT Chrysler (Stellantis N.V.)	0.977	0.021	0.000	0.002	0.960	0.039	0.001	0.001	0.967	0.032	0.001	0.960	0.039	0.001
18	Generali Assicurazioni	0.259	0.733	0.003	0.005	0.165	0.829	0.004	0.001	0.158	0.836	0.006	0.165	0.830	0.004
19	Hera	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.999	0.001	0.000
20	Italgas	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
21	Infrastrutture Wireless Italiane	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.000	0.998	0.002	0.000	0.999	0.001	0.000
22	Interpump Group	0.016	0.691	0.277	0.016	0.009	0.268	0.719	0.004	0.003	0.083	0.914	0.009	0.269	0.722
23	Intesa Sanpaolo	0.996	0.003	0.000	0.001	0.997	0.003	0.000	0.000	0.997	0.003	0.000	0.997	0.003	0.000
24	Leonardo	0.995	0.004	0.000	0.001	0.988	0.011	0.000	0.000	0.993	0.006	0.000	0.989	0.011	0.000
25	Mediobanca	0.999	0.001	0.000	0.000	0.998	0.002	0.000	0.000	0.999	0.001	0.000	0.998	0.002	0.000
26	**Moncler**	0.001	0.013	0.983	0.003	0.000	0.000	1.000	0.000	0.000	0.002	0.997	0.000	0.000	1.000
27	Pirelli	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
28	Prysmian	0.001	0.999	0.000	0.000	0.002	0.996	0.002	0.000	0.001	0.996	0.003	0.002	0.996	0.002
29	Poste Italiane	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.999	0.001	0.000
30	*Ferrari*	0.001	0.002	0.003	0.994	0.002	0.004	0.006	0.988	*trimmed*			*trimmed*		
31	Recordati	0.001	0.012	0.984	0.003	0.002	0.020	0.975	0.003	0.000	0.005	0.995	0.002	0.020	0.978
32	Saipem	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.999	0.001	0.000
33	Snam	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
34	STMicroelectronics	0.003	0.995	0.001	0.001	0.012	0.976	0.011	0.001	0.004	0.988	0.008	0.012	0.977	0.011
35	Tenaris	0.975	0.023	0.000	0.002	0.941	0.056	0.001	0.001	0.967	0.032	0.001	0.942	0.056	0.001
36	Telecom Italia	0.990	0.006	0.000	0.004	0.992	0.006	0.001	0.001	0.992	0.007	0.001	0.993	0.006	0.001
37	Terna	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
38	UniCredit	0.944	0.053	0.001	0.002	0.888	0.109	0.002	0.001	0.920	0.077	0.002	0.889	0.109	0.002
39	Unipol Gruppo	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000

Table 9: Membership degree for the BS-NC-FCMdC, BS-Tr-FCMC and BS-Tr-FCMdC models, with $C = 3$ (for the trimmed models $\alpha = 0.075$)

provided a more detailed understanding. To this purpose, in Figure 6 (a-d), the cubic B-splines corresponding to the medoids (the centroids for the BS-NC-FCMC and BS-Tr-FCMC models) and the outliers are plotted for each method. Figure 6 (e-g), in addition, shows the crisp partition identified by the BS-NC-FCMC model, the BS-NC-FCMdC model and trimmed-based models, respectively. It is worth noting noticing again that the BS-NC-FCMC model differs only for the assignment of unit 15, no more identified as outlier, and unit 22 while the BS-NC-FCMdC one only for the assignment of unit 15.

