ORIGINAL PAPER

Fuzzy clustering of time series based on weighted conditional higher moments

Roy Cerqueti[1,2] · Pierpaolo D'Urso[1] · Livia De Giovanni[3] · Raffaele Mattera[1] · Vincenzina Vitale[1]

Abstract

This paper proposes a new approach to fuzzy clustering of time series based on the dissimilarity among conditional higher moments. A system of weights accounts for the relevance of each conditional moment in defining the clusters. Robustness against outliers is also considered by extending the above clustering method using a suitable exponential transformation of the distance measure defined on the conditional higher moments. To show the usefulness of the proposed approach, we provide a study with simulated data and an empirical application to the time series of stocks included in the FTSEMIB 30 Index.

Keywords Dynamic conditional score · Unsupervised learning · Robust clustering · Fuzzy clustering · Conditional moments · Exponential dissimilarity · Financial time series

✉ Raffaele Mattera
  raffaele.mattera@uniroma1.it

  Roy Cerqueti
  roy.cerqueti@uniroma1.it

  Pierpaolo D'Urso
  pierpaolo.durso@uniroma1.it

  Livia De Giovanni
  ldegiovanni@luiss.it

  Vincenzina Vitale
  vincenzina.vitale@uniroma1.it

1 Department of Social and Economic Sciences, Sapienza University of Rome, Rome, Italy

2 GRANEM, University of Angers, Angers, France

3 Department of Political Sciences, LUISS Guido Carli, Rome, Italy

1 Introduction

Time series clustering methods extend the algorithms for static data in order to take into account the time-varying structure. One of the main issues lies in finding an appropriate dissimilarity measure among time series, given that it is well known that considering a simple Euclidean distance based on raw data is not always the best solution to adopt (Liao 2005; Maharaj et al. 2019).

In general, two main approaches can be identified as alternatives to raw data-based clustering. On one side, we have the feature-based approach that uses time series features instead of raw data for clustering. Examples of interesting features are, among others, the correlation (Mantegna 1999), the distributional characteristics (Bastos and Caiado 2021; Wang et al. 2006), the auto-correlation structure (Alonso and Maharaj 2006; D'Urso and Maharaj 2009), the quantile auto-covariances (Lafuente-Rego and Vilar 2016), the peridiogram ordinates (Caiado et al. 2006, 2020), the cepstral coefficients (D'Urso et al. 2020; Savvides et al. 2008), and the wavelet domain features (Maharaj et al. 2010). On the other side, we have the so-called model-based approach that clusters units based on different model parameters. Examples of statistical models are the ARMA (Maharaj 1996, 2000; Piccolo 1990) or the GARCH processes (Caiado and Crato 2010; D'Urso et al. 2016; Otranto 2008) but also specific probability distributions (Mattera et al. 2021) and splines (Iorio et al. 2016).

Despite the great popularity of conditional variance-based clustering (i.e. GARCH-based approaches), there is abundant literature documenting that time variation characterizes not only the variance but also higher moments (Harvey and Siddique 1999; Jondeau and Rockinger 2012; León et al. 2005). Recently, Cerqueti et al. (2021) proposed a fuzzy *C*-Means clustering approach based on conditional higher moments estimated with the Dynamic Conditional Score model due to Creal et al. (2013). The fuzzy approach is introduced because the identification of a clear boundary between clusters is not easy in many real problems and, in general, soft clustering approaches like the fuzzy *C*-Means (Bezdek 1981) are usually considered for dealing with uncertainty.

We contribute to the previous literature by extending the clustering procedure proposed in Cerqueti et al. (2021) in two ways. First of all, we propose a more general clustering method by introducing a weighting scheme among the conditional moments. Instead of considering a specific conditional moment as in Cerqueti et al. (2021) (i.e. adopting a targeting approach), we optimally combine several conditional higher moments within the clustering algorithm. Second, we make the proposed clustering method more robust against outliers. Notice that, in this paper, we consider as outliers the full time series whose conditional moments' auto-correlation structure is very different from the others, rather than outlying time points in a single time series.

As far as robustness is concerned, it is well known that standard clustering methods may miss to correctly identify the clustering structure when a certain amount of outliers occurs in the data. This fact motivated several authors (e.g. see Wu and Yang 2002; Garcia-Escudero and Gordaliza 2005; D'Urso et al. 2016,

2022) in developing alternative clustering procedures providing a better clustering quality in presence of outliers. In this paper, we follow the approach of Wu and Yang (2002), and use a particular exponential transformation of the distance as a building block of a robust clustering procedure.

To show the usefulness of the methodological proposal, we provide both a simulation study and an application to a real dataset of financial time series. The simulation results confirm that the clustering approaches based on conditional moments provide higher clustering performances than those based on raw time series. Moreover, simulations also show that the clustering approach based on the weighting scheme improves with respect to an approach based on targeting. Most importantly, the robust clustering procedure developed in this paper is more effective in clustering time series in the presence of outliers. The analysis with real data confirms the simulations' evidence.

The paper is structured as follows. In Sect. 2, we introduce the exponential dissimilarity and the Dynamic Conditional Score while, in Sect. 3, we explain the proposed robust clustering approach in detail. Section 4 contains the simulation results and Sect. 5 the application to the financial time series data. In the last section, we highlight the advantages of the new procedure and offer some conclusive remarks.

2 Statistical tools

In what follows, we discuss the main tools required for the implementation of the proposed clustering procedures. In Sect. 2.1 we discuss the estimation of conditional higher moments while, in Sect. 2.2, we present the weighted dissimilarity used for clustering time series by considering many conditional higher moments jointly. The exponential transformation of the proposed weighted dissimilarity, used as a tool for building a robust clustering procedure, is also discussed in Sect. 2.2.

2.1 Conditional moments

There are several ways to estimate time series' conditional moments (for a survey see Soltyk and Chan 2021), like those discussed in Harvey and Siddique (1999) and León et al. (2005). However, the Dynamic Conditional Score (DCS, also called Generalized Autoregressive Score) proposed by Creal et al. (2013) is more recent and constitutes a very general approach. The DCS considers the score function of the predictive model density as the driving mechanism for time-varying parameters and can be formalized as follows.

Let $\mathbf{y}$ be a time series of length T so that the value of the time series at time t, y_t, is generated by the following conditional density $p(\cdot)$:

$$y_t \sim p(y_t|\mathbf{f}_t, \mathcal{F}_t; \boldsymbol{\theta}), \quad t = 1, \dots, T \tag{1}$$

where $\mathbf{f}_t$ is a vector of length K containing K time-varying parameters (i.e. parameters of the conditional distribution) at time t, $\mathcal{F}_t$ is the available information at time t and $\boldsymbol{\theta}$ a vector of static parameters. The length of the vector $\mathbf{f}_t$ depends on the assumption we make about the density (1) so that, for example, in case of the

Gaussian distribution, $K = 2$. By assuming different densities we obtain different K specifications.

