REGULAR ARTICLE

CrossMark

Robust fuzzy clustering based on quantile autocovariances

B. Lafuente-Rego[1] · P. D'Urso[2] · J. A. Vilar[1]

Abstract

Robustness to the presence of outliers in time series clustering is addressed. Assuming that the clustering principle is to group realizations of series generated from similar dependence structures, three robust versions of a fuzzy C-medoids model based on comparing sample quantile autocovariances are proposed by considering, respectively, the so-called metric, noise, and trimmed approaches. Each method achieves its robustness against outliers in different manner. The metric approach considers a suitable transformation of the distance aimed at smoothing the effect of the outliers, the noise approach brings together the outliers into a separated artificial cluster, and the trimmed approach removes a fraction of the time series. All the proposed approaches take advantage of the high capability of the quantile autocovariances to discriminate between independent realizations from a broad range of stationary processes, including linear, non-linear and conditional heteroskedastic models. An extensive simulation study involving scenarios with different generating models and contaminated with outliers is performed. Robustness against (i) outliers generated from different generating patterns, and (ii) outliers characterized by isolated, temporary or persistent level changes is evaluated. The influence of the input parameters required by the different algorithms is analyzed. Regardless of the considered models, the results show that the proposed robust procedures are able to neutralize the effect of the anomalous series preserving the true clustering structure, and fairly outperform other robust algorithms based on alternative metrics. Two applications to financial data sets permit to illustrate the usefulness of the proposed models.

Keywords Time series data · Robust fuzzy C-medoids clustering · Quantile autocovariances · Exponential distance · Noise cluster · Trimming

✉ B. Lafuente-Rego
borja.lafuente@udc.es

Extended author information available on the last page of the article

🖉 Springer

1 Introduction

Time series are in nature complex data objects. They are usually formed by a huge number of records, present dynamic behavior patterns which might change over time, and frequently one must handle realizations of different length. This kind of particularities make it more difficult to perform cluster analysis in a standard way. For example, it is not simple to determine a proper distance between time series exhibiting robustness to the dependence structure, and high dimensionality is an additional obstacle to develop efficient clustering procedures. On the other hand, time series clustering is a central problem in a broad range of applications including economics, finance, econophysics, marketing, environmental sciences, neuroscience, and biomedical sciences, among others (see e.g. Liao 2005; Vilar et al. 2009; D'Urso et al. 2015). These arguments account for the great focus of attraction that this topic has led over the past decades in research, mainly into the fields of Statistics, Data Mining and Artificial Intelligence (Chae et al. 2008; Vilar et al. 2009; Fu 2011; Rani and Sikka 2012; Montero and Vilar 2014; Aghabozorgi et al. 2015; Di Lascio and Giannerini 2016; Lafuente-Rego and Vilar 2016b; Górecki et al. 2018).

Following Caiado et al. (2015), in a methodological point of view, time series clustering methods can be classified into the following categories (for more details, see Caiado et al. 2015; D'Urso et al. 2016):

– Observation-based clustering. Methods based on the actual observed time series (Coppi and D'Urso 2002, 2003, 2006; Coppi et al. 2010; D'Urso 2004, 2005; D'Urso and De Giovanni 2008; D'Urso et al. 2018).
– Feature-based clustering. Methods based on suitable features derived for the time series, i.e., autocorrelation function (Alonso and Maharaj 2006; Caiado et al. 2006, 2009; D'Urso and Maharaj 2009), partial autocorrelation function (Caiado et al. 2009), inverse autocorrelation function (Caiado et al. 2006), quantile autocovariances (Lafuente-Rego and Vilar 2016a, b; Vilar et al. 2018), periodogram and its transformations (Caiado et al. 2009; Maharaj and D'Urso 2011), coherence (Maharaj and D'Urso 2010), cepstral (Maharaj and D'Urso 2011), wavelet decomposition (Maharaj and D'Urso 2010; D'Urso and Maharaj 2012; D'Urso et al. 2014), and oher spectral features (Vilar and Pértega 2004; Vilar et al. 2009; Pértega and Vilar 2010; Vilar et al. 2010).
– Model-based clustering. Assuming that a set of time series generated from the same model have similar patterns, this category encompasses methods based on clustering models fitted to the time series, either by means of parameter estimates or by means of the residuals of the fitted models. In this class, different time series clustering methods are based on ARIMA models (Piccolo 1990; Maharaj 1996, 1999, 2000; Kalpakis et al. 2001; Xiong and Yeung 2004; D'Urso et al. 2013c, b), on GARCH models (Otranto 2008, 2010; Caiado and Crato 2010; D'Urso et al. 2013a, 2016), on functional (James and Sugar 2003; Tarpey and Kinateder 2003; Alonso et al. 2006; Vilar et al. 2010; Slaets et al. 2012; Maharaj et al. 2015; Tsay 2016; Floriello and Vitelli 2017), copulas (Durante et al. 2015; De Luca and Zuccolotto 2017; Disegna et al. 2017), extreme value analysis (D'Urso et al. 2017a) and spline representation of the time series (García-Escudero and Gordaliza 2005).

Regardless of the considered clustering procedure, a very important issue is to provide mechanisms to isolate outliers and neutralize their negative effects in the clustering process. It is well known that the presence of anomalous data can prevent from correctly identifying the hidden clustering structure. If, for example, the clustering principle is to group series coming from the same generating model, then the presence of an anomalous series within a legitimate cluster means to contaminate the cluster, thus contributing to misidentify the dynamic pattern characterizing the cluster. Even though outliers are identified, it is important to assess whether the clustering procedure is affected by their presence. In sum, introducing robust clustering methods for classifying time series is indeed an important matter deserving specific attention from researchers. Three main methodological approaches to perform robust clustering of series have been adopted in the literature, which can be classified as follows.

– Noise approach. To neutralize the negative effects of possible outliers in the clustering process, outlier time series are separated and assigned to an artificial cluster, the so-called noise cluster (see e.g. D'Urso et al. 2013b).
– Metric approach. Distance measures with robust properties are incorporated in the objective function of the clustering method (see e.g. D'Urso et al. 2015).
– Trimmed approach. The clustering method is applied to the time series remaining after a fixed fraction of outlier time series are removed (see e.g. D'Urso et al. 2017b; García-Escudero and Gordaliza 2005).
– Influence weighting approach. A weighting system is incorporated in the clustering method for assigning objectively low weights to outlier time series (see e.g. D'Urso 2005).

In this paper, our main concern is to develop robust clustering procedures to deal with time series. It is assumed that the target is to group series according to the underlying dependence structures, so that similarity between series is understood in terms of similarity between generating processes. This is a very natural criterion, which makes sense if for example one wishes to identify a few temporal behavior patterns hidden behind long time series observed with a high amount of noise. According to the clustering purpose in mind, two important points must be remarked. First, a time series is considered as an outlier when it exhibits an atypical dynamic behavior, different from the rest of identified prototypes. Second, a dissimilarity measure robust to the generating mechanism is necessary in order to attain proper cluster solutions.

Concerning the first point, we assume that the atypical behavior of an outlier can have two possible causes: (i) the outlier series has been generated from a process different from the ones defining the clusters, or (ii) the outlier series comes from one cluster but its behavior has been distorted during a short or long period of time. Robustness against both types of outliers are studied. As regards the robust dissimilarity measure, we propose to use the distance based on comparing sequences of sample quantile autocovariances proposed by Lafuente-Rego and Vilar (2016a). Unlike other extracted features, quantile autocovariances account for important dynamic features of time series and are well-defined for a broad class of processes, including non-linear and heteroskedastic processes (Linton and Whang 2007; Lee and Rao 2012). The results from numerical experiments reported in Lafuente-Rego and Vilar (2016a) showed the high capability of this metric to discriminate between a broad range of dependence

structures, thus illustrating its robustness to deviations from a particular generating model.

Based on the prior arguments, we have considered the feature-based clustering approach (the time series are represented by a set of sample quantile autocovariances) and adopted the Partitioning Around Medoids (PAM) technique in a fuzzy framework to introduce three new robust time series clustering models based on, respectively, the metric, noise, and trimmed approaches.

The proposed models are robust extensions of the fuzzy C-medoid clustering model based on quantile autocovariances (QAF-FCMdC model) introduced by Vilar et al. (2018). Recent works have followed analogous robust approaches but using other distances, such as the distance between autoregressive representations (D'Urso et al. 2013c, 2015, 2017b) and model-based distances considering underlying heteroskedastic models (D'Urso et al. 2016). To gain insight on the behavior of the proposed robust models, a comparison with the mentioned alternatives has been carried out by means of an extensive simulation study including ARMA and GARCH models and in the presence of outliers. Note that the alternative procedures take advantage of being specifically constructed to discriminate between these kinds of processes, and therefore our simulation results provide a realistic measure of the capability of the procedures based on quantile autocovariances. In addition, the usefulness and effectiveness of the proposed robust fuzzy models is also highlighted by considering two study cases with real data.

The rest of the paper is organized as follows. In Sect. 2, the distance considering sample quantile autocovariances is presented and the robust fuzzy clustering models based on this metric are proposed. Specific advantages of the introduced clustering methods are discussed in Sect. 3. Results from a simulation study supporting the good performance of the proposed robust methods are reported in Sect. 4. Section 5 is devoted to presenting two applications of the proposed models on real data involving financial time series, and some concluding remarks are summarized in Sect. 6.

2 Fuzzy clustering based on the QAF-distance: robust approaches

In this section, the distance between time series based on quantile autocovariances is formally defined, the fuzzy C-medoid clustering model considering this distance is presented, and three robust versions considering the metric, noise, and trimmed approaches are proposed.

2.1 Distance between time series based on quantile autocovariances (QAF-distance)

Since we wish to perform clustering governed by similarity between generating models, a dissimilarity measure aimed at capturing high-level dynamic structures describing the global performance of the series is expected to report satisfactory results. With this in mind, Lafuente-Rego and Vilar (2016a) propose to measure the distance between two time series comparing their quantile autocovariance functions.

Let $\{X_t; \ t \in \mathbb{Z}\}$ be a strictly stationary process with marginal distribution function F_X having continuous density f_X. Denote by $q_\alpha = F_X^{-1}(\alpha) = \inf\{q \in \mathbb{R} : F_X(q) \geq \alpha\}$, for $\alpha \in [0, 1]$, the corresponding quantile function. Then, the *quantile autocovariance function of lag l* for the process X_t is defined by

$$\gamma_l(\tau, \tau') = \mathrm{cov}\left\{I\left(X_t \leq q_\tau\right), I\left(X_{t+l} \leq q_{\tau'}\right)\right\},$$

where $I(\cdot)$ denotes the indicator function and $(\tau, \tau') \in [0, 1]^2$ denotes an arbitrary couple of quantile levels. According to this definition, the functions $\gamma_l(\cdot, \cdot), l = 1, 2, \ldots$, account for the serial dependence structure of the time series, reporting information on serial features such as the existence of conditional heteroskedasticity or changes in conditional shapes, which cannot be detected by other feature-based distances proposed in the literature. The sensitivity of the quantile autocovariance functions to capture complex dynamic features suggests that a dissimilarity comparing these functions should exhibit high capability to discriminate between a broad range of generating processes.

Based on these considerations, Lafuente-Rego and Vilar (2016a) define the QAF-distance (d_{QAF}) between a pair of observed times series as the squared Euclidean distance between estimates of their quantile autocovariances over a common range of selected lags and quantiles. Specifically, consider realizations of two time series $X_t^{(1)}$ and $X_t^{(2)}$ of lengths $T^{(1)}$ and $T^{(2)}$, respectively. For prefixed ranges of L lags, $l_1, \ldots, l_L$, and r quantile levels, $0 < \tau_1 < \ldots < \tau_r < 1$, each series $X_t^{(u)}$, $u = 1, 2$, is characterized by the vector $\boldsymbol{\Gamma}^{(u)} = \left(\hat{\gamma}_{l_i}^{(u)}(\tau_j, \tau_{j'})\right), i = 1, \ldots, L, j, j' = 1 \ldots, r$, whose elements $\hat{\gamma}$ denote the sample quantile autocovariance based on the empirical quantiles $\hat{q}_\tau$ obtained from the observed series, i.e.

$$\hat{\gamma}_{l_i}^{(u)}(\tau_j, \tau_{j'}) = \frac{1}{T^{(u)} - l_i} \sum_{t=1}^{T^{(u)} - l_i} I\left(X_t^{(u)} \leq \hat{q}_{\tau_j}\right) I\left(X_{t+l_i}^{(u)} \leq \hat{q}_{\tau_{j'}}\right) - \tau_j \, \tau_{j'}, \quad u = 1, 2.$$

$$(1)$$

Then, the QAF-distance is given by

$$d_{QAF}\left(X_t^{(1)}, X_t^{(2)}\right) = \left\|\boldsymbol{\Gamma}^{(1)} - \boldsymbol{\Gamma}^{(2)}\right\|_2^2$$

$$= \sum_{i=1}^{L} \sum_{j=1}^{r} \sum_{j'=1}^{r} \left(\hat{\gamma}_{l_i}^{(1)}(\tau_j, \tau_{j'}) - \hat{\gamma}_{l_i}^{(2)}(\tau_j, \tau_{j'})\right)^2. \quad (2)$$

For further details on the QAF-distance, including asymptotic considerations and motivating examples for its use in time series clustering, readers are referred to Vilar et al. (2018). It is worthy to notice that cross-dependency is not addressed in this work. All the examined distances focus on comparing individual time evolution patterns, omitting possible cross-sectional dependence between series. Distances based on features such as correlation or cross-correlation might be helpful to deal with cross-dependency, but this issue goes beyond the scope of this work.

The rest of the section is devoted to describe four fuzzy clustering models for classifying time series based on the QAF-distance and the Partitioning Around Medoids (PAM) procedure. The first one (QAF-FCMdC model) is timidly robust (Sect. 2.2), while the other three proposed methods are highly robust to the presence of possible outlier time series in the dataset subjected to clustering. As mentioned, these three robust methods are based, respectively, on metric, noise and trimmed approaches, i.e., the QAF-based fuzzy C-medoids clustering with exponential distance (QAF-FCMdC-Exp) method (Sect. 2.3), the QAF-based fuzzy C-medoids clustering with noise cluster (QAF-FCMdC-NC) method (Sect. 2.4), and the QAF-based trimmed fuzzy C-medoids clustering (QAF-TrFCMdC) method (Sect. 2.5).

Each method overcomes the problems connected to the presence of outliers neutralizing in a different manner the effect of outliers in the clustering process. In detail, the QAF-FCMdC-Exp method achieves its robustness with respect to outliers by inheriting the robustness of the Exponential distance, the QAF-FCMdC-NC method achieves its robustness with respect to outliers by introducing a particular cluster, called noise cluster, represented by a fictitious prototype, called noise prototype, and the QAF-TrFCMdC method achieves its robustness by trimming away a certain fraction of the data.

Hereafter, $S = \left\{ X^{(1)}, \ldots, X^{(n)} \right\}$ denotes the set of n observed time series subjected to clustering, and $\Gamma = \left\{ \Gamma^{(1)}, \ldots, \Gamma^{(n)} \right\}$, $\Gamma^{(i)}$ being the vector of quantile autocovariances estimated from the ith observed series $X^{(i)}$, for $1 \le i \le n$. The specific lags and quantile levels considered to construct $\Gamma^{(i)}$ are established using an optimization algorithm focused on selecting the combination of lags and pairs of quantile levels which minimize the within-group variance and maximize the between-group variance simultaneously (see Sect. 3 in Vilar et al. 2018). For $i, j \in \{1, \ldots, n\}$, the distance between $X^{(i)}$ and $X^{(j)}$ is given by $\left\| \Gamma^{(i)} - \Gamma^{(j)} \right\|_2^2$, such as defined in (2).

2.2 QAF-based fuzzy C-medoids clustering (QAF-FCMdC) method

The QAF-based fuzzy C-medoids clustering (QAF-FCMdC) (Lafuente-Rego and Vilar 2016a) finds the subset of size C of Γ, $\tilde{\Gamma} = \left\{ \tilde{\Gamma}^{(1)}, \ldots, \tilde{\Gamma}^{(C)} \right\}$, and the $p \times C$ matrix of fuzzy coefficients $\Omega = \left(u_{i,c} \right)$ that lead to solve the minimization problem:

$$
\begin{cases}
\min_{\tilde{\Gamma}, \Omega} \sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^m \left\| \Gamma^{(i)} - \tilde{\Gamma}^{(c)} \right\|_2^2 \\
\text{subject to: } \sum_{c=1}^{C} u_{ic} = 1 \text{ and } u_{ic} \ge 0,
\end{cases}
\tag{3}
$$

where $u_{ic} \in [0, 1]$ represents the fuzzy membership degree of the ith time series in the cth cluster, $\tilde{\Gamma}^{(c)}$ is the vector of quantile autocovariances associated to the medoid time series for the cluster c, and $m > 1$ is a weighting exponent that controls the fuzziness of the partition.

Algorithm 1 QAF-FCMdC algorithm

1: Fix C, m and *max.iter*
2: Set $iter = 0$
3: Pick the initial medoids $\widetilde{\boldsymbol{\Gamma}} = \left\{ \widetilde{\boldsymbol{\Gamma}}^{(1)}, \ldots, \widetilde{\boldsymbol{\Gamma}}^{(C)} \right\}$
4: **repeat**
5:     Set $\widetilde{\boldsymbol{\Gamma}}_{\text{OLD}} = \widetilde{\boldsymbol{\Gamma}}$ {Store the current medoids}
6:     Compute u_{ic}, $i = 1, \ldots, n$, $c = 1, \ldots, C$, using (4)
7:     For each $c \in \{1, \ldots, C\}$, determine the index $j_c \in \{1, \ldots, n\}$ satisfying:

$$j_c = \text{argmin}_{1 \le j \le n} \sum_{i=1}^{n} u_{ic}^m \left\| \boldsymbol{\Gamma}^{(i)} - \boldsymbol{\Gamma}^{(j)} \right\|^2$$

8:     **return** $\widetilde{\boldsymbol{\Gamma}}^{(c)} = \boldsymbol{\Gamma}^{(j_c)}$, for $c = 1, \ldots, C$ {Update the medoids}
9:     $iter \leftarrow iter + 1$
10: **until** $\widetilde{\boldsymbol{\Gamma}}_{\text{OLD}} = \widetilde{\boldsymbol{\Gamma}}$ or $iter = max.iter$

The constraints on u_{ic} are standard requirements in fuzzy clustering. In particular, that the sum of the membership degrees for each series equals 1 implies that all of them contribute with the same weight to the clustering process. Parameter m determines the level of fuzziness introduced in the clustering procedure. In the naive case $m = 1$, we have $u_{ic} = 1$ if the ith series is the medoid for the cluster c and 0 otherwise so that the crisp version of the procedure is obtained. As the value of m increases, the boundaries between clusters become softer and therefore the classification is fuzzier.

In a nutshell, the aim of the QAF-FCMdC model is to determine a fuzzy partition into C clusters such that the QAF-distance between the clusters and their prototypes is minimized. The clustering quality strongly depends on the capability of the QAF-metric to identify different dependence structures, but the fuzzy approach means that the non-stochastic uncertainty inherent to the assignment of series to clusters is incorporated to the procedure by means of the membership degrees.

An iterative algorithm that alternately optimizes the membership degrees and the medoids is used to solve the optimization problem in (3). First, the membership degrees are optimized for a set of fixed medoids. The iterative solutions for the membership degrees take the form (Höppner 1999):

$$u_{ic} = \left[\sum_{c'=1}^{C} \left(\frac{\left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|^2}{\left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c')} \right\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \quad \text{for } i = 1, \ldots, n \text{ and } c = 1, \ldots, C.$$

(4)

Then, based on the membership degrees obtained from (4), the C series minimizing (3) are selected as new medoids. This two-step procedure is iterated until there is no change in the medoids or a maximum number of iterations is achieved.