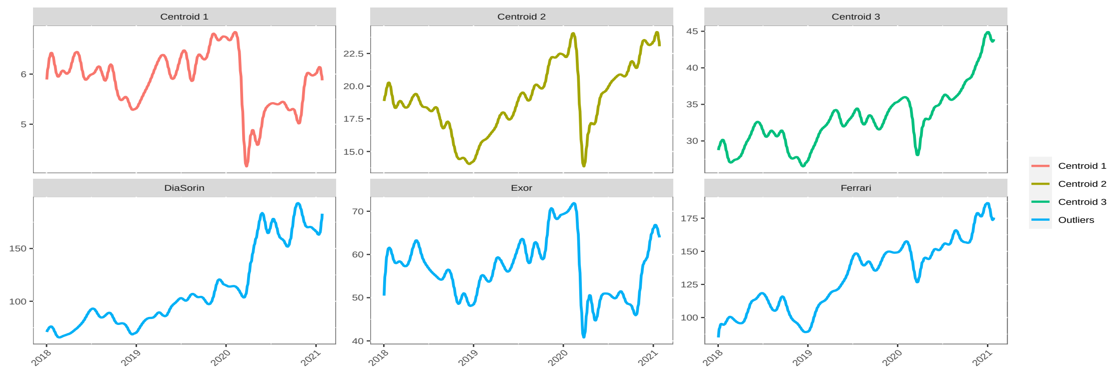


(a) BS-NC-FCMC: centroids and outlier units

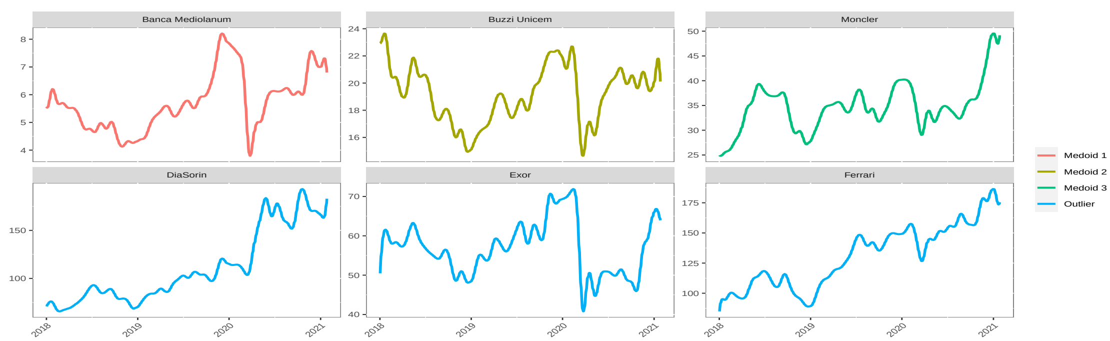


(b) BS-NC-FCMdC: medoids and outlier units

Figure 6: Main results with three clusters

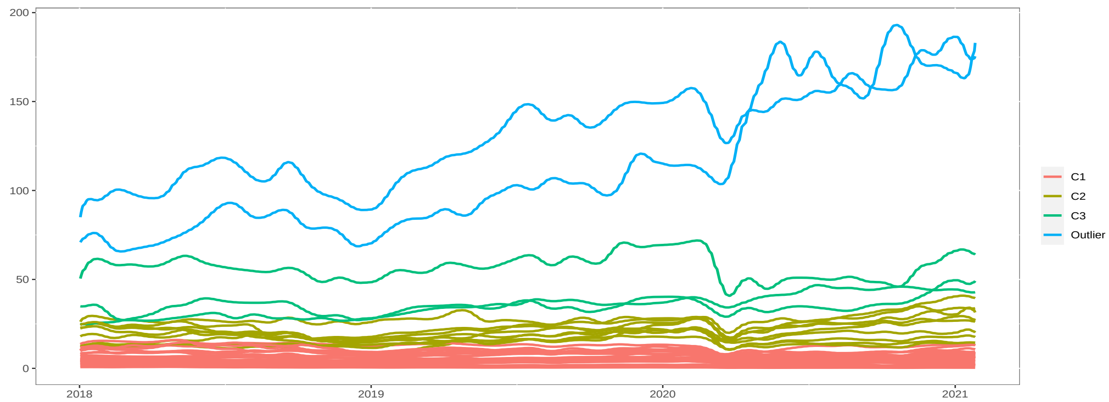


(c) BS-Tr-FCMC: centroids and outlier units

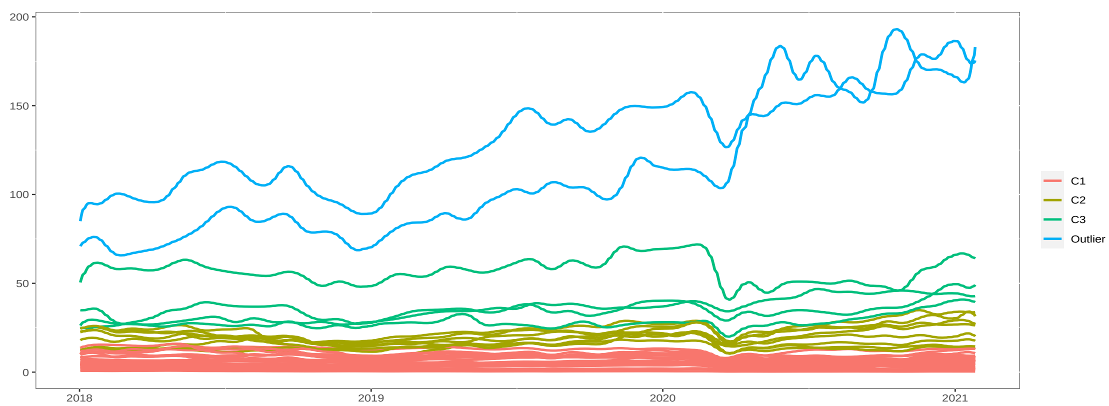


(d) BS-Tr-FCMdC: medoids and outlier units

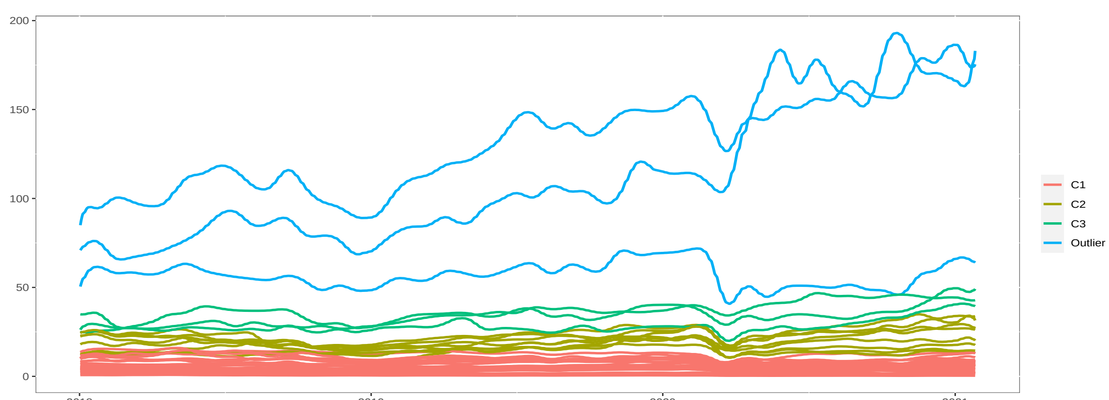
Figure 6: Main results with three clusters



(e) Crisp partition for BS-NC-FCMC



(f) Crisp partition for BS-NC-FCMdC



(g) Crisp partition common to BS-Tr-FCMC and BS-Tr-FCMdC

Figure 5: Main results with three clusters

For sake of completeness, in Appendix A, the same plots have been considered with reference to the partition in two groups[3]. As far as the partition in three groups is concerned, we notice that the medoids of the BS-NC-FCMdC and BS-Tr-FCMdC models are the same and also the centroids identified by the other two models have a perfect matching. The comparison between centroids and medoids shows a certain similarity in the pattern, even if not at all.

Focusing on the results arisen from the BS-Tr-FCMdC method, the three medoids units are *Banca Mediolanum*, *Buzzi Unicem* and *Moncler*, the first two clearly reflecting two falls, at the end of 2018 and on March 12, 2020, the latter due to COVID-19 outbreak. *Banca Mediolanum* is an Italian bank, the parent company of Mediolanum Group, *Buzzi Unicem* is an international group producing and sailing cement, ready-mix concrete and aggregates while *Moncler* is an Italian luxury fashion brand.

The three medoids differ in terms of magnitude of the adjusted closing price and *Moncler* also in terms of trend. Differently with respect to the other two medoids, the *Buzzi Unicem* has not regained the same level of the adjusted closing price recorded at the beginning of 2018. We argue that, with respect to the solution based on two groups, that based on three clusters differs only for the presence of a new cluster with medoid *Moncler*.