The available information set at time t, $\mathcal{F}_t$, is given by the previous realizations of $\mathbf{y}$ and by its time varying parameters. Hence, it is possible to express the DCS of order 1, i.e. DCS(1, 1), for the t-th realization $\mathbf{f}_t$ of the time-varying parameter vector as follows:

$$\mathbf{f}_t = \boldsymbol{\omega} + \mathbf{A}_1 \mathbf{s}_{t-1} + \mathbf{B}_1 \mathbf{f}_{t-1}, \tag{2}$$

where $\boldsymbol{\omega}$ is a real vector and the $\mathbf{A}_1$ and the $\mathbf{B}_1$ are diagonal real matrices whose dimension depends on the number of the time varying parameters. The variable $\mathbf{s}_t$ in (2) is the *scaled* score of the conditional distribution (1) at time t, where the score of the density, called $\boldsymbol{\nabla}_t$:

$$\boldsymbol{\nabla}_t = \frac{\partial \log p(y_t|\mathbf{f}_t, \mathcal{F}_t; \boldsymbol{\theta})}{\partial \mathbf{f}_t}, \tag{3}$$

is scaled by means of a positive definite matrix $\mathbf{S}_t$ such that:

$$\mathbf{s}_t = \mathbf{S}_t \cdot \boldsymbol{\nabla}_t. \tag{4}$$

A common approach of scaling, proposed by Creal et al. (2013), is to consider the inverse of the information matrix of $\mathbf{f}_t$ to a power $\gamma \geq 0$:

$$\mathbf{S}_t = E_{t-1}\left[\boldsymbol{\nabla}_t \boldsymbol{\nabla}_t'\right]^{-\gamma}, \tag{5}$$

where E_{t-1} denotes the expectation at time $t-1$ and the conditional score $\boldsymbol{\nabla}_t$ is defined as in (3). The γ parameter usually takes value $\{0, \frac{1}{2}, 1\}$. When $\gamma = 0$, then $\mathbf{S}_t = I$ is the identity matrix and there is no scaling. Differently, if $\gamma = 1$, then the conditional score $\boldsymbol{\nabla}_t$ is pre-multiplied by the inverse of the Fisher information matrix, while for $\gamma = \frac{1}{2}$ the score $\boldsymbol{\nabla}_t$ is scaled to the square-root of the Fisher information matrix. An appealing feature of the DCS model is that the vector of parameters $\boldsymbol{\theta}$ can be estimated by maximum likelihood (Creal et al. 2013).

According to the model specification (2), we can obtain the time series $\mathbf{f}_t$'s by in sample predictions as follows (Cerqueti et al. 2021):

$$\hat{\mathbf{f}}_t = \hat{w} + \hat{\mathbf{A}}_1 \hat{\mathbf{s}}_{t-1} + \hat{\mathbf{B}}_1 \hat{\mathbf{f}}_{t-1}, \tag{6}$$

where $\hat{\mathbf{f}}_t$ is the vector of the *estimated time-varying parameters* at time t. In the case of the Gaussian distribution:

$$p(y_t|\mathbf{f}_t, \mathcal{F}_t; \boldsymbol{\theta}) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{-(y_t - \mu_t)^2 / 2\sigma_t^2}, \tag{7}$$

we have that the vector of parameters $\mathbf{f}_t = [\mu_t, \sigma_t^2]$ coincides with the first two moments of the distribution. Therefore, in this case, we refer to $\mathbf{f}_t$ as the vector of conditional moments at time t. When the parameters do not coincide with the moments of the distribution, we can obtain the latter by simply replacing the

estimated parameters in the corresponding moments' formula. Let us consider the case of the generalized-t distribution with density:

$$p(y_t|\mathbf{f}_t, \mathcal{F}_t; \boldsymbol{\theta}) = \frac{\Gamma\left(\frac{v_t+1}{2}\right)}{\Gamma\left(\frac{v_t}{2}\right)\sqrt{(v_t-2)\pi\sigma_t^2}}\left(1 + \frac{(y_t - \mu_t)^2}{(v_t-2)\sigma_t^2}\right)^{-\frac{v_t+1}{2}}, \qquad (8)$$

where the vector of the parameters is $\mathbf{f}_t = [\mu_t, \sigma_t^2, v_t]$. Under this distributional assumption, we have that the moments are functions of the parameters. In particular, the first moment is equal to $f_{1,t} = \mu_t$, the second moment is $f_{2,t} = \sigma_t^2 v_t/(v_t - 2)$ and the fourth moment equals $f_{4,t} = 6/(v_t - 4) + 3$ (Jackman 2009). The third moment $f_{3,t}$ – the skewness – is equal to zero $\forall t$ as the distribution is symmetric. The moments can be estimated by replacing the in-sample predictions of the parameters (6) in the corresponding moment formula, i.e. $\hat{f}_{1,t} = \hat{\mu}_t$, $\hat{f}_{2,t} = \hat{\sigma}_t^2 \hat{v}_t/(\hat{v}_t - 2)$ and $f_{4,t} = 6/(\hat{v}_t - 4) + 3$. To avoid abuse of notation, we call the vector of conditional moments $\mathbf{f}_t = [f_{1,t}, f_{2,t}, f_{4,t}]$ and indicate by $\hat{\mathbf{f}}_t = [\hat{f}_{1,t}, \hat{f}_{2,t}, \hat{f}_{4,t}]$ the corresponding estimate. Then, we can then generally consider the following matrix representation of the conditional moments:

$$\hat{\mathbf{F}} = \begin{bmatrix} \hat{f}_{11} & \cdots & \hat{f}_{k1} & \cdots & \hat{f}_{K1} \\ \vdots & \cdots & \hat{f}_{kt} & \cdots & \vdots \\ \hat{f}_{1T} & \cdots & \hat{f}_{kT} & \cdots & \hat{f}_{KT} \end{bmatrix}, \qquad (9)$$

where the element in the t-th row and k-th column is the k-th conditional moment at time t of the time series $\mathbf{y}$. In other words, we can express the time series in terms of a matrix of K conditional moments, that are time series themselves. Since the conditional moments are obtained by the in-sample predictions, they have the same length T of the underlying original time series.

2.2 Dissimilarity measure

As briefly discussed in the introduction, there are several ways of measuring the dissimilarity among time series. Usually, scholars use Euclidean distances computed based on data features (feature-based approaches) or parameters estimated from statistical models (model-based approaches).

In this paper we propose to compute the dissimilarity among time series considering the auto-correlation function (ACF) structure of their estimated conditional higher moments. In particular, in this paper we measure the dissimilarity between two time series as the weighted dissimilarity among their conditional higher moments. The ACF-based distance is commonly used for clustering time series (Caiado et al. 2006; D'Urso and Maharaj 2009). Since the ACF is bounded between − 1 and 1, the ACF-based dissimilarity is free from the problem of the measurement unit. Moreover, it is also well suited for handling conditional higher moments jointly.

Let us consider two time series $\mathbf{y}_i = (y_{i,t} : t = 1, \ldots, T)$ and $\mathbf{y}_j = (y_{j,t} : t = 1, \ldots, T)$, for which we assume a specific probability distribution and estimate K conditional higher moments stored in the matrices $\hat{\mathbf{F}}_i$ and $\hat{\mathbf{F}}_j$, respectively. Let us define $\hat{\rho}_{l,i}^{(k)}$ as the estimated auto-correlation at lag l for the conditional k-th moment of a given i-th time series, obtained with the usual estimator:

$$\hat{\rho}_{l,i}^{(k)} = \frac{\sum_{t=l+1}^{T} \left(\hat{f}_{i,t}^{(k)} - \bar{\hat{f}}_i^{(k)} \right) \left(\hat{f}_{i,t-l}^{(k)} - \bar{\hat{f}}_i^{(k)} \right)}{\sum_{t=1}^{T} \left(\hat{f}_{i,t}^{(k)} - \bar{\hat{f}}_i^{(k)} \right)^2}, \qquad (10)$$

where $\bar{\hat{f}}_i^{(k)}$ is the mean of the estimated k-th conditional moment of the i-th time series, i.e. the k-th column of the matrix $\hat{\mathbf{F}}_i$ as shown in (9). Notice that $\hat{\rho}_{l,i}^{(k)} \in [-1, 1]$ by construction. The squared distance based on weighted conditional higher moments between the two time series $\mathbf{y}_i$ and $\mathbf{y}_j$ can be computed as follows:

$$d^2(\mathbf{y}_i, \mathbf{y}_j) = \sum_{k=1}^{K} \left(w_k d_{i,j}^{(k)} \right)^2 = \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,j}^{(k)} \right)^2, \qquad (11)$$

where $d_{i,j}^{(k)} = \sqrt{\sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,j}^{(k)} \right)^2}$ being the ACF-based Euclidean distance between the time series i and j according to the k-th conditional moment. In this way, we account for the auto-correlation structure of the conditional moments' processes, hence being in line with the fuzzy clustering approach proposed in Cerqueti et al. (2021) and D'Urso and Maharaj (2009).

The k-th weight, w_k, in (11), reflects the relevance of the k-th conditional moment in defining the clusters and, as explained in the next section, it is identified endogenously in the clustering procedure.