The QAF-based fuzzy C-medoids clustering algorithm (QAF-FCMdC) is implemented as outlined in Algorithm 1.

Notice that the PAM-based fuzzy clustering represents a robustification of the fuzzy C-means clustering. Furthermore, robustness is here strengthened by the use of the

QAF-distance since $\hat{\gamma}_{l_i}^{(u)}$ is a robust estimate. Anyway, it is known that the PAM-based fuzzy approach only provides a "timid robustification" (García-Escudero and Gordaliza 1999; García-Escudero et al. 2010).

2.3 QAF-based exponential fuzzy C-medoids clustering (QAF-FCMdC-Exp)

The QAF-based exponential fuzzy C-medoids clustering model (QAF-FCMdC-Exp) considers a new objective function given by:

$$
\begin{cases}
\min_{\tilde{\boldsymbol{\Gamma}}, \Omega} \sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^{m} \left[1 - \exp\left\{ -\beta \left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|_{2}^{2} \right\} \right] \\
\text{subject to: } \sum_{c=1}^{C} u_{ic} = 1 \text{ and } u_{ic} \geq 0,
\end{cases}
\tag{5}
$$

where β is a positive constant.

Following Wu and Yang (2002), the local optimal solution for the objective function (5) is

$$
u_{ic} = \left(\sum_{c'=1}^{C} \left[\frac{1 - \exp\left\{ -\beta \left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|_{2}^{2} \right\}}{1 - \exp\left\{ -\beta \left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c')} \right\|_{2}^{2} \right\}} \right]^{\frac{1}{m-1}} \right)^{-1}.
\tag{6}
$$

As remarked by Wu and Yang (2002), the fuzzy clustering based on the Exponential distance is more robust than the fuzzy clustering based on the Euclidean distance. In fact, Exponential distance gives different weights to each data point, according to whether a data point is noisy or not. In particular, the Exponential distance assigns small weights to outliers and larger weights to those data points laying close to the bulk of the dataset (for more details see Wu and Yang 2002; D'Urso et al. 2015).

The value of β, determined as the inverse of the variability in the data (the more variability in the data, the less the value of β), appropriately affects the membership degrees (6) in terms of robustness to outliers. Thus, parameter β can be selected in the following manner:

$$
\beta = \left(\frac{1}{n} \sum_{i=1}^{n} \left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(k)} \right\|_{2}^{2} \right)^{-1},
\tag{7}
$$

where $\widetilde{\boldsymbol{\Gamma}}^{(k)}$ corresponds to the index k satisfying $k = \text{argmin}_{1 \leq i' \leq n} \sum_{i''=1}^{n} \left\| \boldsymbol{\Gamma}^{(i'')} - \boldsymbol{\Gamma}^{(i')} \right\|_{2}^{2}$. See D'Urso et al. (2015) for more details.

In essence, the QAF-FCMdC-Exp model is aimed to smooth the effect of the anomalous series by adjusting their influence with proper weights. As result, the membership degrees of the outliers are similarly distributed across the clusters but the true clustering structure is not seriously affected by their presence.

Algorithm 2 QAF-based Exponential Fuzzy C-Medoids Clustering (QAF-FCMdC-Exp)

1: Fix C, m and *max.iter*
2: Compute β using (7)
3: Set $iter = 0$
4: Pick the initial medoids $\widetilde{\boldsymbol{\Gamma}} = \left\{ \widetilde{\boldsymbol{\Gamma}}^{(1)}, \ldots, \widetilde{\boldsymbol{\Gamma}}^{(C)} \right\}$
5: **repeat**
6:    Set $\widetilde{\boldsymbol{\Gamma}}_{\text{OLD}} = \widetilde{\boldsymbol{\Gamma}}$ {Store the current medoids}
7:    Compute u_{ic}, $i = 1, \ldots, n, c = 1, \ldots, C$, using (6)
8:    For each $c \in \{1, \ldots, C\}$, determine the index $j_c \in \{1, \ldots, n\}$ satisfying:

$$j_c = \text{argmin}_{1 \le j \le n} \sum_{i=1}^{n} u_{ic}^m \left[1 - \exp \left\{ -\beta \left\| \boldsymbol{\Gamma}^{(i)} - \boldsymbol{\Gamma}^{(j)} \right\|^2 \right\} \right]$$

9:    **return** $\widetilde{\boldsymbol{\Gamma}}^{(c)} = \boldsymbol{\Gamma}^{(j_c)}$, for $c = 1, \ldots, C$ {Update the medoids}
10:    $iter \leftarrow iter + 1$
11: **until** $\widetilde{\boldsymbol{\Gamma}}_{\text{OLD}} = \widetilde{\boldsymbol{\Gamma}}$ or $iter = max.iter$

The QAF-based exponential Fuzzy C-medoids clustering model (QAF-FCMdC-Exp) is implemented as outlined in Algorithm 2.

2.4 QAF-based fuzzy C-medoids clustering with noise cluster (QAF-FCMdC-NC)

The QAF-based fuzzy C-medoids clustering with noise cluster (QAF-FCMdC-NC) can be formalized as follows:

$$\begin{cases} \min_{\tilde{\boldsymbol{\Gamma}}, \Omega} \sum_{i=1}^{n} \sum_{c=1}^{C-1} u_{ic}^m \left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|_2^2 + \sum_{i=1}^{n} \delta^2 \left(1 - \sum_{c=1}^{C-1} u_{ic} \right)^m \\ \text{subject to: } \sum_{c=1}^{C} u_{ic} = 1 \text{ and } u_{ic} \ge 0, \end{cases} \quad (8)$$

where δ is the noise distance, to be set in advance, that plays the role of increasing (decreasing), for high (low) values of δ, the emphasis of the "noise component" in the minimization of the objective function in (8).

Notice that the QAF-FCMdC-NC method introduces an artificial cluster, the noise cluster, whose role is to localize the outliers and place them in a single auxiliary class. The noise cluster is represented by a fictitious prototype time series, "noise prototype time series", that has a constant (noise) distance from every time series. A time series belongs to a real cluster only if its distance from a medoid time series is lower than the noise distance; otherwise, the time series belongs to the noise cluster. We remark that the prototypes of the real clusters are observed time series, while the prototype of the noise cluster is fictitious.

This way, the method provides C clusters, but only $(C-1)$ are "legitimate" clusters, with the extra cluster serving as the noise cluster. The difference $u_{i*} = 1 - \sum_{c=1}^{C-1} u_{ic}$ in the second term of the objective function in (8) expresses the membership degree of the ith time series to the noise cluster, for $i = 1, \ldots, n$. So, by definition, the usual constraint on the membership degrees for the real clusters is here relaxed to

$\sum_{c=1}^{C-1} u_{ic} < 1$, which allows noise time series to have small membership values in good clusters.

The objective function in (8) can be minimized in relation to the membership degrees in a similar manner to the noise clustering method suggested by Dave (1991), obtaining:

$$u_{ic} = \left(\sum_{c'=1}^{C} \left[\frac{\left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|_2^2}{\left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c')} \right\|_2^2} \right]^{\frac{1}{m-1}} + \left[\frac{\left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|_2^2}{\delta^2} \right]^{\frac{1}{m-1}} \right)^{-1}. \qquad (9)$$

Overall, the success of the fuzzy clustering model with noise cluster depends on the appropriate choice of the noise distance δ. If it is too large, the fuzzy clustering model with noise cluster degenerates to the non-robust version of the model and outliers are forced to belong to real clusters; vice versa, if δ is too small, a lot of objects can be considered as noise and misplaced into the noise cluster (Cimino et al. 2005). Although some heuristic solutions to estimate the optimal value of δ have been suggested in the literature, the determination of a suitable value for this parameter is still an open problem.

Davé and Sen (2002) suggest that the value of the distance noise δ should be calculated by considering the dataset statistics, relating its determination to the concept of "scale" in robust statistics (Davé and Krishnapuram 1997). As remarked by Cimino et al. (2005), "unfortunately, the proper estimation of [the] scale is not a trivial task (Davé and Sen 1997) and requires some knowledge of the data, which cannot always be supposed in real clustering applications".

In the initial noise clustering algorithm (Dave 1991), the value of δ was set to:

$$\delta^2 = \lambda \frac{1}{I(C-1)} \sum_{i=1}^{I} \sum_{c=1}^{C-1} \left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|_2^2, \qquad (10)$$

where λ is a scale multiplier to be selected depending on the nature of data.

To obtain the most suitable value of λ (and hence of δ), Cimino et al. (2005) suggest to execute the fuzzy clustering with noise cluster with decreasing values of λ and analyze the distribution of the percentage of objects assigned to the noise cluster. The distribution has an abrupt change of slope (elbow) when the value of the noise distance is so small that objects naturally belonging to real clusters are clustered into the noise cluster. In correspondence to the elbow it is possible to determine the optimal noise distance. The authors approximate the distribution of percentages with a Pareto distribution.

The steps of the algorithm of QAF-FCMdC-NC method are outlined in Algorithm 3.

2.5 QAF-based trimmed fuzzy C-medoids clustering (QAF-TrFCMdC)

The QAF-based trimmed fuzzy C-medoids clustering (QAF-TrFCMdC) achieves its robustness with respect to outliers by trimming away a certain fraction of the data

Algorithm 3 QAF-based Fuzzy C-Medoids Clustering with Noise Cluster (QAF-FCMdC-NC)

1: Fix $C - 1$, m and *max.iter*
2: Set $iter = 0$
3: Pick the initial medoids $\widetilde{\boldsymbol{\Gamma}} = \left\{ \widetilde{\boldsymbol{\Gamma}}^{(1)}, \ldots, \widetilde{\boldsymbol{\Gamma}}^{(C-1)} \right\}$
4: **repeat**
5:　　Set $\widetilde{\boldsymbol{\Gamma}}_{\text{OLD}} = \widetilde{\boldsymbol{\Gamma}}$　　　　　　　　　　　　　　　　　　　{Store the current medoids}
6:　　Compute δ using (10).
7:　　Compute u_{ic}, $i = 1, \ldots, n$, $c = 1, \ldots, C$, using (9)
8:　　For each $c \in \{1, \ldots, C-1\}$, determine the index $j_c \in \{1, \ldots n\}$ satisfying:

$$j_c = \text{argmin}_{1 \leq j \leq n} \sum_{i=1}^{n} u_{ic}^m \left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(j)} \right\|^2$$

9:　　**return** $\widetilde{\boldsymbol{\Gamma}}^{(c)} = \boldsymbol{\Gamma}^{(j_c)}$, for $c = 1, \ldots, C-1$　　　　　　　{Update the medoids}
10:　　　$iter \leftarrow iter + 1$
11: **until** $\widetilde{\boldsymbol{\Gamma}}_{\text{OLD}} = \widetilde{\boldsymbol{\Gamma}}$ or $iter = max.iter$

---

and requires the specification of the "trimming ratio", i.e. the fraction α of the data to be trimmed. Then, all non-trimmed time series are classified according to the QAF-FCMdC model.

Given the trimming ratio α, the QAF-TrFCMdC method can be formalized as the following minimization problem:

$$\begin{cases} \min_{\boldsymbol{Y}, \Omega} \sum_{i=1}^{H(\alpha)} \sum_{c=1}^{C} u_{ic}^m \left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|^2 \\ \text{subject to: } \sum_{c=1}^{C} u_{ic} = 1 \text{ and } u_{ic} \geq 0. \end{cases} \tag{11}$$

where $\boldsymbol{Y}$ ranges on all the subsets of $\boldsymbol{\Gamma}$ of size $H(\alpha) = [n(1 - \alpha)]$. Notice that if $\alpha = 0$, then none of the series is trimmed away from the process and the standard QAF-FCMdC model is obtained.

Just as in the QAF-FCMdC model, the local optimal solution for the estimation of the membership degrees u_{ic} is given by (4), with i ranging in the subset of the non-trimmed series and $c = 1, \ldots, C$. By replacing the expression of the u_{ic} in the objective function (11), we have

$$\sum_{i=1}^{H(\alpha)} \left[\sum_{c=1}^{C} \left(\left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|_2^2 \right)^{1/(1-m)} \right]^{1-m} = \sum_{i=1}^{H(\alpha)} h_i, \tag{12}$$

where

$$h_i = \left[\sum_{c=1}^{C} \left(\left\| \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(c)} \right\|_2^2 \right)^{1/(1-m)} \right]^{1-m}. \tag{13}$$

Algorithm 4 QAF-based Trimmed Fuzzy C-Medoids Clustering (QAF-TrFCMdC)

1: Fix C, m, α and *max.iter*
2: Set $iter = 0$
3: Pick the initial medoids $\widetilde{\boldsymbol{\Gamma}} = \left\{ \widetilde{\boldsymbol{\Gamma}}^{(1)}, \ldots, \widetilde{\boldsymbol{\Gamma}}^{(C)} \right\}$
4: **repeat**
5:     Identify the subset $\boldsymbol{Y}$ formed by the $H(\alpha) = [n(1-\alpha)]$ series closest to the
        medoids, i.e. minimizing $\sum_{i=1}^{H(\alpha)} h_{i:n}$
6:     Set $\widetilde{\boldsymbol{\Gamma}}_{\text{OLD}} = \widetilde{\boldsymbol{\Gamma}}$ {Store the current medoids}
7:     Compute u_{ic}, $i = 1, \ldots, n$, $c = 1, \ldots, C$, using (4)
8:     For each $c \in \{1, \ldots, C\}$, determine the index $j_c \in \{1, \ldots, n\}$ satisfying:
$$j_c = \text{argmin}_{1 \leq j \leq n} \sum_{i=1}^{H(\alpha)} u_{ic}^m \left\| \boldsymbol{\Gamma}^{(i)} - \boldsymbol{\Gamma}^{(j)} \right\|^2$$
9:     **return** $\widetilde{\boldsymbol{\Gamma}}^{(c)} = \boldsymbol{\Gamma}^{(j_c)}$, for $c = 1, \ldots, C$ {Update the medoids}
10:    $iter \leftarrow iter + 1$
11: **until** $\widetilde{\boldsymbol{\Gamma}}_{\text{OLD}} = \widetilde{\boldsymbol{\Gamma}}$ or $iter = max.iter$

Therefore, the objective function for the QAF-TrFCMdC model takes the form

$$\sum_{i=1}^{H(\alpha)} h_{i:n}, \tag{14}$$

where $h_{i:n}$ represents the ith item when h_i, $i = 1, \ldots, n$, are arranged in ascending order.

Notice that, following Krishnapuram et al. (2001), the objective function is defined by considering the Least Trimmed Squares approach. The value of $H(\alpha) < n$ is chosen depending on how many series we would like to eliminate in the clustering process. This allows the clustering procedure to ignore outlier time series while minimizing the objective function. For instance, when $H(\alpha) < n/2$, 50% of the time series are not taken into account in the clustering process, and the objective function is minimized when we pick C medoids in such a way that the sum of harmonic mean (squared) Euclidean distance of 50% of the time series is as small as possible (Krishnapuram et al. 1999, 2001). For further details, see D'Urso and De Giovanni (2014) and D'Urso et al. (2017b).

The QAF-based trimmed fuzzy C-medoids clustering (QAF-TrFCMdC) is implemented as outlined in Algorithm 4.

2.6 Some general remarks

In this section we enumerate some general remarks connected to the clustering methods previously described.

Remark 1 (Local optima) The different algorithms, namely Algorithms 1, 2, 3, and 4, fall in the category of Alternating Cluster Estimation paradigm (Runkler and Bezdek 1999), and it is not guaranteed that the global minimum is reached. Thus, more than one random start is suggested to obtain a stable optimal solution.

Remark 2 (Selection of m) The fuzziness parameter m plays a relevant role in fuzzy clustering. It should be chosen in advance. Although $1 < m < \infty$, values too close to 1 will result in a partition with all membership close to 0 or 1. Excessively large values will lead to disproportionate overlap with all memberships close to $1/C$ (Wedel and Kamakura 1998). Consequently, neither of these types of m is recommended (Arabie et al. 1981). Although there have been some empirical heuristic procedures to determine the value of m (McBratney and Moore 1985; Wedel and Kamakura 1998; D'Urso 2015), there seems to exist no theoretically justifiable manner of selecting m. Usually, $m = 2$ is the most popular choice in fuzzy clustering (Bezdek 1981; Hruschka 1986; Wedel and Kamakura 1998; Yang and Wu 2004; Hwang et al. 2007). Kamdar and Joshi (2000) remarked that "since the medoids always have a membership of 1 in the cluster, raising its membership to the power m has no effect. Thus, when m is high, the mobility of the medoids may be lost [...]. For this reason, a value between 1 and 1.5 for m is recommended". Based on these arguments, the values 1.3, 1.5 and 2 have been considered for m in our simulation study and applications to real data sets. For a discussion and detailed list of references on the choice of m see D'Urso (2015).

Remark 3 (Cluster validation) For selecting the optimal number of clusters we adopt the cluster validity criteria described in Sect. 5, namely the criteria proposed by Xie and Beni (1991) and Kwon (1998).

Remark 4 (Dissimilarity measure) Notice that following the literature on time series clustering based on PAM approach (e.g. D'Urso et al. 2015, 2016, 2017b), the squared Euclidean distance was considered in the QAF-based fuzzy clustering methods (3), (5), (8) and (11). More in general, in a fuzzy framework, we could consider any kind of dissimilarity measure $r\left(\boldsymbol{\Gamma}^{(i)}, \widetilde{\boldsymbol{\Gamma}}^{(c)}\right)$ keeping the properties of the PAM approach (Krishnapuram et al. 1999, 2001).

3 Advantages of the proposed robust time series clustering methods

The proposed clustering methods inherit all the advantages of the theoretical and methodological approaches adopted in the clustering process, namely the fuzzy paradigm, the Partitioning Around Medoids (PAM) technique, the QAF-based distance and the robust clustering. In this section, the appealing properties of these approaches are briefly outlined in order to motivate our approach.

3.1 Advantages connected to fuzzy paradigm

By adopting the fuzzy approach, our three clustering proposals consider a conceptually suitable framework to deal with robustness. Unlike the hard clustering, the fuzzy approach provides valuable insight into the association level of the data objects to each cluster, which is particularly useful when a robust classification against outliers is intended. General advantages connected to the fuzzy approach are given below (Hwang et al. 2007).

1. The fuzzy clustering is attractive because it is easily compatible with distribution-free methods.
2. Due to the difficulty of identifying a clear boundary between clusters in real applications, fuzzy clustering appears more attractive than the hard (non-fuzzy) clustering methods (McBratney and Moore 1985; Wedel and Kamakura 1998).
3. The fuzzy clustering is computationally efficient (McBratney and Moore 1985; Heiser and Groenen 1997).
4. The memberships indicate whether there is a second-best cluster almost as good as the best cluster, a scenario which hard clustering methods cannot uncover (Everitt et al. 2001).

In addition, we have the following specific advantages connected to the clustering of time series (D'Urso 2005; Coppi et al. 2006):

5. Good sensitivity in capturing the dynamic characteristics of the time series. In many cases, since the evolutive behaviors of the time series are switching, the hard clustering methods are likely to miss this underlying structure. The switches can be treated by means of a fuzzy approach.
6. Good adaptivity in defining the prototype time series. This property is better appreciated when the observed time series do not differ too much from each other. In this case, the fuzzy definition of the clusters allows us to single out underlying structures, if these are likely to exist in the given set of time series.