While the role of outlier assigned by all methods to *DiaSorin*, an Italian multinational biotechnology company, and *Ferrari*, the well-know Italian luxury sports car manufacturer, is clearly due to their evident upward trend (also during pandemic) that does not match with that of the other clusters, the role of *Exor N.V.*, the holding company controlled by the Agnelli family whose investments also include *Stellantis*, *Ferrari* and *CNH Industrial*, is controversial; in fact, the trimmed-based models identified it as an outliers, like in the solution with two clusters, while the noise-based models assigned a membership degree higher to the third cluster than to the noise one (0.630 w.r.t 0.316 for BS-NC-FCMC model, 0.6 w.r.t 0.29 for BS-NC-FCMdC model). However, it is worth highlighting that the above three stocks are those with the highest adjusted closing price with respect to the entire sample.

[3] In this case, since there were fuzzy units, they have been plotted in the graphs.

5. Conclusions and further research lines

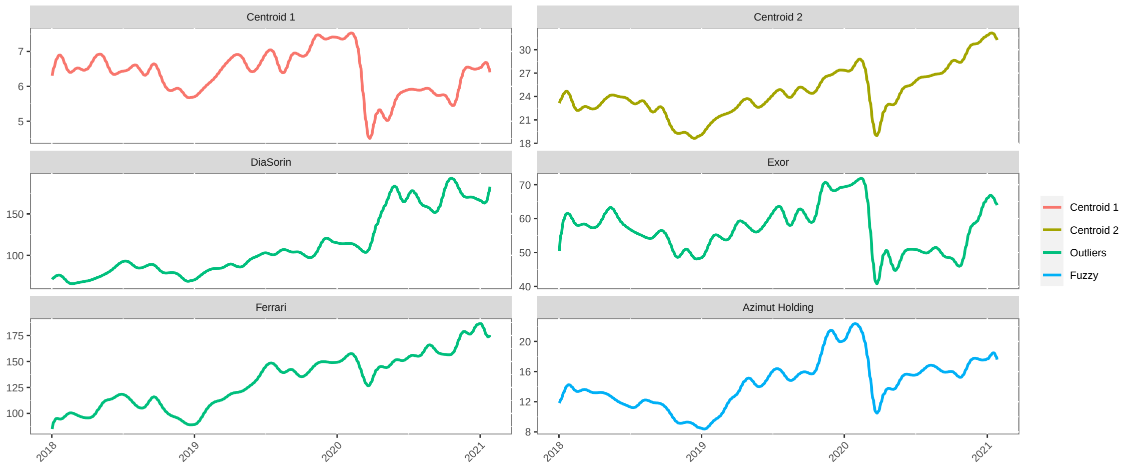
Different approaches for robust fuzzy clustering of time series have been presented and compared. In fact, two different types of robust fuzzy clustering procedures have been considered. One of them discards a fixed fraction α from the (hopefully) most outlying time series, while the other assigns them to a fictitious "noise cluster" by specifying a distance parameter δ^2. The fuzzy C-means and fuzzy C-medoids approach are also considered depending on whether "synthetic" C centroids are allowed or whether the C centroids should coincide with "observed" time series.

A simulation study and a real data example show that they are valuable methods to be considered if outlying and switching time series are expected in the analyzed data set. Unfortunately, the complete absence of outlying and switching time series in real data analysis is the exception rather than the rule, especially when it comes to the unsupervised learning problems addressed in Cluster Analysis.

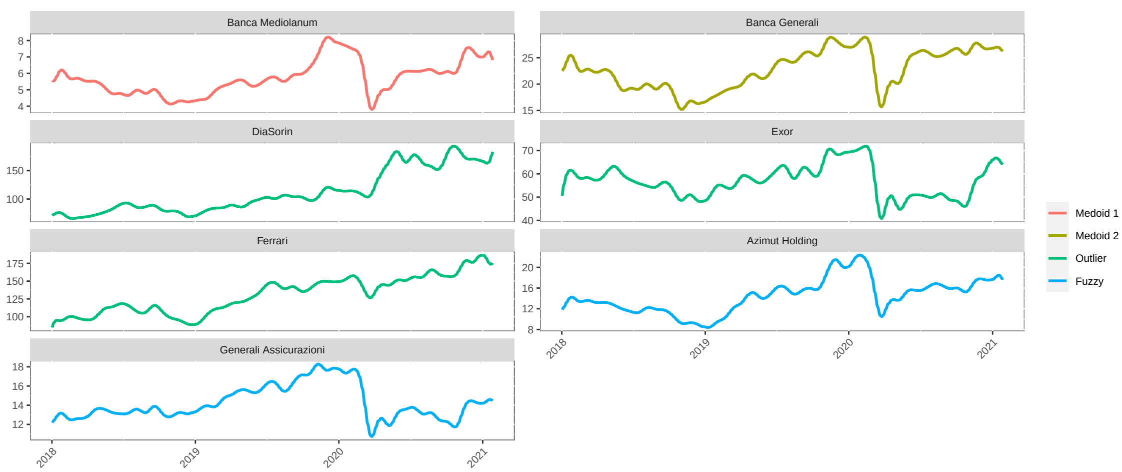
There are many interesting open lines of research that need to be considered further. For instance, more research is needed in order to develop more automatized tools for choosing all the involved parameters, as the total number of clusters C or those parameters α and δ^2 controlling the fraction of time series that are trimmed or included in the noise cluster. However, it is important to note that a completely automated selection of all of them is not expected since they usually depend heavily on the application that the user has in mind. However, as done in [34] for the crisp clustering problem, it would be nice to provide a "reduced" list of sensible fuzzy clustering partitions to help the user make the final decision.