We should notice, however, that standard clustering procedures may not be robust enough when outliers are present. Several alternative approaches have been proposed to overcome this issue, such as the use of trimming (García-Escudero and Gordaliza 1999; Garcia-Escudero and Gordaliza 2005) or the use of a procedure building upon a suitable exponential transformation of the dissimilarity (Wu and Yang 2002; Yang and Wu 2004; D'Urso et al. 2016). Following Wu and Yang (2002), we propose a clustering approach that uses the following exponential transformation of (11) as the building block of a robust procedure:

$$d_{\exp}^2(\mathbf{y}_i, \mathbf{y}_j) = 1 - \exp\left[-\beta d^2(\mathbf{y}_i, \mathbf{y}_j) \right], \qquad (12)$$

where β is a positive constant and $d^2(\mathbf{y}_i, \mathbf{y}_j)$ is the weighted squared distance (11). As shown by Wu and Yang (2002), the use of this dissimilarity in a clustering procedure improves its robustness properties.

To use (12) in practice, we have to specify the value of the parameter β. With this respect, we have to consider that if $\beta \to \infty$, any two statistical units have huge distances and are very different from each other, meaning that each unit has no neighbouring unit in the Euclidean space. Conversely, if $\beta \to 0$, the statistical units have the same distances, so all the units are neighbours in Euclidean space. None of these

solutions is appropriate. Following D'Urso et al. (2016), we determine the value of β as the inverse of a measure of data variability:

$$\beta = \left[\frac{\sum_{i=1}^{N} d^2\left(\mathbf{y}_i, \mathbf{y}_q\right)}{N} \right]^{-1} \quad \text{with} \quad q = \operatorname*{argmin}_{1<i<N} \sum_{j=1}^{N} d^2\left(\mathbf{y}_i, \mathbf{y}_j\right), \qquad (13)$$

where $\mathbf{y}_q$, based on (13), is the time series closest to all the others. The role of (12) is to smooth the dissimilarity in such a way that large values are penalized compared to low ones (see Wu and Yang 2002). When used in a fuzzy clustering procedure, we also get the effect of smoothing the membership degree in such a way that the outlying statistical units belong to all the C clusters with the same membership degree, i.e. they have membership degrees equal (or close) to $1/C$ (see D'Urso et al. 2016).

3 Fuzzy clustering methods

Following a Fuzzy C-Medoids (FCMd) approach, we introduce two main novelties with respect to the previous literature. First, we introduce a weighting scheme for the conditional moments and let the weights be computed optimally within the clustering algorithm. As explained previously, the weights reflect the relevance of each k-th conditional moment in defining the clusters. By letting the algorithm compute the weights, we implicitly assume that the relevance of each conditional moment is unknown ex-ante, and we let the data decide their relevance in clustering.

This approach, called *Weighted Conditional Moments-based Fuzzy C-Medoids* clustering method (WCM-FCMd), is based on the assumption that clustering according to a single conditional moment can be potentially less informative than considering several moments together. Then, we provide a robust version of this weighted method, called *Exponential Weighted Conditional Moments-based Fuzzy C-Medoids* clustering method (Exp-WCM-FCMd), by building a robust procedure based on the exponential transformation of the weighted distance based on the estimated auto-correlations, as explained in the previous section.

We use a Fuzzy C-Medoids (FCMd, Krishnapuram et al. 2001) approach because it is less sensitive to outliers than the usual fuzzy C-Means (Garcia-Escudero and Gordaliza 2005). This happens because the C-Means uses an average—which is notoriously not robust to outliers—as a cluster centroid while the first uses a real time series as prototype. In detail, in the Fuzzy C-Medoids approach, the prototypes of each group, also called the *medoids*, are time series that belong to the sample and are not virtual ones as for the C-Means. Moreover, the possibility of obtaining non-fictitious representative time series in the clusters, as the medoids, is also very appealing because it improves the interpretability of groups.

3.1 Weighted conditional moments-based fuzzy C-medoids clustering method

In the real world, one can be interested in forming clusters according to a specific moment of the distribution. For example, in the case of financial time series, one can be interested in group time series with similar heteroskedastic behaviour (i.e. similar conditional variance) or with similar conditional skewness. For example, the relevance of time varying skewness and kurtosis in finance is a well-acknowledged empirical evidence (e.g. see Harvey et al. 2010; Jondeau and Rockinger 2012).

Nevertheless, if there is no specific interest in a given parameter, the clustering algorithm can benefit from considering all the estimated conditional moments. In doing so, we propose to assign an optimal weight to each moment instead of considering each of them separately.

The resulting fuzzy clustering method can be formalized by minimizing the following objective function:

$$\min : \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^{m} \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2 \qquad (14)$$

under the constraints:

$$\sum_{c=1}^{C} u_{i,c} = 1, \quad u_{i,c} \geq 0, \qquad (15)$$

$$\sum_{k=1}^{K} w_k = 1, \quad w_k \geq 0, \qquad (16)$$

where $u_{i,c}$ denotes the membership degree of the i-th time series to the c-th cluster, the parameter $m > 1$ controls for the fuzziness of the partition, w_k is the weight associated to the k-th moment and is proportional to its relevance in determining the clusters, $\hat{\rho}_{l,i}^{(k)}$ is the k-th conditional moments' autocorrelation at l-th lag and $\hat{\rho}_{l,c}^{(k)}$ represents the autocorrelation of the c-th medoid. Note that both C and m are values fixed by the user.

Theorem 1 *The optimal solution of the problem* (14) *with constraints in* (15) *and* (16) *satisfies*:

$$u_{i,c} = \frac{1}{\sum_{c'=1}^{C} \left[\frac{\sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2}{\sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c'}^{(k)} \right)^2} \right]^{1/(m-1)}}, \quad \forall c = 1, \dots, C \qquad (17)$$

$$w_k = \cfrac{1}{\sum_{k=1}^{K} \left[\cfrac{\sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2}{\sum_{i=1}^{N} \sum_{c'=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c'}^{(k)} \right)^2} \right]}, \quad \forall k = 1, \dots, K \tag{18}$$

Proof The proof adopts the sequential procedure proposed in D'Urso and Massari (2019). First, we prove (17); then, we check the validity of (18).

Let us consider the following Lagrangian function:

$$\mathcal{L}\left(u_{1,1}, \dots, u_{i,c}, \dots, u_{N,C}, w_1, \dots, w_k, \dots, w_K, \lambda, \eta \right) = \mathcal{L}(\cdot)$$
$$= \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2 - \lambda \left(\sum_{c=1}^{C} u_{i,c}^m - 1 \right) - \eta \left(\sum_{k=1}^{K} w_k - 1 \right)$$

The first order conditions are given by:

$$\begin{cases} \frac{\partial \mathcal{L}(\cdot)}{\partial u_{i,c}} = m \cdot u_{i,c}^{m-1} \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2 - \lambda = 0 \\ \frac{\partial \mathcal{L}(\cdot)}{\partial \lambda} = \sum_{c=1}^{C} u_{i,c}^m - 1 = 0 \\ \frac{\partial \mathcal{L}(\cdot)}{\partial w_k} = 2w_k \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2 - \eta = 0 \\ \frac{\partial \mathcal{L}(\cdot)}{\partial \eta} = \sum_{k=1}^{K} w_k - 1 = 0 \end{cases}$$

The membership degrees satisfying the first order conditions are:

$$u_{i,c} = \left(\cfrac{\lambda}{m \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2} \right)^{1/(m-1)}. \tag{19}$$

It is possible to show that:

$$\sum_{c=1}^{C} u_{i,c} = \sum_{c=1}^{C} \left(\cfrac{\lambda}{m \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2} \right)^{1/(m-1)}.$$

From the constraint (15) it follows:

$$1 = \frac{\lambda}{m} \sum_{c=1}^{C} \left(\frac{1}{\sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2} \right)^{1/(m-1)}$$

$$\frac{\lambda}{m} = \frac{1}{\sum_{c=1}^{C} \left(\frac{1}{\sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2} \right)^{1/(m-1)}}. \tag{20}$$

By substituting (20) in (19) we get (17). Then, the weights satisfying the first order conditions are equal to:

$$w_k = \frac{\eta}{2 \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2}. \tag{21}$$

We can show that:

$$\sum_{k=1}^{K} w_k = \sum_{k=1}^{K} \left(\frac{\eta}{2 \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2} \right). \tag{22}$$

From the constraint (16) it follows:

$$1 = \frac{\eta}{2} \sum_{k=1}^{K} \left(\frac{1}{\sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2} \right)$$

$$\frac{\eta}{2} = \frac{1}{\sum_{k=1}^{K} \left(\frac{1}{\sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{l,i}^{(k)} - \hat{\rho}_{l,c}^{(k)} \right)^2} \right)}. \tag{23}$$

By substituting (23) in (21) we get (18).   □

The solutions (17) and (18) have to be found iteratively by adopting a strategy based on Fu's heuristic approach (e.g. see D'Urso et al. 2020; D'Urso and Massari 2019). The algorithm for the implementation of the Weighted Conditional Moments Fuzzy *C*-Medoids (WCM-FCMd) clustering method is shown in the Algorithm 1.