3.2 Advantages connected to partitioning around medoids (PAM) approach

Adopting the PAM approach for clustering time series in a fuzzy manner, the prototypes of each cluster, henceforth medoid time series, are time series actually observed and not "virtual" time series like the "centroids" derived with a fuzzy c-means (Bezdek 1981). Overall, having non-fictitious representative series available makes interpreting the obtained clusters easier, which is often very useful in applications. In fact, "in many clustering problems one is particularly interested in a characterization of the clusters by means of typical or representative objects [time series]. These are objects [time series] that represent the various structural aspects of the set of objects [time series] being investigated. There can be many reasons for searching for representative objects [time series]. Not only can these objects [time series] provide a characterization of the clusters, but they can often be used for further work or research, especially when it is more economical or convenient to use a small set of k objects [C time series in our case] instead of the large set one started off with" (Kaufman and Rousseeuw 2009).

3.3 Advantages connected to feature-based clustering approach and to the use of quantile autocovariances

As the real interest of clustering relies on grouping series with similar dependence structures, the feature- and the model-based approaches are expected to produce good results. The latter require specifying the underlying models and then performing proper fits, which may be restrictive and computationally intensive. By contrast, the feature-

based approach is model-free since it is aimed at measuring a conventional distance between extracted low-dimensional feature vectors. No restrictive assumptions on the generating processes are imposed, although general regularity conditions (e.g. stationarity) are usually required. As result of it, the feature-based approach provides greater flexibility to treat with a range of generating processes. Also, working with feature vectors involves a substantial dimension reduction, which reports additional advantages such as applicability to unbalanced time series and low computational cost. Indeed, these properties are particularly attractive when robustness against atypical dynamic structures is required.

The quantile autocovariances were selected as extracted features due to their enormous potential to discriminate between many kinds of processes. By definition, quantile autocovariances take advantage of the robustness inherent to quantile methods, and hence they encompass a lot of appealing properties, including robustness to the non-existence of moments, treating properly with heavy tailed marginal distributions, detecting nonlinear features and changes in conditional shapes, among others (Lee and Rao 2012; Hagemann 2013; Li 2014; Dette et al. 2015; Kley et al. 2016).

3.4 Advantages connected to robust clustering approach

Unlike the classic fuzzy approach, the three proposed time series clustering methods consider different criteria specifically designed to manage the presence of outlier series. In all cases, the aim is to alleviate the negative effects of outliers in the clustering process. The method based on the metric approach achieves its robustness with respect to outliers by taking into account a robust metric; the method based on the noise approach achieves its robustness with respect to outliers by introducing a noise cluster represented by a noise prototype; the method based on the trimmed approach achieves its robustness with respect to outliers by trimming away a certain fraction of time series.

4 Simulation study

This section reports some results from a broad simulation study conducted to evaluate the clustering performance and accuracy of the proposed methods compared with standard procedures and other robust models based on different metrics. The numerical experiments have been conducted to examine two kinds of robustness. In Sect. 4.1, we focus on robustness against deviations from the generating models, i.e. outliers are time series generated from processes different from the ones defining the clusters. In Sect. 4.2, we pose a much more complex scenario where outliers are series generated from one cluster but distorted during a short or long period of time. As emphasized in Sect. 2, all the simulated scenarios involve independent time series.

4.1 Assessing robustness to the generating models

To gain insight into robustness to the generating models, simulation scenarios considering different time series setups were recreated, namely scenarios involving linear,

non-linear and conditionally heteroskedastic models. At each of these setups, we start with a base scenario formed by two well-separated clusters C_1 and C_2 including four independent time series each, and then the base scenario is successively contaminated with the presence of one and two outlier time series (O_1 and O_2). The specific scenarios and the generation schemes for each scenario are described below.

Clustering of linear models

L.1 Four time series simulated from each of the AR(1) model $X_t = 0.5X_{t-1} + \varepsilon_t$ (cluster C_1) and the MA(1) model $X_t = \varepsilon_t - 0.5\varepsilon_{t-1}$ (cluster C_2).

L.2 The base scenario L.1 plus one outlier time series O_1 simulated from a Gaussian white noise process.

L.3 The scenario L.2 and an additional outlier time series O_2 simulated from the ARMA(1,1) model $X_t = -0.9X_{t-1} + \varepsilon_t + 0.3\varepsilon_{t-1}$.

Clustering of non-linear models

NL.1 Four time series simulated from an exponential autoregressive model of the form

$$X_t = \left(0.3 - 10\exp(-X_{t-1}^2)\right) X_{t-1} + \varepsilon_t \, (\text{cluster } C_1),$$

and four time series simulated from the bilinear model given by

$$X_t = 0.6X_{t-1} - 0.5\varepsilon_{t-1}X_{t-1} + \varepsilon_t \, (\text{cluster } C_2).$$

NL.2 The base scenario NL.1 plus one outlier time series O_1, which consisted of one realization from the non-linear autoregressive model given by

$$X_t = 0.3|X_{t-1}|(3 + |X_{t-1}|)^{-1} + \varepsilon_t.$$

NL.3 The scenario NL.2 plus an additional outlier time series O_2 generated from the non-linear moving average model given by

$$X_t = -0.1\varepsilon_{t-1} + 0.3\varepsilon_{t-1}^2 + \varepsilon_t.$$

Clustering of conditional heteroskedastic models

CH.1 A base scenario formed by two clusters with four time series each generated from ARCH processes $X_t = \sigma_t\varepsilon_t$, where $\sigma_t^2 = 0.1 + \phi X_{t-1}^2$ with $\phi = 0.05$ for cluster C_1 and $\phi = 0.95$ for cluster C_2.

CH.2 The base scenario CH.1 plus one outlier time series O_1 simulated from an exponential GARCH model where the conditional variance is modeled by

$$\ln(\sigma_t^2) = 0.1 + 0.3\varepsilon_{t-1} + 0.7\left[|\varepsilon_{t-1}| - \mathbb{E}(|\varepsilon_{t-1}|)\right].$$

CH.3 The scenario CH.2 plus a second outlier time series O_2 simulated from a GJR-GARCH model of the form

$$\sigma_t^2 = 0.1 + \left[0.1 + 0.6\,I\,(X_{t-1} < 0)\right]X_{t-1}^2 + 0.1\sigma_{t-1}^2.$$

In all cases, the error process ε_t consisted of iid variables following a zero-mean Gaussian distribution with unit variance. All the selected models are strictly stationary. In general, it is not simple to check whether a non-linear time series is strictly stationary. The common practice is to represent the series as a vector-valued Markov chain and to establish the geometrical ergodicity of the induced Markov chain. Then, strict stationarity follows from the fact that an ergodic Markov chain is strictly stationary (Theorem 2.2 in Fan and Yao 2005). Considering mainly this approach, several authors have established constraints on the coefficients of different non-linear models under which a stationary solution is reached. In particular, results by Amendola and Francq (2009) for EXPAR(1) models, Pham and Tran (1981) for bilinear models, An and Huang (1996) for non-linear autoregressive models, Chen et al. (2011) for GJR-GARCH models, and Fan and Yao (2005) for ARCH, GARCH and EGARCH models, ensure stationarity for the specific models forming our simulation scenarios.

To bring insight into the shapes of the true quantile autocovariance functions for the examined models, plots of large sample approximations to these functions were obtained. Specifically, one hundred series of size 1000 were generated from each model and the corresponding sample quantile autocovariances averaged over the 100 replicates. For each $\tau_1 \in \{0.1, 0.5, 0.9\}$, plots of the points $\left\{\hat{\gamma}_1\left(\tau_1, \tau_2\right), \tau_2 = 0.05\, i, \text{ for } i = 1, \ldots, 19\right\}$ joined by lines are shown in Fig. 1.

Plots in Fig. 1 illustrate the capability of the quantile autocovariances to discriminate between the underlying processes. For the linear and non-linear scenarios (Fig. 1a, b, respectively), the theoretical patterns characterizing clusters and outliers exhibit very different curves of quantile autocovariances. As far as the heteroskedastic scenario (Fig. 1c), discrimination between clusters and outliers is also evident if a joint assessment of the plots over the three quantile levels is carried out.

Note that two different lengths of series were considered, namely $T = 150$ and $T = 250$ for the linear and non-linear models, and $T = 1500$ and $T = 2500$ for the case of conditionally heteroskedastic series. Larger realizations are necessary with heteroskedastic models in order to estimate the quantile autocovariances with higher accuracy. Indeed, this limitation also affects other metrics considered in this setup. For instance, estimates for the ARCH/GARCH coefficients are required to measure discrepancy between fitted models, and poor clustering results are obtained if small sample sizes are used due to the high variability of the estimated ARCH/GARCH parameters (Aielli and Caporin 2013).

Another graphical way to visualize both the spatial structure of the generating models and the separability between groups is to perform a metric multidimensional scaling (MDS) based on the pairwise QAF-dissimilarity matrix. For each scenario, 50 and 20 time series were generated from each of the models defining the clusters and the outliers, respectively. The reason to generate 40 outliers was simply to have available a reasonable number of realizations from the underlying process, which enables us to examine how they are spread out across the MDS plot. The two-dimensional scaling based on these realizations was carried out and the corresponding coordinate matrices are displayed in Fig. 2. The stress measures were also calculated to assess the quality of the embeddings, obtaining 2.45% (linear), 3.37% (non-linear) and 13.95% (heteroskedastic) for the small sample sizes, and 2.00% (linear), 2.65% (non-linear) and 8.74% (heteroskedastic) for the large sample sizes. Therefore, regardless of the
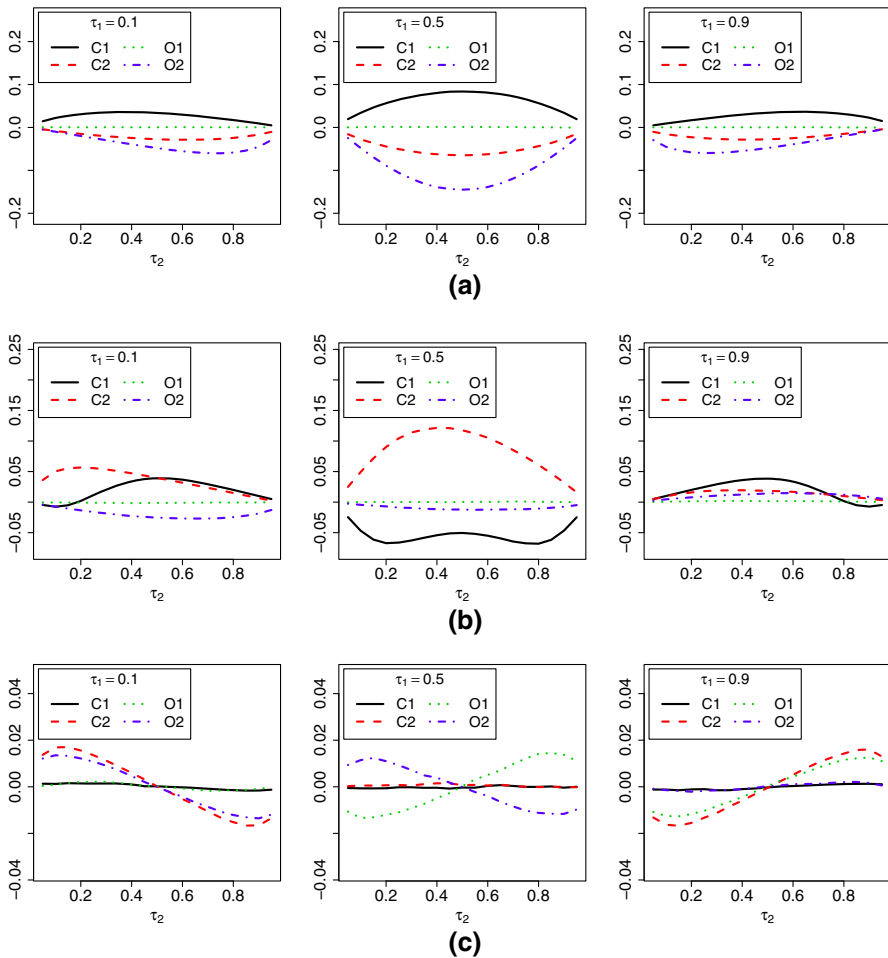
Fig. 1 Large sample approximation of the quantile autocovariances of lag one, $\hat{\gamma}_1 (\tau_1, \tau_2)$, for the models in the linear (**a**), non-linear (**b**) and heteroskedastic (**c**) scenarios. (Color figure online)

sample size, the results are below 5% with linear and non-linear processes, which is considered as an excellent goodness of fit (Dugard et al. 2010, p. 275). In the heteroskedastic case, higher values are generated but always below 15%, the threshold above which the MDS representation is unsatisfactory according to different authors.

The MDS coordinates in Fig. 2 show that the series forming the clusters $\mathcal{C}_1$ (red) and $\mathcal{C}_2$ (black) are grouped into two compact and well-separated clusters, while the outlier time series $\mathcal{O}_1$ (green) and $\mathcal{O}_2$ (blue) tend to be placed at an intermediate location between the clusters, except for the linear scenario where the second outlier, $\mathcal{O}_2$, is situated closer to cluster $\mathcal{C}_2$. Note that the non-linear models selected to generate outlier realizations produce overlapping clusters, while the linear and heteroskedastic models lead to separated groups, although also reasonably equidistant from $\mathcal{C}_1$ and $\mathcal{C}_2$ in the heteroskedastic scenarios. In short, Fig. 2 reveals that the QAF metric should
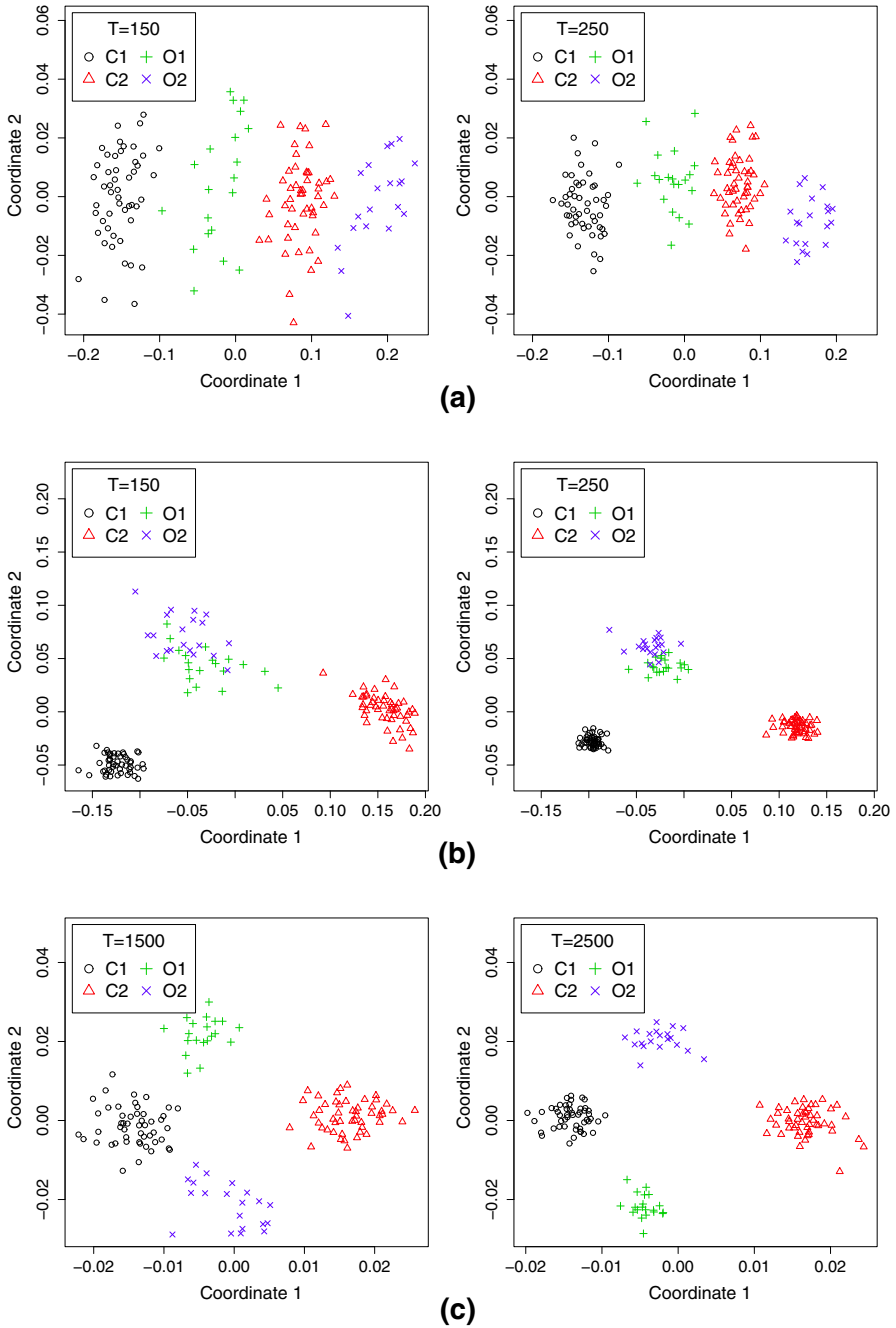
Fig. 2 Two-dimensional scaling configurations based on the QAF distance from the simulated linear (**a**), non-linear (**b**) and heteroskedastic (**c**) models. Three quantiles of levels 0.1, 0.5 and 0.9 and only the lag 1 were considered to compute the QAF distance

provide a useful approach to discriminate between the considered clusters and to detect the outlier time series. As expected, by increasing the length of the time series the gap between groups is more pronounced and, therefore, it will be easier to discriminate between them.

To assess the effectiveness of the proposed approaches in presence of outliers, each simulated dataset was subjected to clustering using the QAF-based fuzzy C-medoids clustering model (QAF-FCMdC) and the robust versions QAF-FCMdC-Exp, QAF-FCMdC-NC and QAF-TrFCMdC. The performance of the QAF metric was also examined by comparison with fuzzy C-medoids algorithms using other distances between fitted models. For the scenarios including ARMA models, we consider the AR distance introduced by Piccolo (1990), which computes the Euclidean distance between estimated coefficients of truncated AR(∞) representations. This way, the fuzzy C-medoids clustering model based on the AR metric, AR-FCMdC (D'Urso et al. 2013c), and the corresponding robust versions AR-FCMdC-Exp (D'Urso et al. 2015), AR-FCMdC-NC (D'Urso et al. 2013b) and AR-TrFCMdC (D'Urso et al. 2017b) were carried out in Scenarios L.1, L.2 and L.3. Analogously, a metric based on the autoregressive representations of GARCH(p,q) processes was employed with the heteroskedastic models. More precisely, a GARCH(p,q) process satisfies $X_t = \sigma_t \varepsilon_t$, where the innovations ε_t are iid variables and the squared conditional variance σ_t^2 follows an ARMA(p,q) model with parameters $(\delta, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$. It can be shown that

$$X_t^2 = \delta + \sum_{i=1}^{p^\star} (\alpha_i + \beta_i) X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \eta_{t-j} + \eta_t, \tag{15}$$

with $p^\star = \max(p, q)$, $\alpha_i = 0$ for $i > p$, $\beta_i = 0$ for $i > q$, and $\eta_t = X_t^2 - \sigma_t^2$ a zero-mean error uncorrelated with the past. Equation (15) establishes an ARMA($p^\star$,q) representation for X_t^2, which can be approximated by an AR(∞) structure with autoregressive coefficients π_u^G given by

$$\pi_u^G = (\alpha_u + \beta_u) + \sum_{j=1}^{\min(q,u)} \beta_j \pi_{u-j}^G \tag{16}$$

where $\pi_0^G = -1$, $\alpha_u = 0$ for $u > p$, and $\beta_u = 0$ for $u > q$. At this point, the GARCH distance is defined by the Euclidean distance between estimators of these new autoregressive coefficients. Based on the GARCH distance, the counterpart fuzzy algorithms GARCH-FCMdC (D'Urso et al. 2013c), GARCH-FCMdC-Exp, GARCH-FCMdC-NC and GARCH-TrFCMdC were carried out in the Scenarios CH.1, CH.2 and CH.3. In sum, we examined the performance of competitors using tailor-made distances for the lineal and heteroskedastic scenarios. Unlike the QAF-based models, it is expected that these model-based approaches get worse in case of model misspecification. However, their use in proper scenarios provide us valuable insight into the robustness of the QAF distance against the generating processes.