Another interesting problem is the optimal determination of the interior knots considered in the B-splines representation. Furthermore, it is also interesting to investigate the performance of other functional bases. As already mentioned, the use of Fourier bases could be interesting with time series of a periodic nature or the use of wavelet bases could be useful to provide resolution both in frequency and in time.

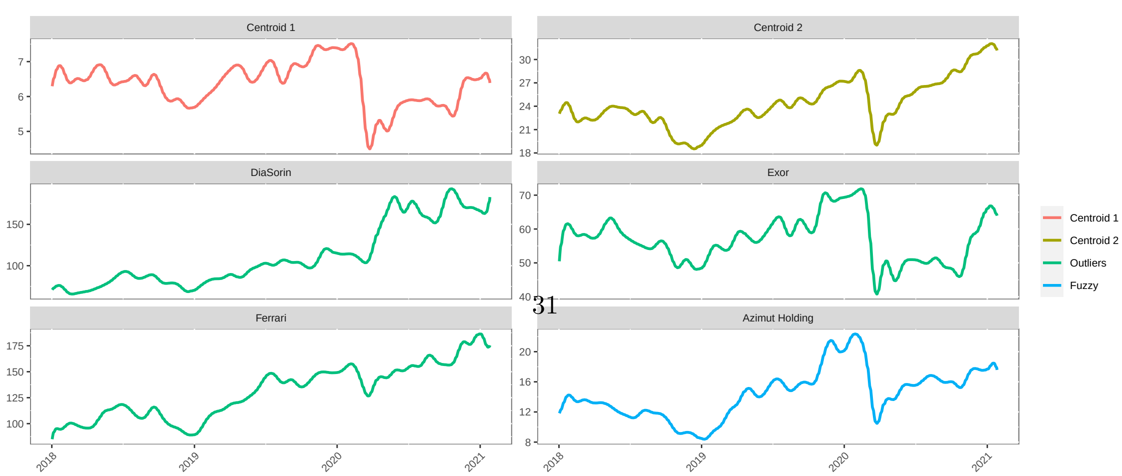
Finally, further improvements in the proposed algorithms can be investigated to provide good initialization procedures that serves to reduce the number of random starts and the number of iterations.