In the paper we initialize the weights according to an equal weighting scheme, i.e. $w_k = 1/K$, while the membership degrees are initialized with random sampling from a Uniform distribution $U[0, 1]$ that are then scaled to fit with the constraint $\sum_{c=1}^{C} u_{i,c} = 1$. However, alternative initialization schemes can be easily considered.

Estimate time-varying parameters and the conditional moments.
Estimate the auto-correlation terms as defined in (10).
Fix the number of clusters C, the maximum iterations $max.iter$ and the fuzzifier m.
Initialize membership degrees $u_{i,c}$ and weights w_k.
Set iter = 0.
Randomly pick the initial medoids $\tilde{C}$.
repeat
    Store the current medoids $\tilde{C}_{old} = \tilde{C}$;
    Update the weights w_k with (18).
    Update the membership degrees $u_{i,c}$ with (17).
    Select the new medoids:
    for $c = 1$ to C **do**:

$$q = \operatorname{argmin}_{1 \leqslant i \leqslant N} \sum_{i'=1}^{N} u_{i'c}^m \left[\sum_{k=1}^{K} w_k \sum_{l=1}^{L} \left(\hat{\rho}_{l,i'}^{(k)} - \hat{\rho}_{l,i}^{(k)} \right)^2 \right]$$

        return $\tilde{C}_c = c_q$
    end for
    $iter \leftarrow iter + 1$.
until $\tilde{C}_{old} = \tilde{C}$ or $iter = max.iter$.

Algorithm 1 WCM-FCMd clustering

Remark It is worth stressing that the proposed approach is a generalization of the clustering method presented in Cerqueti et al. (2021), where one implicitly assigns a weight equal to one to the target conditional moment of interest and weights equal to zero to all the others.

3.2 Exponential weighted conditional moments-based fuzzy C-medoids clustering method

The weighted clustering method (WCM-FCMd) presented in previous section is not robust against outliers. Based on the exponential transformation (12), the *Exponential Weighted Conditional Moments-based Fuzzy C-Medoids* (Exp-WCM-FCMd) clustering method can be formalized by minimizing the following objective function:

$$\min : \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \left[1 - \exp \left(-\beta \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{i,l}^{(k)} - \hat{\rho}_{c,l}^{(k)} \right)^2 \right) \right] \tag{24}$$

under the constraints:

$$\sum_{c=1}^{C} u_{i,c} = 1, \quad u_{i,c} \geq 0 \tag{25}$$

$$\sum_{k=1}^{K} w_k = 1, \quad w_k \geq 0 \tag{26}$$

Theorem 2 *The optimal solution of* (24) *with constraints* (25) *and* (26) *is given by*:

$$u_{i,c} = \frac{1}{\sum_{c'=1}^{C} \left\{ \dfrac{\left[1-\exp\left(-\beta \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{i,l}^{(k)}-\hat{\rho}_{c,l}^{(k)}\right)^2\right)\right]}{\left[1-\exp\left(-\beta \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{i,l}^{(k)}-\hat{\rho}_{c',l}^{(k)}\right)^2\right)\right]} \right\}^{1/(m-1)}}, \quad \forall c = 1, \dots, C \tag{27}$$

$$w_k = \frac{1}{\sum_{k'=1}^{K} \left[\dfrac{\sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{i,l}^{(k)}-\hat{\rho}_{c,l}^{(k)}\right)^2 \exp\left\{-\beta \sum_{k'=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{i,l}^{(k)}-\hat{\rho}_{c,l}^{(k)}\right)^2\right\}}{\sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{l=1}^{L} \left(\hat{\rho}_{i,l}^{(k)}-\hat{\rho}_{c',l}^{(k)}\right)^2 \exp\left\{-\beta \sum_{k'=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{i,l}^{(k)}-\hat{\rho}_{c',l}^{(k)}\right)^2\right\}} \right]}, \quad \forall k = 1, \dots, K \tag{28}$$

The strategy for the proof, in this case, follows the one used for Theorem 1. The proof is shown in the supplementary material.

Also in this case the solutions (27) and (28) have to be found iteratively by adopting a strategy based on Fu's heuristic approach. The algorithm for the implementation of the Exp-WCM-FCMd clustering method is shown in the following.

Estimate time-varying parameters and the conditional moments.
Estimate the auto-correlations terms defined in (10).
Fix the number of clusters C, the maximum iterations $max.iter$ and the fuzzifier m.
Initialize membership degrees $u_{i,c}$ and weights w_k.
Compute β according to (13).
Set $iter = 0$.
Randomly pick the initial medoids $\tilde{C}$.
repeat
    Store the current medoids $\tilde{C}_{old} = \tilde{C}$.
    Update the weights w_k with (28).
    Update the membership degrees $u_{i,c}$ with (27).
    Select the new medoids:
    for $c = 1$ to C **do**:

$$q = \operatorname{argmin}_{1 \leqslant i \leqslant N} \sum_{i'=1}^{N} u_{i'c}^m \left[1 - \exp\left\{-\beta \sum_{k=1}^{K} w_k^2 \sum_{l=1}^{L} \left(\hat{\rho}_{i',l}^{(k)} - \hat{\rho}_{i,l}^{(k)}\right)^2\right\}\right]$$

        return $\tilde{C}_c = c_q$
    end for
    $iter \leftarrow iter + 1$.
until $\tilde{C}_{old} = \tilde{C}$ or $iter = max.iter$.

Algorithm 2 Exp-WCM-FMCd clustering

4 Simulation study

This section shows the results of a simulation study used to evaluate clustering performance and accuracy. Specifically, we first evaluate the differences between the proposed WCM-FCMd and the unweighted approach proposed by Cerqueti et al. (2021), to show that assigning weights to each conditional moment improves the quality of the partition. Secondly, we also compare a benchmark exploiting the auto-correlation structure of the original time series rather than one of the conditional moments, that is, the ACF-based FCMd proposed by D'Urso and Maharaj (2009). This approach, which is similar in spirit to those based on conditional higher moments, is considered for assessing the usefulness of higher moments-based clustering. Then, the performances in the presence of outliers are evaluated as well.

The quality of the clustering methods is evaluated, over 100 trials, by setting C equal to the number of simulated groups and according to different values of fuzziness parameter $m = 1.3, 1.5, 1.7$, through the Fuzzy Adjusted Rand Index (FARI) and the Fuzzy Jaccard Index (FJI) (Campello 2007). Both indices evaluate the agreement between two partitions, in this case between the hard simulated partition and the fuzzy one resulting from applying each specific clustering method. When outliers are included in the simulations, the two indices are computed considering only the non-outlier units.

We consider two experimental designs: in the first one, the time series are generated from two distinct Gaussian-DCS models with density (7) and, in the second one, the series are generated from two t-DCS models (8). Therefore, we have $C = 2$ clusters characterized by conditional moments with different levels of persistence.