According to our clustering aim, the performance and accuracy of each algorithm is evaluated in terms of the percentage of times in which the series generated from the same process are grouped together in the same cluster, with membership degrees close to one for that cluster. Robustness in presence of outliers is examined by analysing the effect of the anomalous series on the membership degrees in the final partition, and also by reporting the percentage of times that the outliers are identified when the noise cluster and the trimmed models are used.

The number of clusters was set correctly at $C = 2$. For each of the nine scenarios, 10 sets of 100 simulations were carried out and subjected to fuzzy clustering with the described algorithms. For each of these 100 trials, the percentage of times that all the series were correctly classified was computed, and then the average percentage of correct classification over the 10 sets was taken as measure of clustering accuracy of the algorithm.

Due to the fact that we are dealing with fuzzy models, it was necessary to specify cut-off values to decide when a specific realization was assigned to a particular cluster. In the baseline scenarios, with no anomalous series, the ith time series is assigned to the cth cluster if its fuzzy membership degree is $u_{ic} > 0.6$. In the scenarios with data contaminated with outliers, the anomalous series were identified following different criteria according to the employed model. By using the noise cluster models, an outlier is correctly classified when it is assigned to the noise cluster. Here the model actually manages three clusters so that we decided to place an outlier within the noise cluster if $u_{ic_{NC}} > 0.5$, with c_{NC} denoting the index of the noise cluster. Our results revealed that a smaller threshold would even provide better results. By performing the standard fuzzy algorithms and the robust versions based on the exponential metric, we assume that the algorithm correctly handles the outliers when their membership degrees are reasonably similar for the two clusters, specifically both of them belonging to the (0.3, 0.7) interval. Lastly, in the case of the trimmed fuzzy, we checked whether the true outliers are trimmed units in the process. It is worthy remarking that these criteria and the selected cut-off values are compatible with the recommendations suggested in the literature (see e.g. D'Urso et al. 2013b, 2015).

In our experiments, three quantiles of levels 0.1, 0.5 and 0.9 and only one lag ($L = 1$, with $l_1 = 1$) were considered to compute the fuzzy algorithms based on the QAF dissimilarity. Certainly, increasing the number of quantiles does not mean an important cost in terms of computing time due to the computational efficiency of the QAF metric. Nevertheless, it was observed that three quantiles were enough to provide satisfactory results. To compute the AR and the GARCH distances, the order of the truncated AR(∞) approximations was determined by the AIC.

In all scenarios, we perform the fuzzy clustering models for several values of the fuzziness parameter m, which has a great influence in the clustering results. While small values of m, close to one, result in partitions with a low level of fuzziness that is with membership degrees close to 1 and 0, large values of m increase the amount of overlapping and the membership degrees are more homogeneously spread across the clusters. Using $m = 1.5$ or $m = 2$ are two popular choices in the literature but, to our knowledge, a theoretically justifiable optimality criterion to select m has not been provided yet. In our experience, high values of m, let us say $m \geq 2$, result in a poor clustering behavior when dealing with the noise cluster based algorithms (this point

is discussed later). Based on the previous considerations and also the arguments in Remark 2, we decided to use the values $m = 1.3, 1.5$ and 2.

As already mentioned, suitable choices of the parameters λ and β are also essential to reach satisfactory results. In fact, it was observed that the optimal selection of these parameters clearly depends on the value considered for m. Therefore, we proceeded to execute our simulations over a range of equally spaced values of λ and β, and the parameters retained were the ones maximizing the percentage of correct classification for each m. All the results reported hereafter correspond to this optimal selection of inputs for the algorithms. This way, we intend to perform fair comparisons, free of the effect of an inappropriate selection of the parameters.

The average percentages of correct classification obtained with the different models in the linear scenarios are shown in Table 1.

As expected, the standard algorithms show a very good behavior in Scenario L.1 without outliers. The two clusters are well-separated and both AR and QAF metrics are able to correctly classify all the series. Also the robust versions FCMdC-Exp and FCMdC-NC work fine in this setup. Adding outlier times series fairly has a disruptive effect on the results, which is clearly more pronounced with two outliers. In particular, AR-FCMdC and QAF-FCMdC present unsatisfactory success percentages for the three values of m, specially in Scenario L.3 where they always failed at correctly identifying both outliers. Actually the non-anomalous series are always well-classified and the failures are caused by the outliers, which are seldom identified. For this reason the best results are reached for the highest value of m, since high values for m imply softer boundaries between clusters, and hence the memberships assigned to the outliers are closer to 0.5. To illustrate these assertions, we have randomly selected one set of 100 trials from the Scenario L.3 and calculated the means and standard deviations of the membership degrees for $m = 2$ and $T = 250$ (the most favorable scenario). The results are displayed in Table 2. It is observed that the eight non-atypical series are always well-grouped. $\mathcal{O}_1$ present average memberships (highlighted in bold) very close to the cut-off values (0.3 and 0.7) and standard deviations large enough to account for a non-negligible number of failures. As far as $\mathcal{O}_2$, it is always assigned to cluster $\mathcal{C}_2$, which is consistent with the plot in the right panel of Fig. 2a, where $\mathcal{O}_2$ realizations are located closer to cluster $\mathcal{C}_2$. Thus, the standard versions erroneously assign the outlier $\mathcal{O}_2$ into the cluster $\mathcal{C}_2$. Smaller values of m led to average memberships higher (lower) than 0.7 (0.3), thus generating worse results.

Regardless of the considered distance, the robust versions based on the exponential metric and the trimmed approach substantially outperform the standard models. With one outlier (L.2) and realizations of length $T = 250$, both models produced excellent success rates, between 96.9 and 99.9%. The results were somewhat worse with two outliers (Scenario L.3) but also satisfactory, particularly using the QAF distance (scores always above 93.2% and 96.4% with QAF-FCMdC-Exp and QAF-TrFCMdC, respectively). The averages and standard deviations of the membership degrees highlighted in italic in Table 2 corroborate the high capability of QAF-FCMdC-Exp to identify the outlier time series. It is also remarkable that the robust QAF-based models performed somewhat better than the AR-based ones despite handling ARMA models. For instance, Fig. 3 shows the evolution of the percentages of correct classification for AR-FCMdC-Exp and QAF-FCMdC-Exp as function of β in Scenario L.3 with

Table 1 Average percentages of correct classification for the simulated linear scenarios

Model	Scenario L.1: no outliers			Scenario L.2: 1 outlier			Scenario L.3: 2 outliers		
	$m = 1.3$	$m = 1.5$	$m = 2$	$m = 1.3$	$m = 1.5$	$m = 2$	$m = 1.3$	$m = 1.5$	$m = 2$
$T = 150$									
AR-based									
AR-FCMdC	100.0	100.0	100.0	20.7	37.4	77.4	0.0	0.0	0.0
AR-FCMdC-Exp	100.0	100.0	99.7	76.2	81.4	88.7	58.2	61.2	68.0
AR-FCMdC-NC	100.0	100.0	99.5	55.7	53.7	33.9	41.3	39.7	26.8
AR-TrFCMdC	–	–	–	68.5	66.0	54.5	57.1	55.4	52.0
QAF-based									
QAF-FCMdC	100.0	100.0	100.0	14.4	28.3	58.3	0.0	0.0	0.0
QAF-FCMdC-Exp	100.0	100.0	100.0	86.5	85.7	90.3	77.3	79.0	84.0
QAF-FCMdC-NC	100.0	100.0	99.9	75.1	74.2	61.4	66.4	65.0	52.8
QAF-TrFCMdC	–	–	–	88.7	89.0	86.6	84.4	83.9	80.9
$T = 250$									
AR-based									
AR-FCMdC	100.0	100.0	100.0	21.8	42.7	81.4	0.0	0.0	0.0
AR-FCMdC-Exp	100.0	100.0	100.0	99.7	99.7	99.9	83.5	86.4	89.3
AR-FCMdC-NC	100.0	100.0	100.0	99.2	99.2	98.0	71.2	69.9	57.2
AR-TrFCMdC	–	–	–	99.3	99.3	99.8	84.7	85.0	81.7
QAF-based									
QAF-FCMdC	100.0	100.0	100.0	20.9	38.8	71.1	0.0	0.0	0.0
QAF-FCMdC-Exp	100.0	100.0	100.0	96.9	97.2	97.9	93.2	93.4	94.9
QAF-FCMdC-NC	100.0	100.0	100.0	97.4	97.1	94.3	88.5	88.9	84.3
QAF-TrFCMdC	–	–	–	99.2	99.5	99.6	96.4	96.6	96.9

Table 2 Mean and standard deviation (in brackets) of membership degrees computed from one randomly selected set of 100 trials in Scenario L.3, with $T = 250$ and $m = 2$

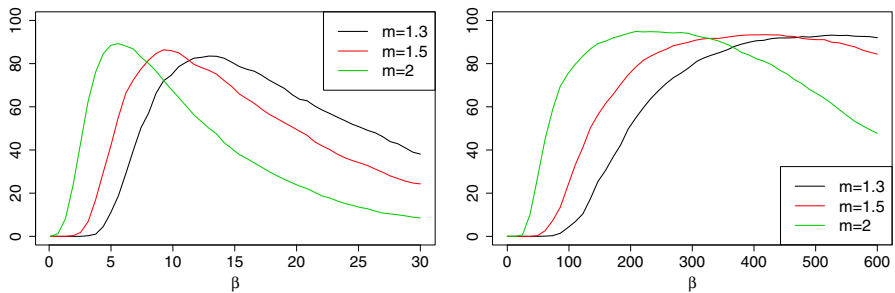|	QAF-FCMdC		QAF-FCMdC-Exp		QAF-FCMdC-NC		
	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$	NC
$X_1 (\in \mathcal{C}_1)$	0.982	0.018	0.882	0.118	0.813	0.012	0.175
	(.015)	(.015)	(.089)	(.089)	(.147)	(.009)	(.138)
$X_2 (\in \mathcal{C}_1)$	0.981	0.019	0.877	0.123	0.796	0.013	0.190
	(.021)	(.021)	(.148)	(.148)	(.154)	(.012)	(.142)
$X_3 (\in \mathcal{C}_1)$	0.982	0.018	0.866	0.134	0.819	0.012	0.169
	(.018)	(.018)	(.147)	(.147)	(.142)	(.010)	(.133)
$X_4 (\in \mathcal{C}_1)$	0.981	0.019	0.866	0.134	0.784	0.014	0.202
	(.018)	(.018)	(.142)	(.142)	(.158)	(.012)	(.147)
$X_5 (\in \mathcal{C}_2)$	0.024	0.976	0.127	0.873	0.014	0.802	0.185
	(.025)	(.025)	(.159)	(.159)	(.012)	(.153)	(.142)
$X_6 (\in \mathcal{C}_2)$	0.024	0.976	0.125	0.875	0.013	0.812	0.175
	(.022)	(.022)	(.130)	(.130)	(.010)	(.133)	(.123)
$X_7 (\in \mathcal{C}_2)$	0.020	0.980	0.127	0.873	0.011	0.837	0.152
	(.022)	(.022)	(.133)	(.133)	(.009)	(.131)	(.122)
$X_8 (\in \mathcal{C}_2)$	0.023	0.977	0.125	0.875	0.013	0.805	0.182
	(.021)	(.021)	(.137)	(.137)	(.010)	(.141)	(.131)
$X_9 (\equiv \mathcal{O}_1)$	**0.446**	**0.554**	*0.456*	*0.544*	0.139	0.231	***0.630***
	(.166)	**(.166)**	*(.062)*	*(.062)*	(.052)	(.087)	***(.051)***
$X_{10} (\equiv \mathcal{O}_2)$	**0.095**	**0.905**	*0.454*	*0.546*	0.024	0.238	***0.738***
	(.031)	**(.031)**	*(.050)*	*(.050)*	(.002)	(.114)	***(.112)***

Fig. 3 Average percentage of correct classification as a function of β by using AR-FCMdC-Exp (left panel) and QAF-FCMdC-Exp (right panel) models in Scenario L.3 with $T = 250$. (Color figure online)

$T = 250$. Besides getting insight into the optimal values for β, Fig. 3 allows us to conclude that QAF-FCMdC-Exp is preferable to AR-FCMdC-Exp for the three values of m if a suitable choice of β, e.g. according to (7), is considered.

Table 3 Percentage of trials where the outliers were correctly trimmed by using AR-TrFCMdC and QAF-TrFCMdC in the linear scenarios L.2 and L.3

	Model	Scenario L.2	Scenario L.3	
		1 outlier	1 outlier	2 outliers
$T = 150$				
$m = 1.3$	AR-TrFCMdC	68.5	18.8	62.1
	QAF-TrFCMdC	88.7	8.6	76.5
$m = 1.5$	AR-TrFCMdC	66.0	20.0	59.3
	QAF-TrFCMdC	89.0	7.4	77.7
$m = 2.0$	AR-TrFCMdC	54.5	23.7	50.3
	QAF-TrFCMdC	86.6	7.8	74.0
$T = 250$				
$m = 1.3$	AR-TrFCMdC	99.3	5.3	89.3
	QAF-TrFCMdC	99.2	1.8	92.4
$m = 1.5$	AR-TrFCMdC	99.3	5.3	88.2
	QAF-TrFCMdC	99.5	1.7	92.5
$m = 2.0$	AR-TrFCMdC	99.8	11.9	80.9
	QAF-TrFCMdC	99.6	2.1	95.1

As far as the trimmed approach is concerned, Table 3 shows that the QAF distance was more efficient than the AR one in removing the true outlier time series in Scenarios L.2 and L.3.

The fuzzy models based on the noise cluster, AR-FCMdC-NC and QAF-FCMdC-NC, reported good results but worse than the ones obtained with the other robust algorithms. In particular, the percentage of success substantially decayed with $m = 2$ and in presence of two outliers. The reason is again that a more balanced distribution of the membership degrees occurs as m increases, thus making more difficult to assign the outliers to the noise cluster. For illustrative purpose only, let us briefly come back to Table 2. As required, the highest average memberships of the outliers with QAF-FCMdC-NC (highlighted in bolditalic) correspond to the noise cluster. Nevertheless they are not significantly greater than the cut-off value, 0.5, and therefore an important number of trials draw out erroneous classification. Likewise Fig. 3, we have depicted the evolution of the percentages of correct classification by using AR-FCMdC-NC and QAF-FCMdC-NC as function of λ in Fig. 4. The poor rates of correct classification with $m = 2$ are evident for all λ, thus concluding that the only way to improve the results is using a less stringent cut-off value. Comparison of the two panels in Fig. 4 also highlights the superiority of the QAF distance to develop the noise cluster fuzzy model.

Lastly, it is worthy to mention that simulations with 25 series per cluster were executed to explore the effect of a larger number of series. In general, the average success percentages were only somewhat lower for the QAF-based models, while they worsened substantially with the AR-distance.
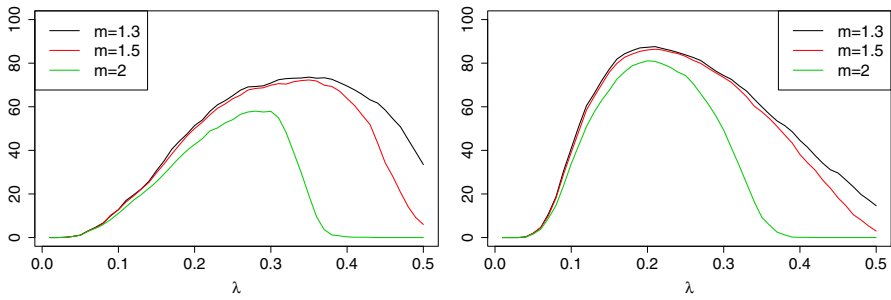
Fig. 4 Average percentage of correct classification as a function of λ by using AR-FCMdC-NC (left panel) and QAF-FCMdC-NC (right panel) models in Scenario L.3 with $T = 250$. (Color figure online)

As far as the scenarios NL.1, NL.2 and NL.3, including non-linear models, the most noticeable fact was the excellent performance showed by the QAF-based models. Although the models based on the AR distance were considered in our experiments, model misspecification heavily affected the results and they have been omitted. Table 4 reports the simulation results for the three non-linear scenarios using the QAF distance and Table 5 presents means and standard deviations of memberships for an arbitrary set of 100 trials in Scenario NL.3. In these new setups, the percentages of correct classification are higher than in the linear scenarios for all models and values of m, particularly by working with the shortest series ($T = 150$). The average percentage of times in which QAF-TrFCMdC trimmed the true outlier in Scenario NL.2 was always above 98.5% for $T = 150$ and 99.8% for $T = 250$, while in Scenario NL.3 the two true outliers were detected above 98.6 and 100% for $T = 150$ and $T = 250$, respectively. It is also noticeable the improvement of the results for the robust model based on the noise cluster. The average membership degrees reported in Table 5 for the outliers time series and graphs in Fig. 5 help us to understand this improvement.