(a) BS-NC-FCMC: centroids, outliers and fuzzy units

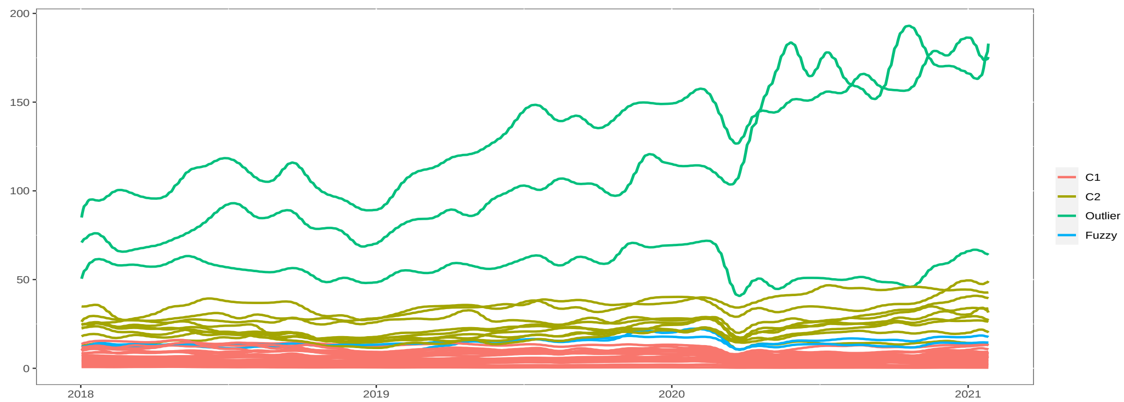


(b) BS-NC-FCMdC and BS-Tr-FCMdC: medoids, outliers and fuzzy units

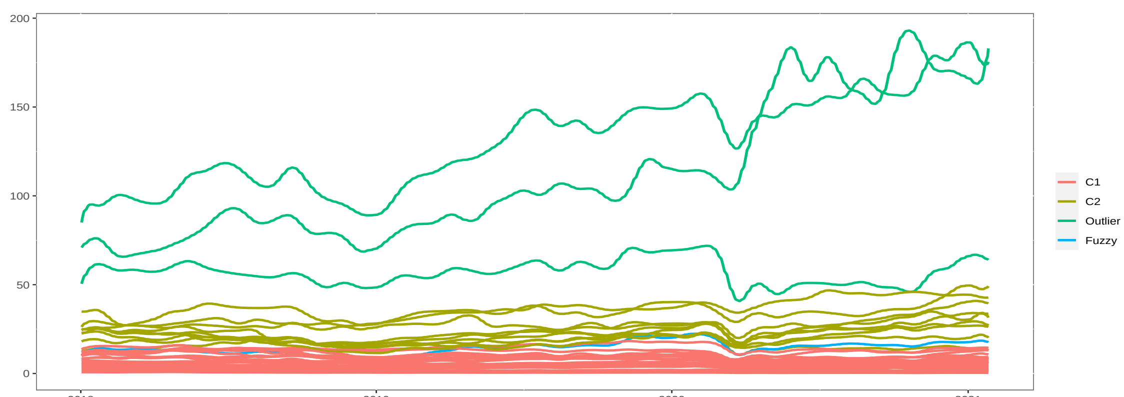


(c) BS-Tr-FCMC: centroids, outliers and fuzzy units

Figure A.6: Main results with two clusters



(d) Crisp partition and fuzzy units common to the BS-NC-FCMdC and BS-Tr-FCMdC models



(e) Crisp partition and fuzzy units common to the BS-NC-FCMC and BS-Tr-FCMC models

Figure A.6: Main results with two clusters

Appendix A. Results for two clusters

References

[1] W. T. Liao, Clustering of time series data: A survey, Pattern Recognition 38 (2005) 1857–1874.

[2] T.-C. Fu, A review on time series data mining, Engineering Applications of Artificial Intelligence 24 (2011) 164–181.

[3] J. Caiado, E. A. Maharaj, P. D'Urso, Time series clustering, in: C. Hennig, M. Meila, F. Murtagh, R. Rocci (Eds.), Handbook of Cluster Analysis, Chapman & Hall, 2014, in press.

[4] S. Aghabozorgi, A. S. Shirkhorshidi, T. Y. Wah, Time-series clustering–a decade review, Information Systems 53 (2015) 16–38.

[5] E. A. Maharaj, P. D'Urso, J. Caiado, Time series clustering and classification, CRC Press, 2019.

[6] P. Montero, J. A. Vilar, TSclust: An R package for time series clustering, Journal of Statistical Software 62(1) (2014) 43.

[7] L. A. García-Escudero, A. Gordaliza, Robustness properties of k means and trimmed k means, Journal of the American Statistical Association 94 (1999) 956–969.

[8] L. A. García-Escudero, A. Gordaliza, C. Matrán, A. Mayo-Iscar, A review of robust clustering methods, Advances in Data Analysis and Classification 4 (2010) 89–109.

[9] G. Ritter, Robust cluster analysis and variable selection, Vol. 137 of Monographs on Statistics and Applied Probability, CRC Press, Boca Raton, FL, 2015.

[10] L. A. García-Escudero, A. Gordaliza, C. Matrán, A. Mayo-Iscar, C. Hennig, Robustness and outliers, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, Taylor & Francis, 2015, Ch. 29, pp. 653–678.