Under Gaussian-DCS, we simulate 25 time series per group (we have $N = 50$) from a Gaussian distribution with different parameters in the two groups. In particular, the 25 time series belonging to the first group are generated by a Gaussian-DCS process with parameters:

$$\boldsymbol{\omega} = \begin{pmatrix} 0.00 \\ 0.42 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1.00 & 0 \\ 0 & 0.30 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 0.80 & 0 \\ 0 & 0.80 \end{pmatrix}, \qquad (29)$$

while the 25 time series belonging to the second one are generated by the following Gaussian-DCS process:

$$\boldsymbol{\omega} = \begin{pmatrix} 0.08 \\ 0.06 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1.00 & 0 \\ 0 & 0.30 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 0.96 & 0 \\ 0 & 0.96 \end{pmatrix}. \qquad (30)$$

Hence, the time series simulated from (29) are characterized by lower persistency than the time series simulated from (30).

Note that the parameters within the $\boldsymbol{\omega}$ vector are optimally determined, after choosing the values in $\mathbf{A}$ and $\mathbf{B}$, such that the simulated time series and their parameters are covariance stationary (for details see Ardia et al. 2019). The parameters in the matrix $\mathbf{A}$ are related to the score-driven component, while those in the matrix $\mathbf{B}$ are associated to the autoregressive part of the process.

Under the generalized-t density, the 25 time series belonging to the first group are generated by a t-DCS process with parameters:

$$\boldsymbol{\omega} = \begin{pmatrix} 0.00 \\ 0.14 \\ -0.47 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1.00 & 0 & 0 \\ 0 & 0.30 & 0 \\ 0 & 0 & 0.30 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 0.80 & 0 & 0 \\ 0 & 0.80 & 0 \\ 0 & 0 & 0.80 \end{pmatrix} \quad (31)$$

while the remaining 25 time series are generated by:

$$\boldsymbol{\omega} = \begin{pmatrix} 0.00 \\ 0.03 \\ -0.09 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1.00 & 0 & 0 \\ 0 & 0.30 & 0 \\ 0 & 0 & 0.30 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 0.96 & 0 & 0 \\ 0 & 0.96 & 0 \\ 0 & 0 & 0.96 \end{pmatrix} \quad (32)$$

To test the robustness of the clustering methods against outliers, anomalous time series are generated from distinct DGPs (for details see the supplementary material) and added to the previous simulated data sets by considering three scenarios characterized by a number of outliers equal to 3, 6, 9, respectively, corresponding to 5%, 10% and 15% of outlier time series in the sample. The outliers are simulated from DGPs alternative to the DCS, precisely several autoregressive, moving average and ARMA processes, whose equations are reported in Sect. 2 of the web supplement. As all the outlier time series are simulated from distinct DGPs, they do not form additional clusters. An example of simulated time series with outliers under Gaussian density is shown in Fig. 1, while the case of the generalized-t is shown in Fig. 2.

4.1 Results under the Gaussian density

In what follows, we evaluate the performance of all clustering methods under consideration based on their capability to assign the 50 time series to the group they belong to. Table 1 provides a summary of the average FARI, while the results in terms of FJI are shown in Table A1 of the supplementary material. The boxplots
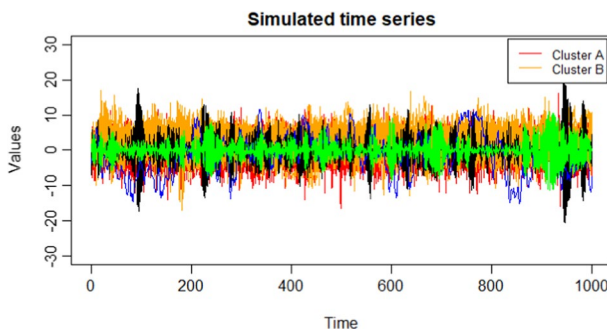

Fig. 1 Example of simulated time series under the Gaussian density. Red time series (Cluster A) are generated from model (29), while orange time series (Cluster B) are generated from (30). The other time series are outliers randomly selected from the processes shown in Sect. 2 of supplementary material (colour figure online)
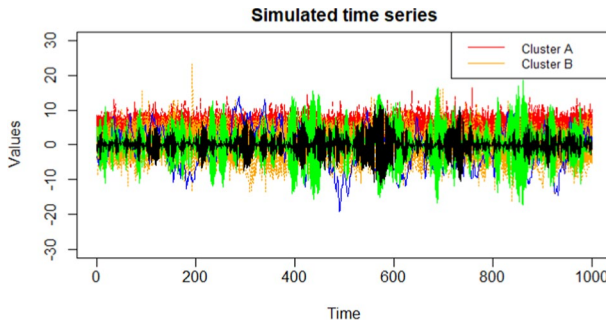
Simulated time series

Fig. 2 Example of simulated time series under the generalized-t density. Red time series (Cluster A) are generated from model (31), while orange time series (Cluster B) are generated from (32). The other time series are outliers randomly selected from the processes shown in Sect. 2 of supplementary material (colour figure online)

showing the distribution of both FARI and FJI are reported in Figs. A1 and A2 of the supplementary material. To make sure that the average values of the alternative clustering procedures are statistically different, we considered t-tests as the clustering procedures are all executed to the same samples, whose results are reported in

Table 1 Average FARI

	Raw	Mean	Var	WCM	Exp-WCM
Scenario I: no outliers					
$m = 1.3$	0.7246	0.8755	0.8388	0.9509	0.9122
$m = 1.5$	0.6484	0.8394	0.7948	0.8964	0.8212
$m = 1.7$	0.6077	0.7823	0.7196	0.8179	0.7216
Scenario II: 5% outliers					
$m = 1.3$	0.7479	0.8469	0.8189	0.9220	0.9637
$m = 1.5$	0.6754	0.7677	0.7663	0.8654	0.9089
$m = 1.7$	0.6522	0.7158	0.7022	0.7763	0.8201
Scenario III: 10% outliers					
$m = 1.3$	0.7151	0.7937	0.7646	0.9204	0.9475
$m = 1.5$	0.6673	0.7364	0.7262	0.8384	0.8973
$m = 1.7$	0.6139	0.6774	0.6486	0.7469	0.8271
Scenario IV: 15% outliers					
$m = 1.3$	0.7239	0.7641	0.7311	0.9086	0.9427
$m = 1.5$	0.6492	0.7173	0.6820	0.8250	0.8969
$m = 1.7$	0.5937	0.6589	0.6250	0.7318	0.8212

Gaussian density

"Raw" is the ACF-FCMd clustering method proposed in D'Urso and Maharaj (2009), "Mean" is the is the clustering procedure of Cerqueti et al. (2021) with conditional mean targeting, while "Var" with conditional variance targeting. "WCM" is the WCM-FCMd clustering procedure discussed in Sect. 3.1, while "Exp-WCM" is the robust procedure presented in Sect. 3.2

Tables A2–A5 included in Sect. 3 of the supplementary material. We notice that the differences in the average FARI and FJI included in Tables A2–A5 are all significant at the 5% level.

In the baseline scenario without outliers, the proposed WCM-FCMd provides the highest performances compared with the alternatives. More in detail, with $m = 1.3$, we have that the WCM-FCMd has an average FARI of 0.95 (see Scenario I of Table 1), while the two targeting approaches have similar performances, that is 0.87 for mean targeting and 0.84 for variance targeting. All the approaches based on the conditional higher moments improve the benchmark based on the raw time series data. As expected, the higher the value of the fuzziness parameter m, the lower the value of the ARI, for all methods, since we are comparing a crisp partition with a fuzzy one; however, the same previous findings still hold also for higher values of the fuzziness parameter. In the baseline scenario, the Exp-WCM-FCMd also performs better than the benchmarks; in fact, the average FARI is always greater than both targeting approaches and the raw time series clustering method. Its lower performance compared with WCM-FCMd is motivated by the fact the exponential transformation smooths the membership degrees too much. The results obtained in terms of FJI are substantially the same as those of FARI.

The WCM-FCMd performs well in the absence of outliers. The Exp-WCM-FCMd also performs better than the benchmarks based on targeting, but worse than the WCM-FCMd. To study how the presence and the number of outliers affect the clustering results, we introduce several outliers into the previously simulated data sets by increasing their number from 3 to 6 and then to 9. The outlier time series are simulated by separate ARMA processes (see Sect. 2 of the supplementary material).