Simulation results from the heteroskedastic scenarios CH.1, CH.2 and CH.3 based on the GARCH and QAF distances are shown in Tables 6, 7, and 8 and Figs. 6 and 7. As already mentioned, conditional heteroskedasticity induces a more complex scenario because of the simulated realizations from GARCH processes are characterized by high dispersion for small sample sizes (Aielli and Caporin 2013). Table 6 corroborates this feature since success rates comparable to the ones obtained in the linear and non-linear scenarios are only attained with $T = 2500$. It is worthy mentioning that these sample sizes are frequently considered in the literature by working with heteroskedastic processes. Notice also that the membership degrees for the non-anomalous series in Table 7 are moderately further from 0 and 1 than in previous analyses, thus emphasizing the major difficulty of clustering under heteroskedasticity. In fact, non-anomalous series were sometimes missclassified using the GARCH distance. This assertion is easily understood by comparing the outputs in Tables 6 and 8. It is observed in Table 8 that GARCH-TrFCMdC and QAF-TrFCMdC present similar percentages of success by trimming the true outliers, but in contrast QAF-TrFCMdC exhibits higher average percentages of correct classification in Table 6.
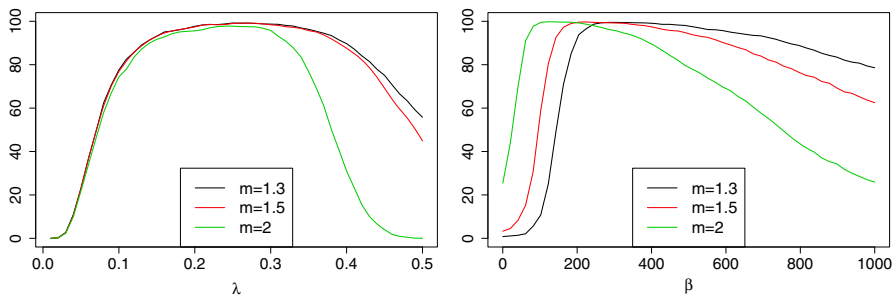
Again the main conclusion is that the QAF-based models fairly outperform the GARCH-based ones. While the latter are affected by the inaccurate estimation of

Table 4 Average percentages of correct classification for the simulated non-linear scenarios

Model	Scenario NL.1: no outliers			Scenario NL.2: 1 outlier			Scenario NL.3: 2 outliers		
	$m = 1.3$	$m = 1.5$	$m = 2$	$m = 1.3$	$m = 1.5$	$m = 2$	$m = 1.3$	$m = 1.5$	$m = 2$
$T = 150$									
QAF-based									
QAF-FCMdC	100	100	100	13.2	24.4	57.8	0.8	3.4	27.7
QAF-FCMdC-Exp	100	100	100	97.3	97.8	98.8	97.5	98.4	99.2
QAF-FCMdC-NC	100	100	100	95.6	95.0	90.7	95.4	95.0	88.4
QAF-TrFCMdC	–	–	–	99.5	99.5	98.5	99.6	99.5	98.6
$T = 250$									
QAF-based									
QAF-FCMdC	100	100	100	11.7	24.0	60.1	0.7	3.3	25.4
QAF-FCMdC-Exp	100	100	100	99.7	99.7	99.9	99.5	99.7	99.8
QAF-FCMdC-NC	100	100	100	99.6	99.6	97.9	99.2	99.2	97.8
QAF-TrFCMdC	–	–	–	100	99.9	99.8	100	100	100

Table 5 Mean and standard deviation (in brackets) of membership degrees computed from one randomly selected set of 100 trials in Scenario NL.3, with $T = 250$ and $m = 2$

	QAF-FCMdC		QAF-FCMdC-Exp		QAF-FCMdC-NC		
	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$	NC
$X_1 (\in \mathcal{C}_1)$	0.985	0.015	0.937	0.063	0.860	0.010	0.130
	(.014)	(.014)	(.057)	(.057)	(.129)	(.010)	(.119)
$X_2 (\in \mathcal{C}_1)$	0.985	0.015	0.928	0.072	0.844	0.011	0.145
	(.014)	(.014)	(.059)	(.059)	(.128)	(.010)	(.118)
$X_3 (\in \mathcal{C}_1)$	0.987	0.013	0.940	0.060	0.867	0.009	0.124
	(.013)	(.013)	(.052)	(.052)	(.120)	(.008)	(.112)
$X_4 (\in \mathcal{C}_1)$	0.985	0.015	0.936	0.064	0.852	0.010	0.138
	(.013)	(.013)	(.059)	(.059)	(.110)	(.007)	(.103)
$X_5 (\in \mathcal{C}_2)$	0.008	0.992	0.037	0.963	0.006	0.908	0.086
	(.008)	(.008)	(.038)	(.038)	(.006)	(.079)	(.073)
$X_6 (\in \mathcal{C}_2)$	0.008	0.992	0.037	0.963	0.006	0.920	0.074
	(.008)	(.008)	(.033)	(.033)	(.005)	(.066)	(.062)
$X_7 (\in \mathcal{C}_2)$	0.008	0.992	0.034	0.966	0.005	0.921	0.074
	(.008)	(.008)	(.030)	(.030)	(.005)	(.070)	(.065)
$X_8 (\in \mathcal{C}_2)$	0.007	0.993	0.028	0.972	0.005	0.931	0.064
	(.007)	(.007)	(.030)	(.030)	(.005)	(.077)	(.072)
$X_9 (\equiv \mathcal{O}_1)$	0.333	0.667	0.436	0.564	0.105	0.225	0.670
	(.099)	(.099)	(.039)	(.039)	(.026)	(.047)	(.029)
$X_{10} (\equiv \mathcal{O}_2)$	0.296	0.704	0.437	0.563	0.084	0.214	0.701
	(.079)	(.079)	(.026)	(.026)	(.017)	(.038)	(.028)

Fig. 5 Average percentage of correct classification in Scenario NL.3 with $T = 250$ for QAF-FCMdC-NC (left panel) and QAF-FCMdC-Exp (right panel) models as function of β and λ, respectively. (Color figure online)

the GARCH parameters, the former take advantage of the capability of the QAF distance to detect changes in conditional shapes and to deal with heavy-tailed marginal distributions. As in the above scenarios, the robust models, particularly QAF-FCMdC-Exp and QAF-TrFCMdC, led to the best results in presence of outliers regardless of

Table 6 Average percentages of correct classification for the simulated conditional heteroskedastic scenarios

Model	Scenario CH.1: no outliers			Scenario CH.2: 1 outlier			Scenario CH.3: 2 outliers		
	$m = 1.3$	$m = 1.5$	$m = 2$	$m = 1.3$	$m = 1.5$	$m = 2$	$m = 1.3$	$m = 1.5$	$m = 2$
$T = 1500$									
GARCH-based									
GARCH-FCMdC	63.0	63.0	62.8	11.8	22.8	43.1	2.4	5.8	21.7
GARCH-FCMdC-Exp	63.0	63.0	62.6	53.8	53.7	53.0	45.8	45.1	41.5
GARCH-FCMdC-NC	63.0	62.9	61.9	51.1	54.0	53.9	46	45.9	46.1
GARCH-TrFCMdC	–	–	–	58.3	58.0	57.7	50.5	50.5	50.5
QAF-based									
QAF-FCMdC	99.5	98.9	96.9	34.7	56.9	86.4	10.2	29.2	75.8
QAF-FCMdC-Exp	99.5	98.9	96.9	84.4	88.0	89.9	75.7	79.1	83.2
QAF-FCMdC-NC	99.5	98.6	85.7	69.4	59.2	17.3	54.0	42.4	11.9
QAF-TrFCMdC	–	–	–	87.5	85.5	79.0	76.4	73.7	66.4
$T = 2500$									
GARCH-based									
GARCH-FCMdC	69.8	69.8	69.5	20.3	33.8	58.1	3.0	10.8	34.1
GARCH-FCMdC-Exp	69.8	69.8	69.4	63.1	63.2	63.8	59.6	59.9	59.9
GARCH-FCMdC-NC	69.8	69.8	69.4	62.8	63.4	63.1	58.7	58.5	58.7
GARCH-TrFCMdC	–	–	–	66.7	66.7	66.7	63.1	63.1	63.1
QAF-based									
QAF-FCMdC	99.9	100.0	100.0	30.5	57.1	93.8	7.9	29.0	81.6
QAF-FCMdC-Exp	100.0	100.0	100.0	96.0	97.4	98.0	90.9	92.9	93.9
QAF-FCMdC-NC	99.9	100.0	98.4	90.7	87.4	54.9	78.2	73.2	40.6
QAF-TrFCMdC	–	–	–	98.8	98.2	97.1	95.4	95.0	91.9

Table 7 Mean and standard deviation (in brackets) of membership degrees computed from one randomly selected set of 100 trials in Scenario CH.3, with $T = 2500$ and $m = 2$

	QAF-FCMdC		QAF-FCMdC-Exp		QAF-FCMdC-NC		
	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$	NC
X_1 ($\in \mathcal{C}_1$)	0.888	0.112	0.813	0.187	0.781	0.088	0.130
	(.020)	(.020)	(.020)	(.020)	(.020)	(.016)	(.008)
X_2 ($\in \mathcal{C}_1$)	0.886	0.114	0.808	0.192	0.780	0.087	0.133
	(.023)	(.023)	(.015)	(.015)	(.015)	(.013)	(.006)
X_3 ($\in \mathcal{C}_1$)	0.884	0.116	0.808	0.192	0.773	0.087	0.139
	(.026)	(.026)	(.014)	(.014)	(.022)	(.019)	(.008)
X_4 ($\in \mathcal{C}_1$)	0.892	0.108	0.808	0.192	0.777	0.092	0.131
	(.021)	(.021)	(.016)	(.016)	(.028)	(.023)	(.011)
X_5 ($\in \mathcal{C}_2$)	0.125	0.875	0.218	0.782	0.097	0.750	0.153
	(.022)	(.022)	(.011)	(.011)	(.015)	(.022)	(.011)
X_6 ($\in \mathcal{C}_2$)	0.126	0.874	0.223	0.777	0.093	0.746	0.162
	(.022)	(.022)	(.016)	(.016)	(.016)	(.022)	(.012)
X_7 ($\in \mathcal{C}_2$)	0.127	0.873	0.221	0.779	0.098	0.735	0.167
	(.018)	(.018)	(.010)	(.010)	(.020)	(.022)	(.008)
X_8 ($\in \mathcal{C}_2$)	0.122	0.878	0.211	0.789	0.087	0.763	0.149
	(.024)	(.024)	(.017)	(.017)	(.015)	(.022)	(.009)
$X_9 \equiv \mathcal{O}_1$	0.573	0.427	0.516	0.484	0.302	0.222	0.476
	(.011)	(.011)	(.002)	(.002)	(.009)	(.005)	(.006)
$X_{10} \equiv \mathcal{O}_2$	0.565	0.435	0.517	0.483	0.313	0.233	0.454
	(.005)	(.005)	(.003)	(.003)	(.004)	(.003)	(.004)

the fuzziness parameter. In this case, the model based on the noise cluster showed worse results, specially in Scenario CH.3 with two outlier time series (see Table 7 and Fig. 7).

4.2 Assessing robustness against distorted individual time series

In this section we assess a different kind of robustness. Specifically, outliers are generated using the pattern of one cluster but they are distorted by structural changes during a more or less long-term period. Following this principle, four outlier types are commonly considered in the time series literature, namely *Additive Outliers* (AO), *Transitory Change* (TC) outliers, *Level Shift* (LS) outliers and *Innovative Outliers* (IO) (see Tsay 1986; Peña 2011, among others). An AO occurs when the series receives a sharp impact due to an unexpected event at a particular time point, which is reflected by an isolated spike in its temporal dynamic. For a TC outlier, the effect of the mentioned impact does not disappear immediately but it dies out gradually with time. Unlike AO and TC outliers, which exhibit transitory level changes, LS outliers present a permanent level change starting from the time point where the initial shock occurs

Table 8 Percentage of trials where the outliers were correctly trimmed by using GARCH-TrFCMdC and QAF-TrFCMdC in the heterokedastic scenarios CH.2 and CH.3

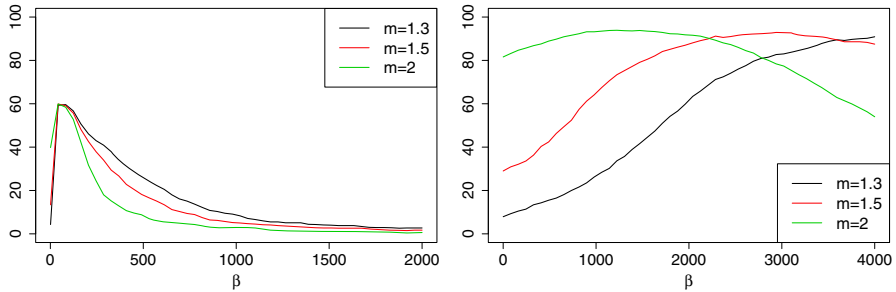|	Model	Scenario CH.2	Scenario CH.3	
		1 outlier	1 outlier	2 outliers
$T = 1500$				
$m = 1.3$	GARCH-TrFCMdC	86.6	9.6	70.4
	QAF-TrFCMdC	87.6	9.6	76.4
$m = 1.5$	GARCH-TrFCMdC	85.9	9.8	70.0
	QAF-TrFCMdC	86.0	10.2	73.7
$m = 2.0$	GARCH-TrFCMdC	85.4	9.2	70.3
	QAF-TrFCMdC	81.9	12.3	67.6
$T = 2500$				
$m = 1.3$	GARCH-TrFCMdC	93.8	5.7	81.8
	QAF-TrFCMdC	98.8	1.5	95.4
$m = 1.5$	GARCH-TrFCMdC	93.6	5.6	81.8
	QAF-TrFCMdC	98.2	1.7	95.0
$m = 2.0$	GARCH-TrFCMdC	93.1	5.3	82.2
	QAF-TrFCMdC	97.2	2.5	92.0

Fig. 6 Average percentage of correct classification as a function of β by using GARCH-FCMdC-Exp (left panel) and QAF-FCMdC-Exp (right panel) models in Scenario CH.3 with $T = 2500$. (Color figure online)

and continuing until the end of the observation period. Lastly, the model for an IO is based on contaminating the distribution of the noise of the underlying process. More precisely, the outliers are constructed as follows.

AO $\quad X_t^\star = X_t + \omega$ if $t = t_0$ and $X_t^\star = X_t$ otherwise

TC $\quad X_t^\star = X_t$ if $t < t_0$ and $X_{t_0+k}^\star = X_{t_0} + \delta^k \omega$, for $k = 0, 1, \ldots, T - t_0$

LS $\quad X_t^\star = X_t + \omega$, if $t \geq t_0$ and $X_t^\star = X_t$, if $t < t_0$

IO $\quad X_t^\star$ is generated as X_t but with innovations ε_t following a non-Gaussian distribution F_0
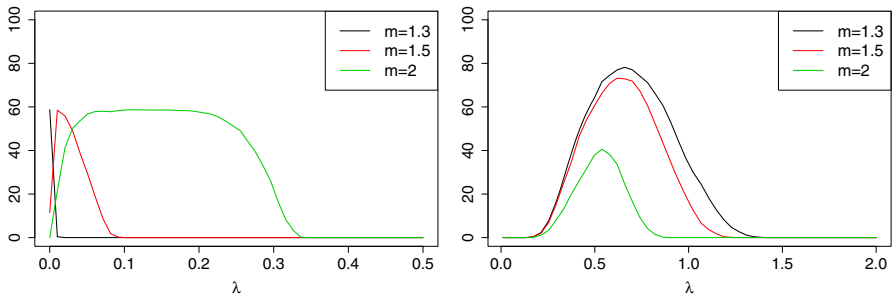
Fig. 7 Average percentage of correct classification as a function of λ by using GARCH-FCMdC-NC (left panel) and QAF-FCMdC-NC (right panel) models in Scenario CH.3 with $T = 2500$. (Color figure online)

where $X_t^\star$ denotes the contaminated series, X_t is an arbitrary realization coming from cluster $\mathcal{C}_1$, t_0 is the particular time point where the shock occurs, $\omega > 0$ quantifies the effect of the initial impact and $0 < \delta < 1$ regulates the propagation of this effect in subsequent observations. Note that AO and LS are limit scenarios of TC as δ tends to 0 and 1, respectively.

To examine robustness against these types of outliers, the simulation scheme in the above section has been performed with scenarios including one outlier from each type (AO, TC, LS and IO). Thus, for example, clustering of linear models is carried out in four different scenarios consisting of the base scenario L.1 plus one outlier of type AO (Scenario L.4), TC (L.5), LS (L.6) and IO (L.7). In the four scenarios, the outlier time series $X_t^\star$ is constructed by contaminating a realization X_t generated from the underlying process in cluster $\mathcal{C}_1$. This same scheme was considered for the setups involving non-linear and conditionally heteroskedastic models, thus recreating scenarios with outliers NL.4-NL.7 and CH.4-CH.7, respectively. The parameters required for defining the outliers were always $t_0 = T/2$, $\omega = 5$, $\delta = 0.9$ and F_0 the exponential distribution for innovations in the case of IO. The remaining simulation features (clustering models, cut-off values for fuzzy membership degrees, sample sizes, input parameters, …) are the ones used in Sect. 4.1. In particular, we use $m = 1.3$, 1.5 and 2, but for simplicity and clarity of presentation, we only show the results for $m = 1.3$ and 2. The average percentages of correct classification for the different models in the new scenarios are given in Tables 9, 10 and 11, for the linear, non-linear and conditionally heteroskedastic setups, respectively.

To summarize the main conclusions from the new experiments, we first focus on the success percentages with outliers generated by changes in the level (AO, TC, and LS). Overall, the success percentages are very high when the effect of the change is held until the end of the observation period (LS outliers). With temporary effects, i.e. for AO and TC outliers, the success percentages are low, which is expected because this is a very hard problem. In fact, regardless of the metric, the non-robust versions have never been able to detect AO and TC outliers in linear and non-linear setups (Tables 9 and 10). Only the high capability of the QAF-metric to discriminate between heteroskedastic models allowed to attain reasonable success rates with the standard fuzzy algorithm (Table 11). It is also observed that the percentages of correct classification tend to get worse by increasing the length of the series. This is also expected since the estimation

Table 9 Average percentages of correct classification for the simulated linear scenarios L.4, L.5, L.6 and L.7, formed by the base scenario L.1 plus one outlier of type AO (L.4), TC (L.5), LS (L.6) and IO (L.7)

Model	Scenario L.4: AO		Scenario L.5: TC		Scenario L.6: LS		Scenario L.7: IO	
	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$
$T = 150$								
AR-based								
AR-FCMdC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AR-FCMdC-Exp	5.0	7.3	5.8	9.0	32.4	45.4	3.6	5.4
AR-FCMdC-NC	3.4	2.1	4.1	3.0	19.6	15.4	2.4	1.9
AR-TrFCMdC	6.1	6.0	9.4	9.6	37.5	41.3	5.4	5.3
QAF-based								
QAF-FCMdC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
QAF-FCMdC-Exp	10.9	17.6	26.0	35.4	98.8	98.9	21.8	35.6
QAF-FCMdC-NC	5.8	4.5	12.8	9.4	98.9	96.8	10.6	6.8
QAF-TrFCMdC	11.5	11.3	27.6	29.7	99.7	99.9	25.2	26.2
$T = 250$								
AR-based								
AR-FCMdC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AR-FCMdC-Exp	3.4	5.7	5.3	8.3	44.3	58.2	3.2	4.9
AR-FCMdC-NC	1.7	1.4	3.6	2.8	27.9	22.8	2.4	1.8
AR-TrFCMdC	5.1	5.2	8.3	8.4	50.1	52.3	4.5	4.6
QAF-based								
QAF-FCMdC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
QAF-FCMdC-Exp	4.0	17.9	10.8	33.7	99.1	99.6	15.9	53.5
QAF-FCMdC-NC	4.9	3.7	9.9	8.3	99.4	99.3	19.4	15.5
QAF-TrFCMdC	11.7	11.5	24.1	24.3	99.9	99.9	38.7	40.2

Table 10 Average percentages of correct classification for the simulated non-linear scenarios NL.4, NL.5, NL.6 and NL.7, formed by the base scenario NL.1 plus one outlier of type AO (NL.4), TC (NL.5), LS (NL.6) and IO (NL.7)

Model	Scenario NL.4: AO		Scenario NL.5: TC		Scenario L.6: LS		Scenario NL.7: IO	
	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$
$T = 150$								
QAF-based								
QAF-FCMdC	0.0	0.0	0.0	0.4	0.0	0.0	31.7	83.2
QAF-FCMdC-Exp	4.6	6.4	37.9	47.7	72.6	80.7	96.8	98.1
QAF-FCMdC-NC	1.5	1.3	24.0	20.1	55.0	48.3	95.0	89.0
QAF-TrFCMdC	3.7	3.5	44.1	43.6	78.8	80.8	99.2	99.1
$T = 250$								
QAF-based								
QAF-FCMdC	0.0	0.0	0.0	0.0	0.0	0.0	40.4	93.7
QAF-FCMdC-Exp	1.7	5.8	23.7	38.6	92.3	95.2	99.9	100
QAF-FCMdC-NC	1.8	1.5	15.6	14.1	82.8	79.0	99.6	98.3
QAF-TrFCMdC	4.1	3.2	30.9	30.6	94.5	96.3	100	99.9