[11] C. Abraham, P.-A. Cornillon, E. Matzner-Løber, N. Molinari, Unsupervised curve clustering using b-splines, Scandinavian Journal of Statistics 30 (2003) 581–595.

[12] L. A. García-Escudero, A. Gordaliza, A proposal for robust curve clustering, Journal of Classification 22 (2005) 185–201.

[13] J. A. Cuesta-Albertos, A. Gordaliza, C. Matrán, Trimmed k-means: An attempt to robustify quantizers, The Annals of Statistics 25 (1997) 553–576.

[14] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms, Plenum Press, New York, 1981.

[15] X. L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (1991) 841–847.

[16] R. J. G. B. Campello, E. R. Hruschka, A fuzzy extension of the silhouette width criterion for cluster analysis, Fuzzy Sets and Systems 157 (2006) 2858–2875.

[17] L. A. García-Escudero, A. Gordaliza, C. Matrán, Trimming tools in exploratory data analysis, Journal of Computational and Graphical Statistics 12 (2003) 434–449.

[18] J. Kim, R. Krishnapuram, R. Davé, Application of the least trimmed squares technique to prototype-based clustering, Pattern Recognition Letters 17 (1996) 633–641.

[19] R. Krishnapuram, A. Joshi, O. Nasraoui, L. Yi, Low-complexity fuzzy relational clustering algorithms for web mining, IEEE Transactions on Fuzzy Systems 9 (2001) 595–607.

[20] P. J. Rousseeuw, A. M. Leroy, Robust regression and outlier detection, John Wiley and Sons, New York, 1987.

[21] R. Maronna, Principal components and orthogonal regression based on robust scales, Technometrics 47 (2005) 264–273.

[22] H. Fritz, L. A. García-Escudero, A. Mayo-Iscar, Robust constrained fuzzy clustering, Information Sciences 245 (2013) 38–52.

[23] T. A. Runkler, J. C. Bezdek, Alternating cluster estimation: A new tool for clustering and function approximation, IEEE Transactions on Fuzzy Systems 7 (1999) 377–393.

[24] R. N. Davé, Characterization and detection of noise in clustering, Pattern Recognition Letters 12 (1991) 657–664.

[25] Y. Ohashi, Fuzzy clustering and robust estimation, in: Ninth Meeting of SAS Users Group Int., 1984.

[26] R. N. Davé, S. Sen, Noise clustering algorithm revisited, in: Fuzzy Information Processing Society, 1997 Annual Meeting of the North American, NAFIPS'97, IEEE, 1997, pp. 199–204.

[27] P. D'Urso, L. De Giovanni, R. Massari, Trimmed fuzzy clustering of financial time series based on dynamic time warping, Annals of Operations Research (2019). doi:10.1007/s10479-019-03284-1.

[28] P. D'Urso, L. De Giovanni, R. Massari, Robust fuzzy clustering of multivariate time trajectories, International Journal of Approximate Reasoning 99 (2018) 12 – 38.

[29] E. A. Maharaj, P. D'Urso, Fuzzy clustering of time series in the frequency domain, Information Sciences 181 (2011) 1187 – 1211.

[30] R. J. Campello, E. R. Hruschka, A fuzzy extension of the silhouette width criterion for cluster analysis, Fuzzy Sets and Systems 157 (2006) 2858–2875.

[31] S.-L. Chang, C.-Y. Chien, H.-C. Lee, C. Lin, Historical high and stock index returns: Application of the regression kink model, Journal of International Financial Markets, Institutions and Money 52 (C) (2018) 48–63.

[32] M. Rechenthin, W. N. Street, P. Srinivasan, Stock chatter: Using stock sentiment to predict price direction, Algorithmic Finance 2 (3-4) (2013) 169–196.

[33] P. D'Urso, R. Massari, C. Cappelli, L. De Giovanni, Autoregressive metric-based trimmed fuzzy clustering with an application to pm10 time series, Chemometrics and Intelligent Laboratory Systems 161 (2017) 15–26.

[34] A. Cerioli, L. A. García-Escudero, A. Mayo-Iscar, M. Riani, Finding the number of normal groups in model-based clustering via constrained likelihoods, Journal of Computational and Graphical Statistics 27 (2018) 404–416.