We start considering the case of 3 outlying time series (corresponding to about the 5% of contamination). The results in terms of average FARI in presence of 3 outliers are shown in Table 1 (Scenario II). In this case, the Exp-WCM-FCMd performs better than the WCM-FCMd in terms of average FARI. Furthermore, both the clustering approaches provide a higher clustering quality than the benchmark based on targeting, so we can conclude that weight-based clustering methods outperform the others also in presence of outliers. By inspecting with more detail the results of Table 1 (Scenario II), we find that for $m = 1.3$ the Exp-WCM-FCMd has an average FARI value equal to 0.96 while the WCM-FCMd a value of 0.92. Then, conditional mean targeting provides better performances than conditional variance targeting (0.85 versus 0.82 of the average FARI) but such performances are far from those of the proposed weighted clustering methods. Then, increasing the value of the fuzziness parameter, the performances of all the clustering approaches decrease but the models' ranking does not change.

We then investigate what happens with an increasing number of outliers. The aforementioned results, indeed, hold for a relatively small amount of outliers that are 5% of the dataset size. In what follows, we consider 6 outliers (the 10% of contamination). The clustering performances, measured in terms of FARI, are reported in Table 1 (Scenario III). With 10% of outliers, the difference in the clustering performances between the weighted and the robust weighted clustering procedures becomes larger than the case with 5% of outliers. The improvement in terms of clustering accuracy is more evident as the proportion of outliers increases in the dataset.

Table 2 Average FARI

	Raw	Mean	Var	Shp	WCM	Exp-WCM
Scenario I: no outliers						
$m = 1.3$	0.6655	0.7608	0.7389	0.4181	0.9166	0.8769
$m = 1.5$	0.6078	0.7650	0.7068	0.4110	0.8059	0.8372
$m = 1.7$	0.5559	0.6893	0.6520	0.4006	0.7456	0.7577
Scenario II: 5% outliers						
$m = 1.3$	0.6503	0.7801	0.7084	0.4176	0.8446	0.9201
$m = 1.5$	0.6490	0.7809	0.6622	0.4079	0.7709	0.8485
$m = 1.7$	0.6268	0.7043	0.6067	0.3980	0.7381	0.7422
Scenario III: 10% outliers						
$m = 1.3$	0.3860	0.6840	0.6309	0.4174	0.7048	0.8564
$m = 1.5$	0.3653	0.6250	0.5660	0.4114	0.6324	0.7271
$m = 1.7$	0.3361	0.5334	0.5315	0.4060	0.6553	0.7399
Scenario IV: 15% outliers						
$m = 1.3$	0.2925	0.5518	0.6054	0.4216	0.5828	0.7552
$m = 1.5$	0.2745	0.5240	0.5427	0.4079	0.5442	0.6431
$m = 1.7$	0.2907	0.4267	0.5192	0.3982	0.5538	0.6492

Generalized-t density

"Raw" is the ACF-FCMd clustering method proposed in D'Urso and Maharaj (2009), "Mean" is the is the clustering procedure of Cerqueti et al. (2021) with conditional mean targeting, while "Var" with conditional variance targeting and "Shp" conditional shape targeting. "WCM" is the WCM-FCMd clustering procedure discussed in Sect. 3.1, while "Exp-WCM" is the robust procedure presented in Sect. 3.2

The highest performances are obtained, for all the clustering approaches, in the simulated scenario with a low value of the fuzziness parameter $m = 1.3$. However, by increasing m the overall ranking among clustering approaches does not change.

Similar evidence, indeed, can be found by increasing the number of outliers to 9, i.e. with a 15% of contamination. In the simulated scenarios involving outlier time series, the results obtained in terms of FJI are substantially the same as those of FARI.

4.2 Results under the generalized-t density

In what follows, we evaluate the performance of the considered clustering methods under the assumption of generalized-t density. In this case we have three targeting-based clustering approaches, while the two weighted clustering procedures involve weighting the three moments of the distribution. Table 2 provides a summary of the average values of the FARI, while the FJI results are in Table A6 of the supplementary material. The boxplots summarizing the distributions of the two indices are reported in Figs. A3 and A4 of the supplementary material. To make sure that the average values of the alternative clustering procedures

are statistically different, also in this case we considered t-tests, whose results are shown in Tables A7–A10 included in the supplementary material. The differences in the average FARI and FJI included in Tables A7–A10 are all significant at the 5% level. In the baseline scenario without outliers (see Scenario I of Table 2), also under this distributional assumption we have that the proposed WCM-FCMd provides the highest performances compared with the alternatives with $m = 1.3$, but with increasing fuzziness the WCM-FCMd and the Exp-WCM-FCMd perform similarly. The average FARI of the WCM-FCMd is 0.92, which is higher than the 0.87 achieved with the Exp-WCM-FCMd considering $m = 1.3$. By increasing the fuzziness to $m = 1.5$, however, the WCM-FCMd has an average FARI of 0.81, which is lower than 0.84. The two clustering methods perform similarly for $m = 1.7$, instead. Then, we consider simulated scenarios including contamination. In these cases, we observe an increasing performance with the Exp-WCM-FCMd compared with the WCM-FCMd procedure. Let us focus first on the case of 3 outliers, that is 5% of contamination. The average FARI is shown in Table 2 (see Scenario II). With $m = 1.3$ the Exp-WCM-FCMd achieves an average FARI of 0.92, that is larger than the 0.84 of the WCM-FCMd. The other methods perform poorer, with values between 0.42 and 0.78. By increasing m, the ranking of the clustering methods remains the same, as the Exp-WCM-FCMd is the best method with an average FARI of 0.84, higher than the 0.77 of the WCM-FCMd. With $m = 1.7$, however, the two weighted methods perform similarly. By increasing the degree of contamination from 3 to 6 and 9 outliers – i.e. to 10% and 15% respectively – we observe that the difference in the performances between WCM-FCMd and Exp-WCM-FCMd becomes larger. We refer to the last two scenarios (Scenario III and IV) in Table 2. With $m = 1.3$ and 10% of contamination, the Exp-WCM-FCMd has an average FARI of 0.85, which is very high compared with 0.7 of the WCM-FCMd. In the case of 15% of contamination the difference between the two models further increases as the Exp-WCM-FCMd has an average of 0.76 while the WCM-FCMd only 0.58. The difference between the two clustering methods remains high also with increasing fuzziness m. In sum, also under the generalized-t density we get evidence favouring the use of a weighted clustering procedure instead of those based on targeting.

5 Application to financial time series

In what follows we propose an application of the weighted clustering procedures based on conditional higher moments to financial time series data. The presence of conditional high order moments is a well acknowledged fact in finance (e.g. see Jondeau and Rockinger 2012; Soltyk and Chan 2021), so that the proposed clustering procedures are well suited for handling financial time series. Although it is known that not only the second moment varies over the time, the clustering approaches based on conditional volatility—such as the GARCH-based clustering approaches—are the most widely used. For this reason, we claim that considering time variation

in high order moments is particularly useful especially when financial time series have to be clustered.

For the experiment with real data, we select stocks included in the FTSE MIB 30 Index, including the 30 most capitalized stocks in Italy. The list of selected stocks and their tickers is reported in Table A11 of the supplementary material. We consider the daily returns of the stocks during the last thre years, that is between 1-st October 2020 and 30-th October 2022. On the basis of the discussion highlighted so far, we exploit the clustering structure due to the time variation in higher conditional moments instead of clustering stocks on the basis of the raw returns' time series. Let us discuss the results related to the developed clustering methods in the case of Gaussian density specification. The considered time series and the estimated conditional moments are shown in Figs. A4 and A5 of the supplementary material. Considering the differences across the stocks' higher moments (a brief discussion is provided in the supplementary material), it is natural to expect that clustering stocks on the basis of both conditional moments would be more suitable than an approach based on raw data. Figure A8 in the supplement material exploits differences in terms of conditional moments' auto-correlation structure.