Table 11 Average percentages of correct classification for the simulated non-linear scenarios CH.4, CH.5, CH.6 and CH.7, formed by the base scenario CH.1 plus one outlier of type AO (CH.4), TC (CH.5), LS (CH.6) and IO (CH.7)

| Model	Scenario CH.4: AO		Scenario CH.5: TC		Scenario CH.6: LS		Scenario CH.7: IO	
	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$	$m = 1.3$	$m = 2$
$T = 1500$								
GARCH-based								
GARCH-FCMdC	0.4	1.0	0.1	1.0	0.3	0.9	0.1	0.2
GARCH-FCMdC-Exp	12.4	13.8	45.1	48.2	10.4	15.2	5.0	9.5
GARCH-FCMdC-NC	10.8	11.6	28.3	40.2	1.3	8.3	0.6	2.8
GARCH-TrFCMdC	19.9	20.0	53.6	53.2	15.2	15.2	14.3	14.3
QAF-based								
QAF-FCMdC	0.0	1.5	1.5	5.0	39.9	36.6	39.3	37.3
QAF-FCMdC-Exp	5.9	14.5	39.5	55.8	61.2	48.4	61.8	47.6
QAF-FCMdC-NC	3.5	0.6	24.8	5.6	44.9	42.4	45.5	41.0
QAF-TrFCMdC	8.6	8.1	42.6	39.5	99.0	94.9	99.2	96.6
$T = 2500$								
GARCH-based								
GARCH-FCMdC	0.0	0.0	0.1	0.2	0.0	0.3	0.0	0.1
GARCH-FCMdC-Exp	10.1	14.4	49.5	56.2	16.3	25.8	7.6	13.8
GARCH-FCMdC-NC	2.9	9.2	18.1	45.2	0.9	14.5	0.3	6.1
GARCH-TrFCMdC	20.1	20.1	59.6	59.6	26.1	26.1	20.5	20.6
QAF-based								
QAF-FCMdC	0.0	0.2	0.0	1.4	41.2	41.7	39.9	40.4
QAF-FCMdC-Exp	1.8	12.6	9.2	40.9	67.5	60.7	63.7	58.8
QAF-FCMdC-NC	3.3	1.3	16.2	6.4	46.1	42.5	43.1	40.8
QAF-TrFCMdC	7.9	7.6	28.4	28.6	99.8	99.3	100	99.7
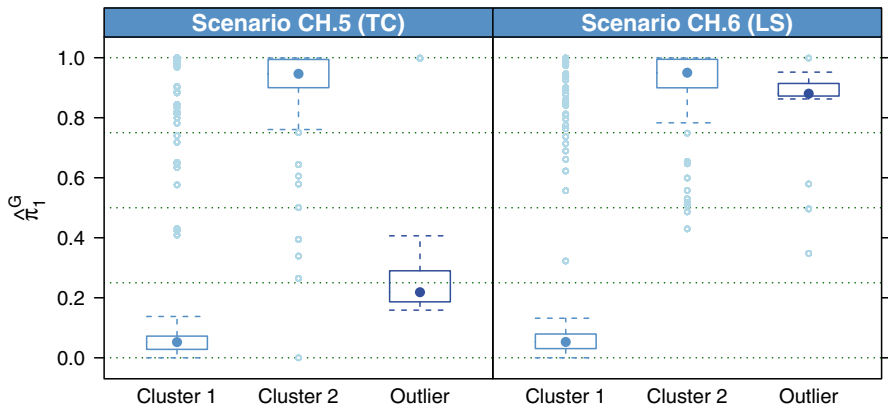
Fig. 8 Boxplots based on the estimated autoregressive coefficients for the AR(1) representation of the squared series in the conditional heteroskedastic scenarios CH.5 and CH.6 scenarios, involving TC and LS outliers, respectively

of the features used to discriminate is less affected by the one-shot impact when the length of the series grows.

Working with linear models (Table 9), the robust versions of the QAF-based algorithms fairly outperform their AR-based counterparts. The differences are particularly significant in the Scenario L6 considering LS outliers. The longer the impact effect the higher the success percentage, thus accounting for the poorest results with AO, the slight improvement with TC outliers and the excellent outcomes close to one hundred percent with LS outliers. Again the versions based on the noise cluster draw out the lowest success rates, but this can be corrected by using a less stringent cut-off value, as was already argued in Sect. 4.1. Similar results are observed by clustering non-linear series in Table 10. Compared to the linear scenario, the percentages of correct classification based on the QAF metric are somewhat lower with AO and LS outliers, but in contrast, they are fairly higher with TC outliers (Scenario NL.5).

In the scenario involving conditional heteroskedastic models (Table 11), the QAF-based algorithms perform as in the non-linear scenario, achieving similar success percentages. It is interesting to observe that the trimmed approach produces here better results than the other robust versions when LS outliers are present. Nevertheless, the most noticeable point is that the robust algorithms based on the GARCH metric exhibit better behavior to identify the AO and TC outliers than the QAF-based counterparts. Surprisingly, the success rates get worse in the most simple scenario involving LS outliers, where again the QAF metric reports better results. A careful analysis of the results allows to understand this unexpected behavior. The permanent level change generated by the shift produces an estimated autoregressive coefficient $\hat{\pi}_1^G$ [see (16)] for the ARMA representation of the squared series very close to the one defining the underlying pattern in cluster $\mathcal{C}_2$. Hence, the LS outlier usually receives a high membership degree for $\mathcal{C}_2$, thus being classified in $\mathcal{C}_2$ and causing a failed trial. For illustrative purpose, boxplots in Fig. 8 show the distributions of the estimates of π_1^G in our simulations for the series forming Scenarios CH.5 and CH.6. While the TC outliers in CH.5 produce estimates $\hat{\pi}_1^G$ reasonably equidistant from the ones for clusters $\mathcal{C}_1$

and $\mathcal{C}_2$, the $\widehat{\pi}_1^G$ corresponding to LS outliers in CH.6 are much closer to cluster $\mathcal{C}_2$. This behavior corroborates that, unlike the QAF-based models, the approaches based on the GARCH distance are extremely sensitive to model misspecification.

As far as the IO series based on contaminating the noise, the QAF-based procedures again report the best results. The average percentages of correct classification are close to 100% for the three robust versions and the two sample sizes in the non-linear scenario. With conditional heteroskedastic series, the success percentages are also almost the one hundred percent with the trimmed version, and they decay to around 60% with the QAF-FCMdC-Exp model and 44% with the QAF-FCMdC-NC model, but in all cases far above the poor results obtained by the GARCH-based models (always below 21%). The effect of IO is more intricate with the linear processes. In fact, both AR and QAF distances lead to poor results, thus exhibiting a low discriminant power in this scenario. In any case, the robust QAF-based versions are clearly the best ones.

5 Applications

There are many fields where time series clustering is a valuable analysis tool, sometimes playing a central role. Finance is a very good example because often the interest focuses on identifying assets with similar behavior over time for investment and risk analysis (Peng et al. 2011; Kou et al. 2014; Bastos and Caiado 2014; Durante et al. 2014). In this section, two specific applications involving realizations of financial time series are carried out. In both cases, our analysis is not aimed at deriving economic implications, but at illustrating the usefulness of the proposed fuzzy clustering approaches to identify: (i) homogeneous groups with similar stochastic dependence patterns, and (ii) isolated time series exhibiting atypical dependence structures.

5.1 Robust clustering of daily returns of Euro exchange rates

The first database consists of a set of series of the daily closing values of Euro exchange rates against twenty-eight international currencies, collected from 1st January 2010 to 28th February 2014 ($T = 1520$). All the data have been sourced from the website of the Bank of Italy[1] and graphs of the single series are depicted in Fig. 9.

Note that all series are non-stationary in mean. Nevertheless, the analysis of this kind of series is often focused on identifying similar volatility patterns. The volatility measures the degree of variation of a trading price series over time, and therefore high volatility means that the trading price is spread out over a large range of values. Thus, in the present application, volatility refers to the fluctuations of the Euro price with regard to other international currencies. Following the usual approach to model volatility, the series of nominal exchange rates are transformed to obtain series of daily log returns, i.e. series formed by the first differences of the natural logarithm of the nominal exchange rates. These new series are depicted in Fig. 10. Overall, the profiles of the transformed series show a high amount of heteroskedasticity, combining

---

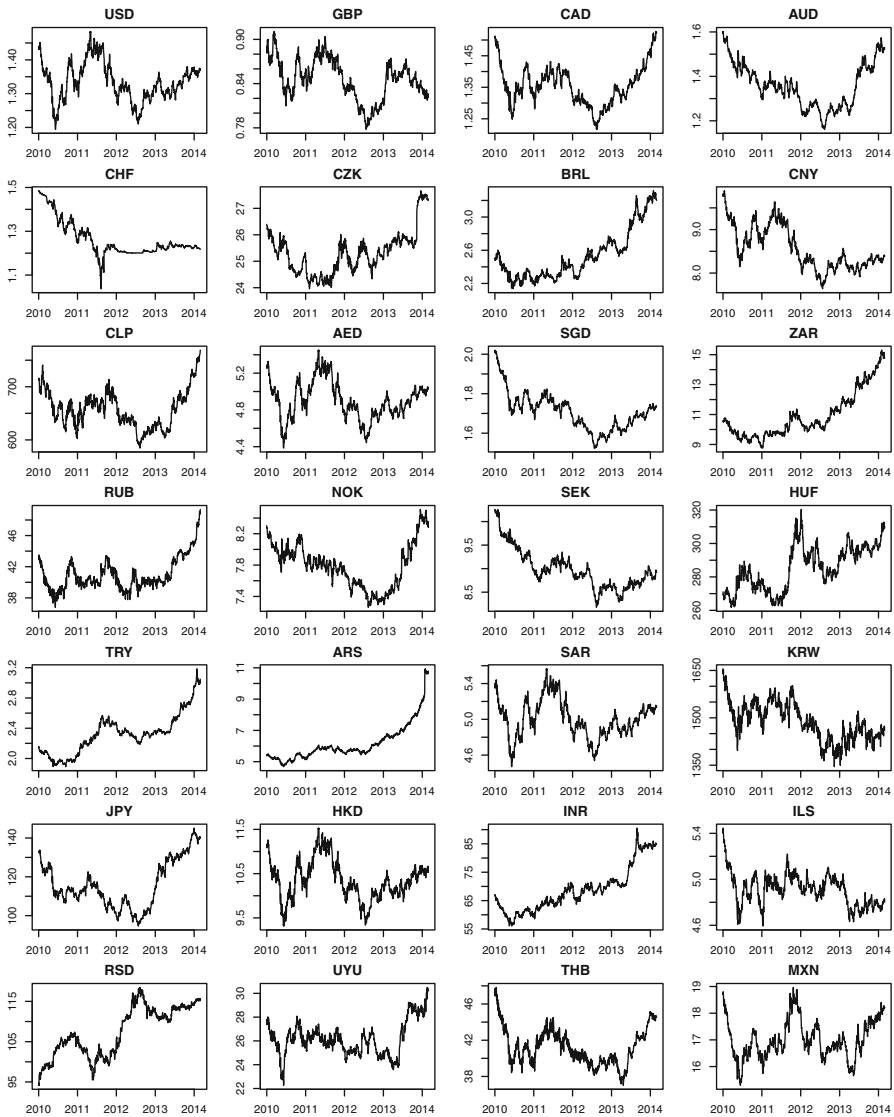[1] https://www.bancaditalia.it/compiti/operazioni-cambi/archivio-cambi/index.html.

Fig. 9 Daily closing values of Euro exchange against 28 currencies

periods of fast and pronounced fluctuations with periods of lower volatility. This phenomenon of volatility clustering is a regular feature by observing exchange rate returns, and it is frequently approached by modeling the underlying process with an ARCH or GARCH-type model. In this scenario of change in variance over time, we intend to take advantage from the capability of the quantile autocovariances to discriminate between conditionally heteroskedastic models. We expect that the robust fuzzy clustering algorithms based on d_{QAF} allow us to determine groups of currencies
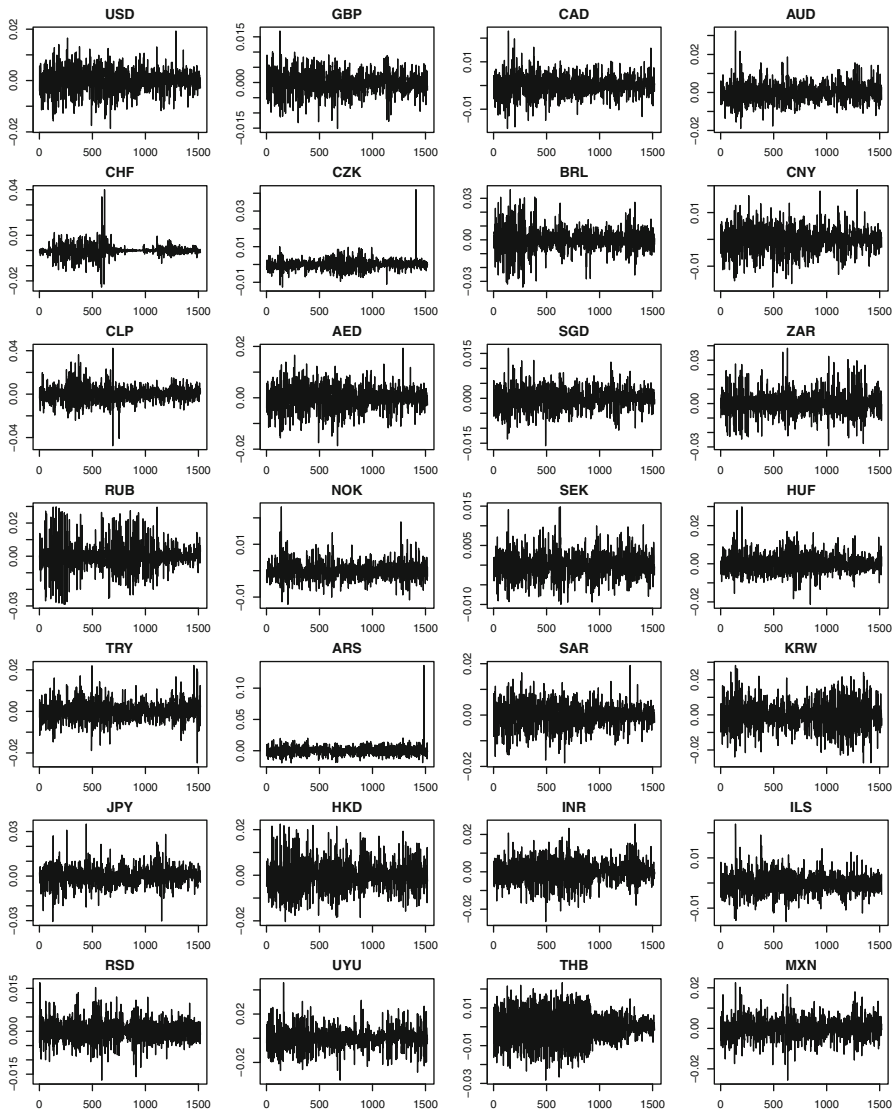
Fig. 10 Daily returns of Euro exchange against 28 currencies

following a similar behavior pattern and identify the ones presenting anomalous or atypical performances.

Just as in simulations, the metric d_{QAF} was constructed using three quantiles of levels 0.1, 0.5 and 0.9, and only one lag ($L = 1$, with $l_1 = 1$). Also, in line with the range of values considered for the fuzziness parameter m in simulations, we select the values $m = 1.3$ and $m = 1.7$. Both values produced very similar results, in particular drawing the same number of outlier time series for all the considered fuzzy models. For it, only the results for $m = 1.7$ are here included.
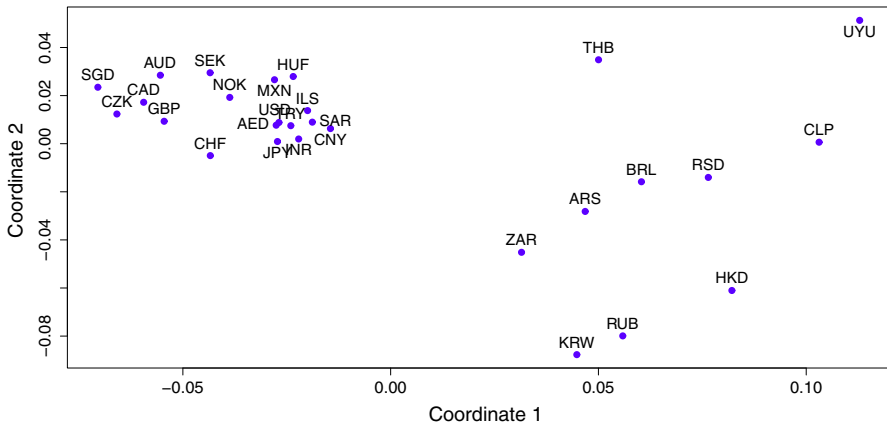
Fig. 11 Two-dimensional scaling configurations based on the QAF-distance for the daily returns of Euro exchange against 28 currencies

A two-dimensional metric scaling (MDS) based on the pairwise QAF-dissimilarity matrix was carried out to have available a projection of the Euro exchange rates on a two-dimensional map preserving the distances as well as possible. The stress measure was 6.421% and the placement of the series in the two-dimensional scaling map is shown in Fig. 11.

Figure 11 shows the existence of a reasonably compact cluster formed by eighteen series including the Euro exchange rates against the major international currencies and those linked to the US dollar, such as the Canadian dollar (CAD) and the Great Britain pound (GBP), among others. The remaining ten objects are more spread out. At least the Uruguayan peso (UYU) and the Thailand baht (THB) appear to be isolated, well-separated from the remaining currencies, and they could be identified as atypical time series. South African rand (ZAR), Argentine peso (ARS), Brazilian real (BLR), Serbian dinar (RSD), and Chilean peso (CLP) are placed close to each other, and they could constitute another cluster. The three remaining currencies, namely Hong Kong dollar (HKD), Russian ruble (RUB) and South Korean won (KRW) are somewhat separated from the latter group and they could be joined to this group or form a third cluster. For comparison purpose, a two-dimensional scaling based on the AR metric was also performed. The resulting plot exhibits a configuration with much greater dispersion and without identifying well-separated groups, which is quite unrealistic in the analyzed problem. These worse results are indeed expected because the AR metric relies on autoregressive fits, which are fairly inappropriate to model the conditionally heteroskedastic series in study.

Two different criteria to determine the optimal number of clusters C were considered, namely those values of C minimizing the Xie and Beni (1991) and Kwon (1998) indexes. To simplify the definition of both indexes, let us denote by H_{ic} the squared Euclidean distance between the sequence of estimated quantile autocovariances for the ith series and the average sequence for the cth cluster, that is
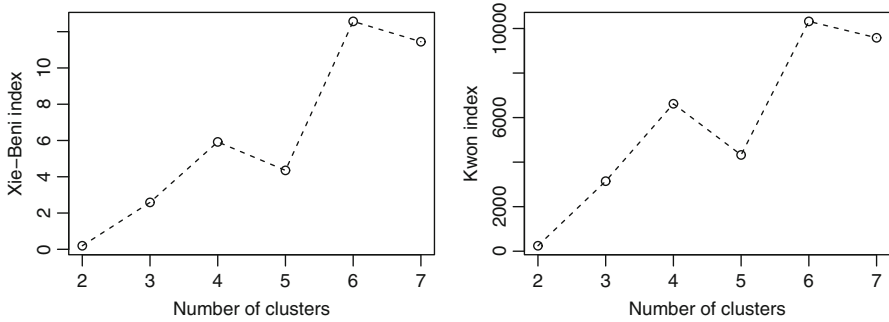
Fig. 12 Xie–Beni and Kwon indexes for different sizes of partition using QAF-FCMdC

$$H_{ic} = \sum_{k=1}^{L} \sum_{j=1}^{r} \sum_{j'=1}^{r} \left(\hat{\gamma}_{l_k}^{(i)}(\tau_j, \tau_{j'}) - \overline{\hat{\gamma}}_{l_k}^{(c)}(\tau_j, \tau_{j'}) \right)^2. \tag{17}$$

The Xie–Beni index for a partition into C clusters is defined as the ratio between the total variance and the minimum separation between clusters, i.e.

$$XB(C) = \frac{\displaystyle\sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^m H_{ic}}{n \displaystyle\min_{c \neq c'} H_{cc'}}. \tag{18}$$

Note that minimizing the numerator of $XB(\cdot)$ in (18) is the goal of the QAF-FCMdC algorithm. On the other hand, the denominator of $XB(\cdot)$ measures how separated are the clusters, thus the Xie–Beni index decreases with the separation between clusters.

The Kwon index provides a correction of the Xie–Beni index by penalizing the decreasing tendency when the number of clusters becomes very large and close to the number of time series. Specifically, the Kwon index is defined as follows.

$$K(C) = \frac{\displaystyle\sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^m H_{ic} + \frac{1}{C} \sum_{c=1}^{C} \sum_{c'=1}^{C} H_{cc'}}{n \displaystyle\min_{c \neq c'} H_{cc'}}. \tag{19}$$

The values obtained for both indexes using QAF-FCMdC are depicted in Fig. 12. In both cases the lowest value is attained for $C = 2$ clusters, with a substantial increase when three or more clusters are considered. Similar results were obtained by using the robust versions of the model, and therefore both criteria lead to conclude the existence of two major groups.

The value for the parameter β required by the QAF-FCMdC-Exp model was determined using (7) in Sect. 2.3, resulting $\beta = 1095.649$. To set δ in the QAF-FCMdC-NC model, we follow the approach suggested by Cimino et al. (2005), which consists of

successively executing the fuzzy QAF-FCMdC-NC algorithm for decreasing values of δ, recording the percentage of series assigned to the noise cluster, and selecting the value of δ producing an abrupt change of slope (elbow) in this percentage. The idea is gradually reducing δ until a proper threshold is found out because of excessively small values of δ lead to assign non-anomalous objects into the cluster noise. According to this criterion, $\delta = 0.4$ was selected. As far as the QAF-TrFCMdC model, the trimming ratio α minimizing the Xie–Beni and the Kwon indexes over a grid of possible values for α was considered as the optimal choice, resulting $\alpha = 0.1621$, i.e. five time series were trimmed.

Table 12 shows the membership degrees obtained by using the standard and robust fuzzy methods. For each single series, the bolditalicized entries enhance the highest membership degrees obtained with each procedure, i.e. the cluster assignments from a crisp perspective. The memberships showed in bold font for a particular robust procedure indicate time series identified as outlier. The currencies' names in bold font refer to series identified as outliers by the three robust methods. When only one or two robust procedures achieved that conclusion, the currency is written in italic font.

Overall, the obtained partition with the standard fuzzy model QAF-FCMdC is consistent with the plot displayed in Fig. 11. The medoid time series are the Emiratri dirham (AED), for the most compact cluster (C_2) grouping eighteen currencies, and the Brazilian real (BRL) for the cluster C_1 exhibiting higher spread. It is noticeable that most of the currencies are assigned to one cluster with high membership degrees ($u_{ic} \geq 0.7$), the only exception being the Thailand baht (THB), which was located in C_1 with membership 0.639. Nevertheless, Fig. 11 suggests that THB is too far from the time series forming C_2 and hence the Thailand baht should be considered as an outlier. In short, QAF-FCMdC seems to work reasonably fine, but it does not allow us to identify currencies showing an atypical behavior.

The partition obtained with QAF-FCMdC-Exp determines the existence of four outlier time series by splitting their membership degrees uniformly across the clusters, namely the Uruguayan peso (UYU), the Thailand baht (THB), the South Korean won (KRW) and the Russian ruble (RUB). These four currencies are also allocated together into the noise cluster with memberships $u_{in_C} > 0.6$ when the QAF-FCMdC-NC model is considered. The South African rand (ZAR) and the Hong Kong dollar (HKD) are also added to the noise cluster on the basis of much weaker memberships, particularly the former with memberships for C_1 and the noise cluster hardly discernible, 0.458 and 0.463, respectively. Note that consideration of these isolated objects modifies the C_2 medoid, now resulting the Norwegian krone (NOK) which seems to be a more representative prototype than AED in Fig. 11. The fuzzy QAF-TrFCMdC model draw out very similar results. Considering a trimmed ratio of $\alpha = 0.1621$, five Euro exchange currencies are trimmed away, namely the same four outliers identified by the other two robust methods plus HKD. Actually, a small reduction of the trimmed ratio allows to cancel this additional outlier so that in essence the three robust methods allow us to obtain similar conclusions.

Table 12 Membership degrees for the fuzzy clustering models based on quantile autocovariances by considering a two-cluster partition

	QAF-FCMdC		QAF-FCMdC-Exp		QAF-FCMdC-NC			QAF-TrFCMdC		
	C_1	C_2	C_1	C_2	C_1	C_2	NC	C_1	C_2	Tr
USD	0.001	*0.999*	0.100	*0.900*	0.034	*0.801*	0.165	0.012	*0.988*	N
GBP	0.030	*0.970*	0.087	*0.913*	0.016	*0.807*	0.177	0.036	*0.964*	N
CAD	0.030	*0.970*	0.084	*0.916*	0.013	*0.813*	0.174	0.045	*0.955*	N
AUD	0.040	*0.960*	0.077	*0.923*	0.013	*0.828*	0.159	0.044	*0.956*	N
CHF	0.042	*0.958*	0.145	*0.855*	0.037	*0.697*	0.266	0.025	*0.975*	N
CZK	0.040	*0.960*	0.111	*0.889*	0.018	*0.755*	0.228	0.044	*0.956*	N
BRL	*1.000*	0.000	*1.000*	0.000	*1.000*	0.000	0.000	*0.938*	0.062	N
CNY	0.019	*0.981*	0.171	*0.829*	0.070	*0.684*	0.246	0.028	*0.972*	N
CLP	*0.946*	0.054	*0.802*	0.198	*0.573*	0.024	0.403	*0.953*	0.047	N
AED	0.000	*1.000*	0.085	*0.915*	0.028	*0.829*	0.143	0.012	*0.988*	N
SGD	0.044	*0.956*	0.159	*0.841*	0.022	*0.652*	0.326	0.067	*0.933*	N
ZAR	*0.799*	0.201	*0.720*	0.280	0.458	0.079	*0.463*	*0.691*	0.309	N
RUB	*0.809*	0.191	*0.636*	0.364	0.281	0.044	*0.674*	–	–	**Y**
NOK	0.021	*0.979*	0.000	*1.000*	0.000	*1.000*	0.000	0.013	*0.987*	N
SEK	0.045	*0.955*	0.017	*0.983*	0.003	*0.963*	0.034	0.030	*0.970*	N
HUF	0.066	*0.934*	0.059	*0.941*	0.017	*0.877*	0.106	0.012	*0.988*	N
TRY	0.012	*0.988*	0.085	*0.915*	0.031	*0.832*	0.137	0.004	*0.996*	N
ARS	*0.911*	0.089	*0.858*	0.142	*0.705*	0.038	0.257	*0.693*	0.307	N
SAR	0.004	*0.996*	0.140	*0.860*	0.053	*0.730*	0.218	0.019	*0.981*	N
KRW	*0.744*	0.256	*0.591*	0.409	0.214	0.051	*0.735*	–	–	**Y**
JPY	0.015	*0.985*	0.113	*0.887*	0.035	*0.770*	0.194	0.018	*0.982*	N

Table 12 continued

	QAF-FCMdC		QAF-FCMdC-Exp		QAF-FCMdC-NC			QAF-TrFCMdC		
	C_1	C_2	C_1	C_2	C_1	C_2	NC	C_1	C_2	Tr
HKD	*0.904*	0.096	*0.723*	0.277	0.426	0.031	**0.543**	–	–	**Y**
INR	0.020	*0.980*	0.103	*0.897*	0.033	*0.790*	0.177	0.027	*0.973*	N
ILS	0.022	*0.978*	0.076	*0.924*	0.029	*0.853*	0.119	0.000	*1.000*	N
RSD	*0.964*	0.036	*0.894*	0.106	*0.768*	0.020	0.212	*1.000*	0.000	N
UYU	*0.812*	0.188	*0.581*	*0.419*	0.193	0.039	**0.767**	–	–	**Y**
THB	*0.639*	0.361	*0.582*	*0.418*	0.239	0.105	**0.655**	–	–	**Y**
MXN	0.045	*0.955*	0.058	*0.942*	0.017	*0.877*	0.106	0.022	*0.978*	N

5.2 Robust clustering of daily stocks returns in IBEX-35 index

The second application considers daily returns of stocks included in the IBEX-35, which groups the thirty-five companies with the highest liquidity and trading volume in the Spanish stock market. Specifically, we select the daily returns of twenty-four stocks located in the TOP-30 ranking according to the finance section of the Yahoo website.[2] The period of observation of the series spans from 1st January 2008 to 19th December 2016, thus resulting realizations of length $T = 2337$ which are depicted in Fig. 13. Note that different levels of heteroskedasticity are again observed, which are more pronounced for some of the stocks.

This database was previously subjected to fuzzy clustering in Vilar et al. (2018) considering the QAF-FCMdC, the AR-FCMdC and the GARCH-FCMdC models. According to the Xie–Beni index, the optimal value for the number of clusters is $C = 3$. This conclusion was also corroborated with the criterion of selecting the number of clusters maximizing the so-called Fuzzy Silhouette Width (Campello and Hruschka 2006). In fact, the highest fuzzy silhouette index corresponded to $C = 3$ and $m = 2$ so that the clustering algorithms were executed in Vilar et al. (2018) using these values. The results showed that only the QAF-FCMdC model produces a fuzzy partition congruent with the estimated volatility patterns.

Specifically, the 3-cluster solution generated by the QAF-FCMdC model consisted of a large cluster, $\mathcal{C}_1$, gathering together most of the stocks, a cluster $\mathcal{C}_3$ grouping the company Arcelormittal-MTE together with the smaller banks Banco Popular-POP, Banco Sabadell-SAB and Bankinter-BKT, and a cluster $\mathcal{C}_2$ formed by two important companies of the consumer goods industry (Viscofan-VIS and Inditex-ITX), the only insurance company (Mapfre-MAP), and a technological company related to the travel sector (Amadeus-AMS). However, beyond interpretations in financial terms (e.g. company size or business sector), we observed that the resulting groups are characterized by presenting similar fluctuation patterns, with some bumps of different size in similar periods of time. This can be seen in Fig. 14, where the estimated volatility curves grouped by the identified clusters are shown.

Here, we go one step further by checking for the presence of stocks with atypical temporal behavior. First, as in the above study case, we carry out a two-dimensional scaling based on the pairwise QAF-dissimilarity matrix. The resulting MDS plot in Fig. 15, with an associated stress value of 4.275%, allows us to visually inspect the proximity between the time series in terms of the QAF-distance. Note that the labels have been colored according to the cluster solution obtained with the QAF-FCMdC model.

Figure 15 shows that Banco Sabadell-SAB fairly constitutes an isolated point located very far from the bulk of the data points. Therefore, the MDS plot suggests that SAB is an outlier and it should not be included in the cluster $\mathcal{C}_3$ together with MTS and POP (the two nearest series). This conjecture is supported by Fig. 14c, where a flat volatility profile is depicted for SAB. Also the Viscofan-VIS stock, placed into $\mathcal{C}_2$, can be seen as a potential outlier. It is the second furthest data point in the MDS plot, and its volatility profile does not show the peaks of the rest of members in $\mathcal{C}_2$.
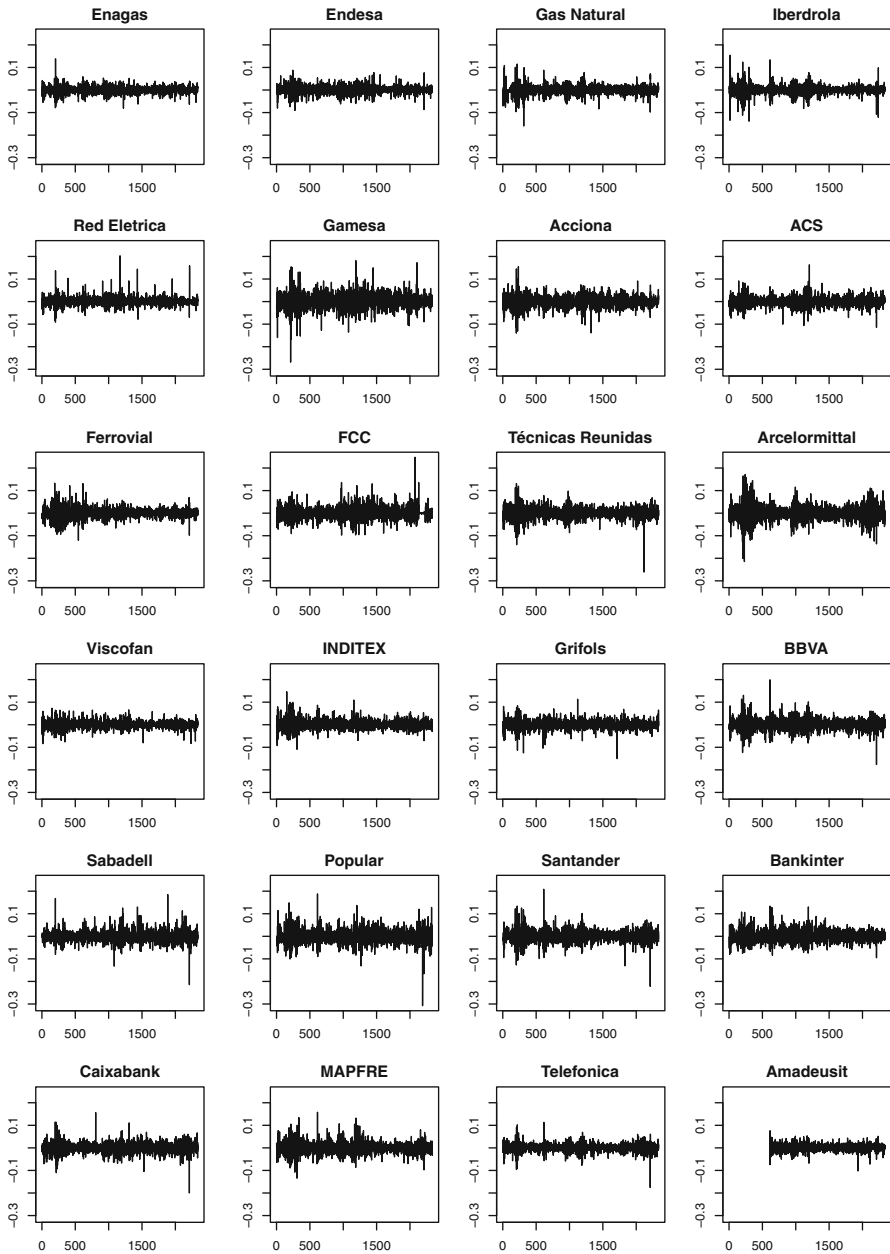
---

[2] https://finance.yahoo.com/.

Fig. 13 Daily returns of 24 stocks included in the IBEX-35. Sample period: 1st January 2008 to 19th December 2016
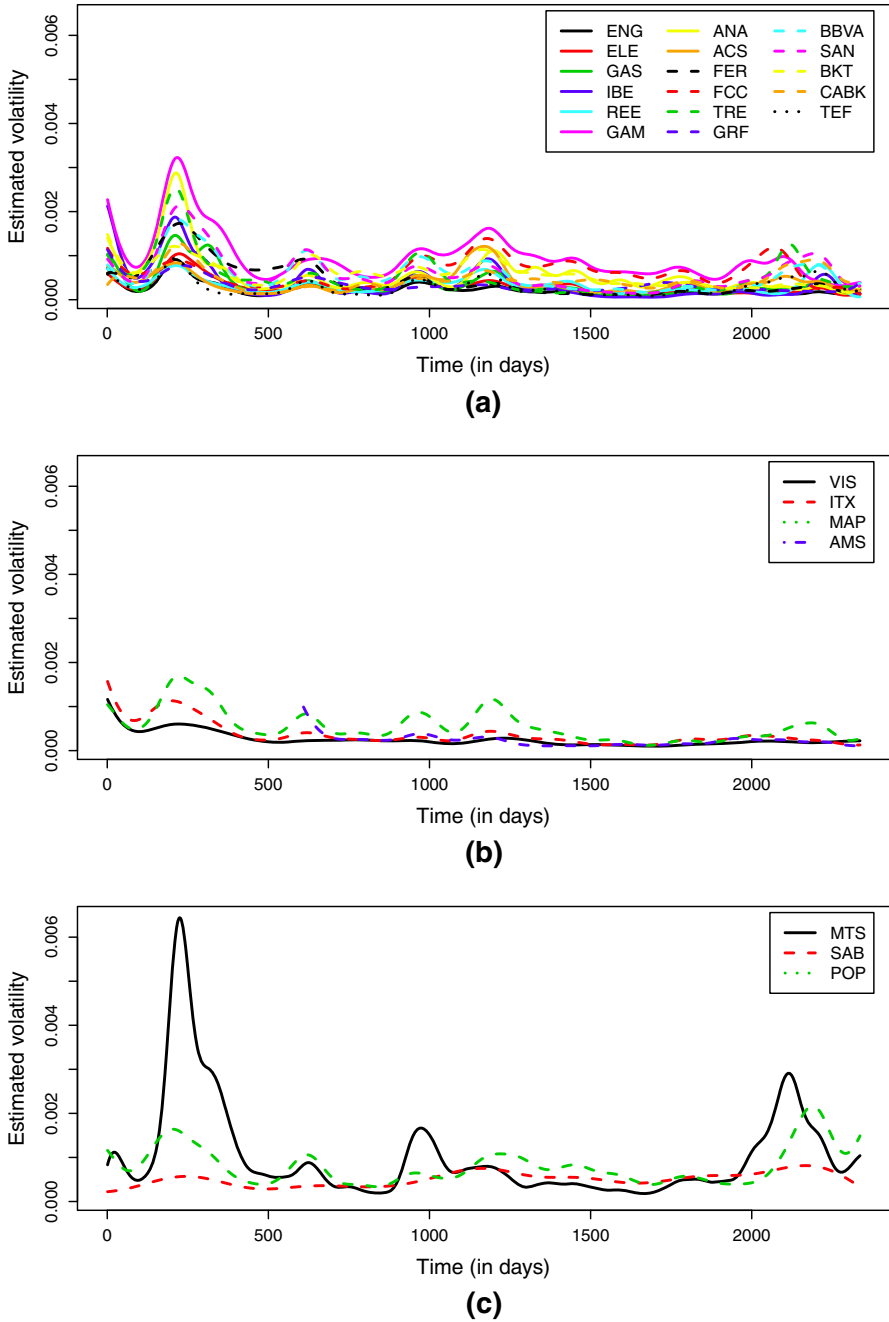
Fig. 14 Nonparametric estimators of the volatility for the daily returns of the 24 analyzed stocks grouped according to the cluster solution provided by the QAF-FCMdC model: $\mathcal{C}_1$ (**a**), $\mathcal{C}_2$ (**b**) and $\mathcal{C}_3$ (**c**). (Color figure online)
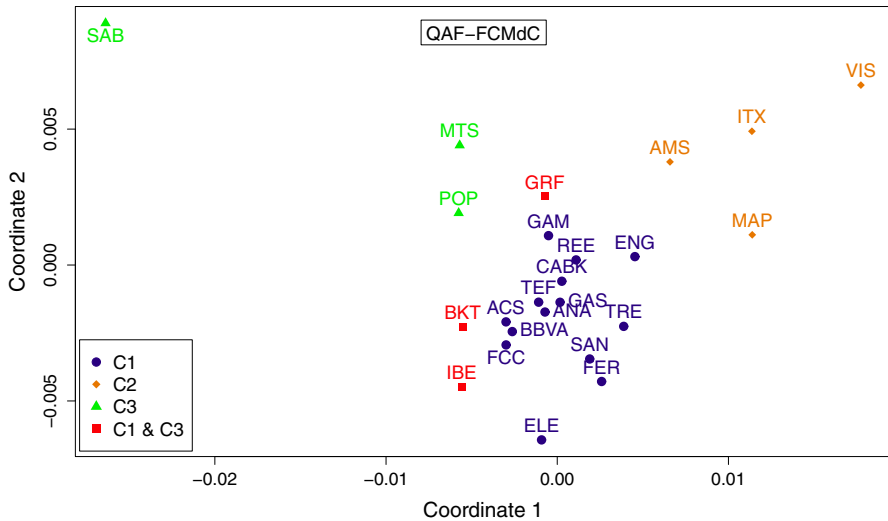
Fig. 15 Two-dimensional scaling configurations based on the QAF-distance for the daily returns of the 24 analyzed stocks. (Color figure online)

According to these considerations, it is interesting to examine the clustering results produced by the robust versions of the QAF-FCMdC model. Table 13 provides the membership degrees obtained with the QAF-FCMdC model and its robust versions. As in Table 12, the highest membership degrees with each procedure are enhanced using bolditalicized entries. Since we manage a 3-cluster solution, the highest membership can be spread out across two clusters, thus indicating a fuzzy allocation between them. For this reason, we use a bolditalicized entries as long as $u_{ic} > 0.35$.