In what follows, we apply the proposed clustering methods—WCM-FCMd and Exp-WCM-FCMd—to the dataset. The partitions obtained with the two proposed clustering methods are also compared with the literature approaches based on targeting.

In order to apply the clustering procedures, both the number of clusters C and the fuzziness parameter m have to be set. We choose the combination of m and C that maximizes the Fuzzy Silhouette (FS) index (Campello 2007). About the choice of the fuzziness parameter, according to the literature (e.g. Kamdar and Joshi 2000; Maharaj et al. 2019) we consider a value of m falling between 1 and 1.5. Based on the maximization of the FS index, we select $m = 1.5$ and $C = 2$ for the proposed weighted clustering methods and $m = 1.3$ and $C = 2$ for the methods employing targeting (Cerqueti et al. 2021).

Figure 3 shows the value of the Fuzzy Silhouette index for $C \in \{2, 3, 4, 5, 6\}$ and $m = 1.5$ for our proposals and $C \in \{2, 3, 4, 5, 6\}$ and $m = 1.3$ for the benchmarks. Accordingly, we choose $C = 2$ for all the clustering approaches.

The membership degrees matrices obtained with the WCM-FCMd and Exp-WCM-FCMd clustering methods are reported in Table 3. The membership degree matrices highlight important differences in the partition with respect to the four outlier time series in the sample. Indeed, Exp-WCM-FCMd assigns a membership equal to $1/C$ (0.5, in this case) to the four time series A2A, BPE, DIA and STLA, while the WCM-FCMd assigns a higher membership to these series. The differences across the two methods also concern the set of employed weights. The Exp-WCM-FCMd assigns a weight of 0.75 to the mean and 0.25 to the variance, while the WCM-FCMd method assigns 0.53 to the mean and 0.47 to the variance.

Figure 4 shows the auto-correlation structure—in terms of the Box-Pierce Q statistics—of the clusters identified with the Exp-WCM-FCMd method. The left panel shows the entire dataset, while the right panel removes the A2A, BPE, DIA and STLA time series focusing on the block of stocks. The points are coloured in
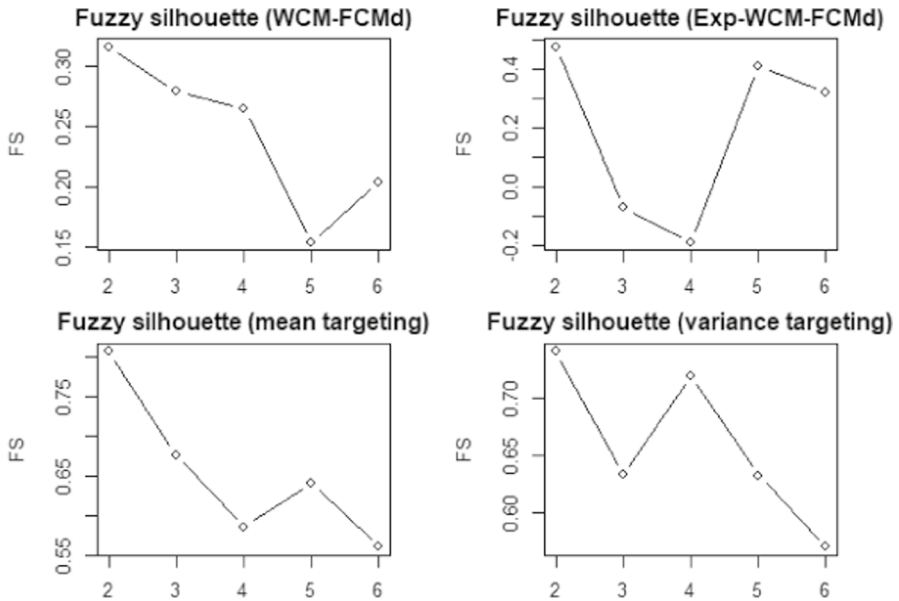
Fig. 3 Fuzzy Silhouette (FS) of Campello (2007) according to several choices of the number of clusters *C*

blue or red according to the partition provided by Exp-WCM-FCMd. In accordance with previous studies, in black are the units with membership degrees lower than 0.7 (Dembele and Kastner 2003; Belacel et al. 2004). The corresponding medoids are identified by a square and a triangle. By looking at Fig. 4, we notice that both ENI and FBK are possible additional outliers, recognized by Exp-WCM-FCMd and not by the WCM-FCMd. The medoids' ACF related to both conditional mean and conditional variance are shown in Figs. A7–A10 in the supplementary material.

The membership degree matrices associated with the targeting approaches are shown in Table A.13 of the web supplement. By looking at the obtained partitions, we observe that both targeting methods are affected by the presence of outliers. More details can be found in Sect. 6.4 of the web supplement. The present application with real data shows why considering both dimensions in clustering time series of financial type is important: focusing on a single dimension can lead to misleading results. Furthermore, our results provide evidence about the usefulness of a robust procedure in the presence of outliers. Indeed, without a robust procedure, clustering methods may fail in discovering the clustering structure of a dataset, which is the main drawback of the targeting approach.

Table 3 Membership degree of the WCM-FCMd clustering methods—standard and exponential approaches

	WCM-FCMd		Exp-WCM-FCMd	
	STM.MI	CPR.MI	MONC.MI	LDO.MI
A2A.MI	0.3168	0.6832	0.4999	0.5001
AMP.MI	0.0505	0.9495	0.0489	0.9511
ATL.MI	0.6457	0.3543	0.9278	0.0722
BAMI.MI	0.2614	0.7386	0.1652	0.8348
BMED.MI	0.4429	0.5571	0.0814	0.9186
BPE.MI	0.4578	0.5422	0.4961	0.5039
CNHI.MI	0.9508	0.0492	0.9593	0.0407
CPR.MI	**0.0000**	**1.0000**	0.0846	0.9154
DIA.MI	0.4855	0.5145	0.4747	0.5253
ENEL.MI	0.8783	0.1217	0.9755	0.0245
ENI.MI	0.1276	0.8724	0.4227	0.5773
FBK.MI	0.7925	0.2075	0.6161	0.3839
G.MI	0.5301	0.4699	0.0601	0.9399
INW.MI	0.8351	0.1649	0.7018	0.2982
ISP.MI	0.1725	0.8275	0.0806	0.9194
LDO.MI	0.1989	0.8011	**0.0000**	**1.0000**
MB.MI	0.8355	0.1645	0.9658	0.0342
MONC.MI	0.9839	0.0161	**1.0000**	**0.0000**
NEXI.MI	0.9761	0.0239	0.9793	0.0207
PRY.MI	0.9338	0.0662	0.9595	0.0405
PST.MI	0.7964	0.2036	0.9083	0.0917
RACE.MI	0.9748	0.0252	0.9861	0.0139
REC.MI	0.0537	0.9463	0.3120	0.6880
SRG.MI	0.0898	0.9102	0.1794	0.8206
STLA.MI	0.3885	0.6115	0.5000	0.5000
STM.MI	**1.0000**	**0.0000**	0.9560	0.0440
TEN.MI	0.6376	0.3624	0.9408	0.0592
TIT.MI	0.3934	0.6066	0.3710	0.6290
TRN.MI	0.9859	0.0141	0.9748	0.0252
UNI.MI	0.7493	0.2507	0.1588	0.8412

The column header reports the medoid of the group, whose membership degrees are highlighted in bold

6 Conclusions

In this paper, we consider the conditional higher moments of the time series as interesting features for clustering. This choice is motivated by the abundant evidence documenting the time variation characterizing the higher moments of the time series. In particular, we consider fuzzy clustering techniques to deal with the uncertainty in the assignment of the units to the clusters.
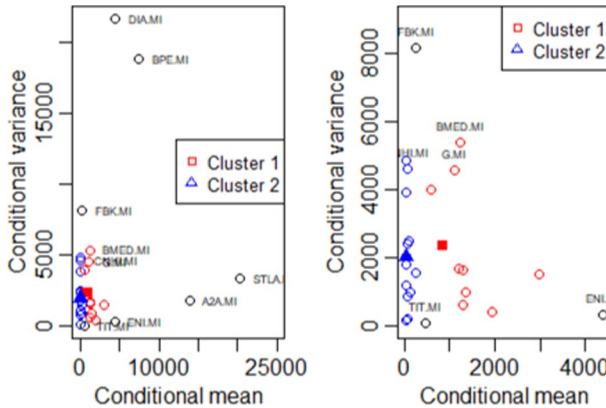
Fig. 4 Auto-correlation structure (Box-Pierce *Q* statistics) of the clusters identified with the Exp-WCM-FCMd method

We contribute to previous literature by proposing two new clustering methods for time series that share a similar conditional distribution. First, we define a clustering approach that assigns optimal weights to different conditional moments. Indeed, it is reasonable to assume that each conditional moment is relevant in explaining the entire data distribution. Therefore, we do not consider each conditional moment separately as previous papers do. Then, we also develop a robust weighted clustering procedure by considering a weighted exponential dissimilarity measure. The introduction of a robust clustering method is motivated by the fact that, in many real-world examples, outliers or noisy data can reduce the performance of the standard clustering approaches, both in terms of the number of groups to be identified and of the assignment of the units to their cluster.

The usefulness of the proposed approach is demonstrated through both a study with simulated data and an empirical application to FTSEMIB 30 time series. Overall, in the simulation study, the proposed weighted clustering approach is characterized by better performances than the considered alternative procedures, based on conditional moments targeting and auto-correlation structure of the original time series. Moreover, the robust approach ensures that a suitable clustering structure can be recovered also when a considerable amount of outliers is present in our data. Through a simulation study, we show that the robust method is the only one that can maintain the same level of accuracy as the level of contamination in the data increases.

The proposed procedures, like other distribution-based clustering approaches, require the specification of a probability distribution. This choice is however not straightforward when dealing with real data. A suggestion for the users of the proposed clustering procedures is to select a general distribution family—such as the generalized-t—which encompasses other well-known distributions as special cases. However, in the case of long time series, it can also be reasonable to assume a Gaussian distribution. The issue related to the selection of the distribution family for clustering purposes is however an important topic which deserves future research.

Additional future research directions can be outlined as follows. First, it is possible to develop alternative robust clustering approaches considering, for example, the presence of a noise cluster or a trimmed clustering method. Second, a mixture approach based on the DCS method can be developed and compared with the clustering procedures proposed in this paper. Then, also the case with multivariate time series data can be analyzed in future work. In the end, we can mention the interesting application of the proposed clustering method, which can be used with the aim of building portfolios of stocks for investment purposes.

References

Alonso AM, Maharaj EA (2006) Comparison of time series using subsampling. Comput Stat Data Anal 50(10):2589–2599

Ardia D, Boudt K, Catania L (2019) Generalized autoregressive score models in r: the gas package. J Stat Softw 88(6):1–28

Bastos JA, Caiado J (2021) On the classification of financial data with domain agnostic features. Int J Approx Reason 138:1–11

Belacel N, Čuperlović-Culf M, Laflamme M, Ouellette R (2004) Fuzzy j-means and VNS methods for clustering genes from microarray data. Bioinformatics 20(11):1690–1701

Bezdek JC (1981) Objective function clustering. Pattern recognition with fuzzy objective function algorithms, pp 43–93. Springer

Caiado J, Crato N (2010) Identifying common dynamic features in stock returns. Quant Finance 10(7):797–807

Caiado J, Crato N, Peña D (2006) A periodogram-based metric for time series classification. Comput Stat Data Anal 50(10):2668–2684

Caiado J, Crato N, Poncela P (2020) A fragmented-periodogram approach for clustering big data time series. Adv Data Anal Classif 14(1):117–146

Campello RJ (2007) A fuzzy extension of the rand index and other related indexes for clustering and classification assessment. Pattern Recognit Lett 28(7):833–841

Cerqueti R, Giacalone M, Mattera R (2021) Model-based fuzzy time series clustering of conditional higher moments. Int J Approx Reason 134:34–52

Creal D, Koopman SJ, Lucas A (2013) Generalized autoregressive score models with applications. J Appl Econom 28(5):777–795

Dembele D, Kastner P (2003) Fuzzy c-means method for clustering microarray data. Bioinformatics 19(8):973–980

D'Urso P, De Giovanni L, Massari R (2016) Garch-based robust clustering of time series. Fuzzy Sets Syst 305:1–28

D'Urso P, De Giovanni L, Massari R, D'Ecclesia RL, Maharaj EA (2020) Cepstral-based clustering of financial time series. Expert Syst Appl 161:113705

D'Urso P, De Giovanni L, Vitale V (2022) Spatial robust fuzzy clustering of covid 19 time series based on b-splines. Spat Stat 49:100518

D'Urso P, Maharaj EA (2009) Autocorrelation-based fuzzy clustering of time series. Fuzzy Sets Syst 160(24):3565–3589

D'Urso P, Massari R (2019) Fuzzy clustering of mixed data. Inf Sci 505:513–534

García-Escudero LÁ, Gordaliza A (1999) Robustness properties of k means and trimmed k means. J Am Stat Assoc 94(447):956–969

Garcia-Escudero LA, Gordaliza A (2005) A proposal for robust curve clustering. J Classif 22(2):185–201

Harvey CR, Liechty JC, Liechty MW, Müller P (2010) Portfolio selection with higher moments. Quant Finance 10(5):469–485

Harvey CR, Siddique A (1999) Autoregressive conditional skewness. J Financ Quant Anal 34:465–487

Iorio C, Frasso G, D'Ambrosio A, Siciliano R (2016) Parsimonious time series clustering using p-splines. Expert Systems Appl 52:26–38

Jackman S (2009) Bayesian analysis for the social sciences. Wiley

Jondeau E, Rockinger M (2012) On the importance of time variability in higher moments for asset allocation. J Financ Econom 10(1):84–123

Kamdar T, Joshi A (2000) On creating adaptive web servers using weblog mining. UMBC Student Collection

Krishnapuram R, Joshi A, Nasraoui O, Yi L (2001) Low-complexity fuzzy relational clustering algorithms for web mining. IEEE Trans Fuzzy Syst 9(4):595–607

Lafuente-Rego B, Vilar JA (2016) Clustering of time series using quantile autocovariances. Adv Data Anal Classif 10(3):391–415

León Á, Rubio G, Serna G (2005) Autoregresive conditional volatility, skewness and kurtosis. Q Rev Econom Finance 45(4–5):599–618

Liao TW (2005) Clustering of time series data—a survey. Pattern Recognit 38(11):1857–1874

Maharaj EA (1996) A significance test for classifying arma models. J Stat Comput Simul 54(4):305–331

Maharaj EA (2000) Cluster of time series. J Classif 17(2):297–314

Maharaj EA, D'Urso P, Caiado J (2019) Time series clustering and classification. CRC Press

Maharaj EA, D'Urso P, Galagedera DU (2010) Wavelet-based fuzzy clustering of time series. J Classif 27(2):231–275

Mantegna RN (1999) Hierarchical structure in financial markets. Eur Phys J B Condens Matter Complex Syst 11(1):193–197

Mattera R, Giacalone M, Gibert K (2021) Distribution-based entropy weighting clustering of skewed and heavy tailed time series. Symmetry 13(6):959

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

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

Savvides A, Promponas VJ, Fokianos K (2008) Clustering of biological time series by cepstral coefficients based distances. Pattern Recognit 41(7):2398–2412

Soltyk SJ, Chan F (2021) Modeling time-varying higher-order conditional moments: a survey. J Econ Surv. https://doi.org/10.1111/joes.12481

Wang X, Smith K, Hyndman R (2006) Characteristic-based clustering for time series data. Data Min Knowl Discov 13(3):335–364

Wu K-L, Yang M-S (2002) Alternative c-means clustering algorithms. Pattern Recognit 35(10):2267–2278

Yang M-S, Wu K-L (2004) A similarity-based robust clustering method. IEEE Trans Pattern Anal Mach Intell 26(4):434–448

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.