The results from Table 13 reveal that the standard fuzzy model QAF-FCMdC does not identify atypical time series since no stock presents membership degrees uniformly distributed over the three clusters. Only Grifols-GRF, Bankinter-BTK and Iberdrola-IBE present a fuzzy allocation between $\mathcal{C}_1$ and $\mathcal{C}_3$, while the remaining stocks are assigned to one cluster with high membership degrees.

The robust QAF-FCMdC-NC algorithm produces a very similar result but it is capable to detect outliers. Sabadell-SAB is fairly identified as outlier since it is included in the noise cluster with a high membership 0.816. On the other hand, Viscofan-VIS has a fuzzy allocation between the cluster $\mathcal{C}_2$ and the noise cluster with membership degrees 0.505 and 0.405, respectively. Therefore, VIS can be also seen as a potential outlier. As for the remaining stocks, minor differences are observed. Overall, the membership degrees slightly decrease in favor of the noise cluster, but the clusters $\mathcal{C}_1$ and $\mathcal{C}_3$ are formed by the same elements as in the case of QAF-FCMdC, including the same centroids. Therefore, QAF-FCMdC-NC leads to an intuitive classification, reasonably congruent with the estimated volatility curves, and robust enough to detect the atypical series without perverting the clustering process.

The models QAF-FCMdC-Exp and QAF-TrFCMdC lead to different results. QAF-FCMdC-Exp identifies SAB as outlier and retains the same composition for the cluster $\mathcal{C}_2$. However, clusters $\mathcal{C}_1$ and $\mathcal{C}_3$ are now hardly discernible since most of the stocks

Table 13 Membership degrees for the fuzzy clustering models considering $C = 3$ clusters and $m = 2$

		QAF-FCMdC			QAF-FCMdC-Exp			QAF-FCMdC-NC				QAF-TrFCMdC			
		C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	NC	C_1	C_2	C_3	Tr
Sector: energy															
Enagás	ENG	*0.558*	0.275	0.167	*0.405*	0.217	*0.378*	*0.410*	0.202	0.122	0.266	*0.455*	0.322	0.223	N
Endesa	ELE	*0.684*	0.072	0.244	*0.454*	0.080	*0.466*	*0.501*	0.053	0.179	0.267	*0.525*	0.068	*0.407*	N
Gas Natural	GAS	*0.946*	0.013	0.041	*0.700*	0.014	0.286	*0.906*	0.012	0.040	0.042	*0.814*	0.014	0.172	N
Iberdrola	IBE	*0.579*	0.053	*0.369*	*0.471*	0.090	*0.439*	*0.413*	0.038	0.263	0.287	0.270	0.029	*0.701*	N
Red Eletrica	REE	*0.814*	0.054	0.132	*0.499*	0.044	*0.456*	*0.716*	0.048	0.116	0.120	*0.702*	0.055	0.243	N
Sector: materials, industry and construction															
Gamesa	GAM	*0.703*	0.079	0.218	*0.512*	0.087	*0.402*	*0.538*	0.061	0.166	0.235	*0.490*	0.069	*0.442*	N
Acciona	ANA	*1.000*	0.000	0.000	*1.000*	0.000	0.000	*1.000*	0.000	0.000	0.000	*1.000*	0.000	0.000	N
Act. Const. y Servicios	ACS	*0.816*	0.020	0.164	*0.535*	0.024	*0.441*	*0.738*	0.018	0.149	0.095	0.000	0.000	*1.000*	N
Ferrovial	FER	*0.754*	0.095	0.151	*0.484*	0.083	*0.433*	*0.587*	0.074	0.118	0.221	*0.614*	0.109	0.277	N
Fomento de const. y contr.	FCC	*0.821*	0.025	0.154	*0.560*	0.031	*0.408*	*0.723*	0.022	0.136	0.119	0.132	0.005	*0.864*	N
Técnicas Reunidas	TRE	*0.710*	0.147	0.142	*0.494*	0.129	*0.377*	*0.526*	0.109	0.106	0.259	*0.538*	0.196	0.266	N
Arcelormittal	MTS	0.084	0.020	*0.896*	0.382	0.145	*0.473*	0.075	0.018	*0.809*	0.098	*0.435*	0.094	*0.471*	N
Sector: consumer goods															
Viscofan	VIS	0.087	*0.849*	0.064	0.151	*0.697*	0.151	0.052	*0.505*	0.038	*0.405*	–	–	–	Y
Inditex	ITX	0.000	*1.000*	0.000	0.000	*1.000*	0.000	0.000	*1.000*	0.000	0.000	0.097	*0.833*	0.070	N
Grifols	GRF	*0.539*	0.07	*0.392*	0.387	0.068	*0.546*	*0.449*	0.058	0.326	0.166	*0.583*	0.073	0.344	N
Sector: financial services															
BBVA	BBVA	*0.860*	0.017	0.123	*0.561*	0.020	*0.419*	*0.790*	0.016	0.113	0.081	0.064	0.002	*0.934*	N
Sabadell	SAB	0.316	0.168	*0.516*	0.336	0.327	0.337	0.058	0.031	0.095	*0.816*	–	–	–	Y
Popular	POP	0.000	0.000	*1.000*	0.390	0.096	*0.513*	0.000	0.000	*1.000*	0.000	*0.414*	0.058	*0.528*	N
Santander	SAN	*0.801*	0.076	0.123	*0.558*	0.077	0.365	*0.631*	0.06	0.097	0.212	*0.588*	0.087	0.325	N

Table 13 continued

		QAF-FCMdC			QAF-FCMdC-Exp			QAF-FCMdC-NC				QAF-TrFCMdC			
		C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	NC	C_1	C_2	C_3	Tr
Bankinter	BKT	**0.501**	0.040	**0.459**	**0.465**	0.074	**0.461**	**0.394**	0.031	**0.361**	0.214	0.210	0.019	**0.772**	N
Caixabank	CABK	**0.860**	0.037	0.103	**0.597**	0.038	**0.365**	**0.765**	0.033	0.092	0.110	**0.652**	0.036	0.313	N
Sector: insurance															
MAPFRE	MAP	0.118	**0.821**	0.061	0.141	**0.722**	0.137	0.087	**0.605**	0.045	0.263	0.000	**1.000**	0.000	N
Sector: technology and telecommunications															
Telefonica	TEF	**0.955**	0.006	0.039	0.000	0.000	**1.000**	**0.933**	0.006	0.038	0.023	**0.858**	0.007	0.136	N
Amadeus	AMS	0.184	**0.708**	0.107	0.190	**0.618**	0.192	0.138	**0.531**	0.080	0.251	0.242	**0.599**	0.159	N

are equidistant from both clusters. This way, a coherent classification is also obtained although an interpretation in terms of two clusters would be more feasible (in fact, $C = 2$ is optimal when the Xie–Beni index is examined using QAF-FCMdC-Exp). Lastly, QAF-TrFCMdC performs quite similar as QAF-FCMdC-Exp. Based on the Xie–Beni index, the optimal trimmed ratio is $\alpha = 0.104$, i.e. two time series are trimmed away. Not surprisingly QAF-TrFCMdC determines that SAB and VIS are atypical series, the cluster $\mathcal{C}_2$ (now without VIS) is correctly identified, and clusters $\mathcal{C}_1$ and $\mathcal{C}_3$ again exhibit a high amount of overlapping.

6 Concluding remarks

We have proposed three robust versions of the Fuzzy C-Medoid Clustering model for time series based on sample quantile autocovariances (QAF-FCMdC). The main objective has been to develop structure-based clustering algorithms capable to neutralize the disruptive effect of the outliers, i.e. of isolated time series generated from dependence models different from the ones determining the true fuzzy clustering structure. Three generalizations of robust techniques considered in the clustering literature to manage the presence of anomalous data have been introduced, namely the metric approach considering a more robust metric (QAF-FCMdC-Exp), the noise approach introducing an artificial noise cluster (QAF-FCMdC-NC), and the trimmed approach based on trimming away a small fraction of series (QAF-TrFCMdC). In essence, our proposal intends to take advantage of the nice properties of the involved methodological tools, including the versatility of the fuzzy logic by permitting overlapping clusters, the interest in selecting a subset of observed series like prototypes in accord with the PAM technique, and in particular, the high capability of the QAF-metric to discriminate between complex dependence structures such as non-linear or conditionally heteroskedastic models.

The proposed approaches have been evaluated by means of a broad simulation study involving scenarios with different generating models and contaminated with outliers. Regardless of the considered models, our experiments have shown that the fuzzy robust procedures reported satisfactory results. Unlike the standard algorithms, all the robust versions worked fine in presence of outliers when an optimal selection of the input parameters was carried out. Furthermore, the robust methods were able to neutralize the effect of the anomalous series preserving the true clustering structure. It is worthy to remark that our approach outperformed other robust clustering algorithms based on alternative metrics and specifically designed to classify the models subjected to clustering. No significant differences have been observed between the behaviors of the robust versions, but it is important to emphasize that all the robust procedures are particularly sensitive to the choice of the input parameters. For illustrative purpose, the proposed clustering techniques have been applied to two specific real data sets consisting of realizations of financial time series. Our analysis showed as the proposed robust clustering approaches were able to identify series exhibiting atypical dynamic behaviors.

In conclusion, the experiments with simulated and real data sets have shown robustness against extreme observations of the proposed clustering procedures, thus

providing valuable tools to perform time series clustering. Anyway, we think that there are still interesting points to be considered in further research. For instance, although the QAF-distance reports interesting properties and has shown robustness against the underlying models, introducing suitable approaches to encompass non-stationary models has great interest in applications. Also, an exhaustive study about the optimal selection of the input parameters required for the robust algorithms is also desirable due to the great sensitivity of the procedures to these parameters. Extensions considering versions of the quantile autocovariance metric in the frequency domain are also of interest. In particular, a metric such as the L^2-distance between copula ranks periodograms or between smoothed versions of them (Dette et al. 2015) would take advantage of automatically considering all the lags. An additional point to address in future work is the cross-dependency issue. Traditional measures comparing features as coherence or cross-correlation or also quantile-based procedures (Baruník and Kley 2015) could be taken into consideration.

References

Aghabozorgi S, Shirkhorshidi AS, Wah TY (2015) Time-series clustering a decade review. Inf Syst 53(C):16–38

Aielli GP, Caporin M (2013) Fast clustering of GARCH processes via gaussian mixture models. Math Comput Simul 94:205–222

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

Alonso AM, Berrendero JR, Hernández A, Justel A (2006) Time series clustering based on forecast densities. Comput Stat Data Anal 51(2):762–776

Amendola A, Francq C (2009) Concepts and tools for nonlinear time-series modelling. Wiley, New York, pp 377–427

An HZ, Huang FC (1996) The geometrical ergodicity of nonlinear autoregressive models. Stat Sin 6(4):943–956

Arabie P, Carroll JD, DeSarbo WS, Wind YJ (1981) Overlapping clustering: a new method for product positioning. J Mark Res 18(3):310–317

Baruník J, Kley T (2015) Quantile cross-spectral measures of dependence between economic variables. arXiv:1510.06946

Bastos JA, Caiado J (2014) Clustering financial time series with variance ratio statistics. Quant Financ 14(12):2121–2133

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, Norwell, MA

Caiado J, Crato N (2010) Identifying common dynamic features in stock returns. Quant Financ 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, Peña D (2009) Comparison of times series with unequal length in the frequency domain. Commun Stat Simul Comput 38(3):527–540

Caiado J, Maharaj E, D'Urso P (2015) Time series clustering. In: Hennig C, Meila M, Murtagh F, Rocci R (eds) Handbook of cluster analysis. Chapman and Hall/CRC, Boca Raton, pp 241–264

Campello R, Hruschka E (2006) A fuzzy extension of the sihouette width criterion for cluster analysis. Fuzzy Sets Syst 157(21):2858–2875

Chae SS, Kim C, Kim JM, Warde WD (2008) Cluster analysis using different correlation coefficients. Stat Pap 49(4):715–727

Chen C, So M, Liu FC (2011) A review of threshold time series models in finance. Stat Interface 4:167–181

Cimino M, Frosini G, Lazzerini B, Marcelloni F (2005) On the noise distance in robust fuzzy c-means. Proc World Acad Sci Eng Technol 1:361–364

Coppi R, D'Urso P (2002) Fuzzy K-means clustering models for triangular fuzzy time trajectories. Stat Methods Appt 11(1):21–40

Coppi R, D'Urso P (2003) Three-way fuzzy clustering models for LR fuzzy time trajectories. Comput Stat Data Anal 43(2):149–177

Coppi R, D'Urso P (2006) Fuzzy unsupervised classification of multivariate time trajectories with the Shannon entropy regularization. Comput Stat Data Anal 50(6):1452–1477

Coppi R, D'Urso P, Giordani P (2006) Fuzzy C-medoids clustering models for time-varying data. In: Bouchon-Meunier B, Coletti G, Yager S (eds) Modern information processing: from theory applications. Elsevier, New York, pp 195–206

Coppi R, D'Urso P, Giordani P (2010) A fuzzy clustering model for multivariate spatial time series. J Classif 27(1):54–88

Dave RN (1991) Characterization and detection of noise in clustering. Pattern Recognit Lett 12(11):657–664

Davé RN, Krishnapuram R (1997) Robust clustering methods: an unified view. IEEE Trans Fuzzy Syst 5:270–293

Davé RN, Sen S (1997) Noise clustering algorithm revisited. In: IEEE Fuzzy information processing society, 1997 annual meeting of the North American, NAFIPS'97, pp 199–204

Davé RN, Sen S (2002) Robust fuzzy clustering of relational data. IEEE Trans Fuzzy Syst 10(6):713–727

De Luca G, Zuccolotto P (2017) Dynamic tail dependence clustering of financial time series. Stat Pap 58(3):641–657

Dette H, Hallin M, Kley T, Volgushev S (2015) Of copulas, quantiles, ranks and spectra: an l_1-approach to spectral analysis. Bernoulli 21(2):781–831

Di Lascio FML, Giannerini S (2016) Clustering dependent observations with copula functions. Stat Pap https://doi.org/10.1007/s00362-016-0822-3

Disegna M, D'Urso P, Durante F (2017) Copula-based fuzzy clustering of spatial time series. Spat Stat 21(Part A):209–225

Dugard P, Todman JB, Staines H (2010) Approaching multivariate analysis: a practical introduction, 2nd edn. Routledge, London

Durante F, Pappadà R, Torelli N (2014) Clustering of financial time series in risky scenarios. Adv Data Anal Classif 8(4):359–376

Durante F, Pappadà R, Torelli N (2015) Clustering of time series via non-parametric tail dependence estimation. Stat Pap 56(3):701–721

D'Urso P (2004) Fuzzy C-means clustering models for multivariate time-varying data: different approaches. Int J Uncertain Fuzz 12(03):287–326

D'Urso P (2005) Fuzzy clustering for data time arrays with inlier and outlier time trajectories. IEEE Trans Fuzzy Syst 13(5):583–604

D'Urso P (2015) Fuzzy clustering. In: Hennig C, Meila M, Murtagh F, Rocci R (eds) Handbook of cluster analysis. Chapman and Hall/CRC, Boca Raton

D'Urso P, De Giovanni L (2008) Temporal self-organizing maps for telecommunications market segmentation. Neurocomputing 71(13):2880–2892

D'Urso P, De Giovanni L (2014) Robust clustering of imprecise data. Chemometr Intell Lab Syst 136:58–80

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

D'Urso P, Maharaj EA (2012) Wavelets-based clustering of multivariate time series. Fuzzy Sets Syst 193:33–61

D'Urso P, Cappelli C, Di Lallo D, Massari R (2013a) Clustering of financial time series. Physica A 392(9):2114–2129

D'Urso P, De Giovanni L, Massari R, Di Lallo D (2013b) Noise fuzzy clustering of time series by autoregressive metric. Metron 71(3):217–243

D'Urso P, Di Lallo D, Maharaj EA (2013c) Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks. Soft Comput 17(1):83–131

D'Urso P, De Giovanni L, Maharaj EA, Massari R (2014) Wavelet-based self-organizing maps for classifying multivariate time series. J Chemom 28(1):28–51

D'Urso P, De Giovanni L, Massari R (2015) Time series clustering by a robust autoregressive metric with application to air pollution. Chemometr Intell Lab Syst 141:107–124

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, Maharaj EA, Alonso AM (2017a) Fuzzy clustering of time series using extremes. Fuzzy Sets Syst 318(Supplement C):56–79. https://doi.org/10.1016/j.fss.2016.10.006

D'Urso P, Massari R, Cappelli C, De Giovanni L (2017b) Autoregressive metric-based trimmed fuzzy clustering with an application to PM 10 time series. Chemometr Intell Lab Syst 161:15–26

D'Urso P, Giovanni LD, Massari R (2018) Robust fuzzy clustering of multivariate time trajectories. Int J Approx Reason 99:12–38

Everitt B, Landau S, Leese S (2001) Clust Anal. Arnold Press, London

Fan J, Yao Q (2005) Nonlinear time series: nonparametric and parametric methods. Springer series in statistics, Springer, New York

Floriello D, Vitelli V (2017) Sparse clustering of functional data. J Multivar Anal 154:1–18

Fu TC (2011) A review on time series data mining. Eng Appl Artif Intell 24(1):164–181

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

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

García-Escudero LA, Gordaliza A, Matrán C, Mayo-Iscar A (2010) A review of robust clustering methods. Adv Data Anal Classif 4(2):89–109

Górecki T, Krzyśko M, Waszak Ł, Wołyński W (2018) Selected statistical methods of data analysis for multivariate functional data. Stat Pap 59(1):153–182

Hagemann A (2013) Robust spectral analysis. arXiv:1111.1965v1

Heiser WJ, Groenen PJF (1997) Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. Psychometrika 62(1):63–83

Höppner F (1999) Fuzzy cluster analysis: methods for classification, data analysis and image recognition. Wiley, New York

Hruschka H (1986) Market definition and segmentation using fuzzy clustering methods. Int J Res Market 3(2):117–134

Hwang H, Desarbo WS, Takane Y (2007) Fuzzy clusterwise generalized structured component analysis. Psychometrika 72(2):181–198

James GM, Sugar CA (2003) Clustering for sparsely sampled functional data. J Am Stat Assoc 98(462):397–408

Kalpakis K, Gada D, Puttagunta V (2001) Distance measures for effective clustering of ARIMA time-series. In: Proceedings IEEE international conference on data mining, 2001 (ICDM 2001), pp 273–280

Kamdar T, Joshi A (2000) On creating adaptive web servers using weblog mining. Technical report TR-CS- 00-05, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County

Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis, vol 344. Wiley, New York

Kley T, Volgushev S, Dette H, Hallin M (2016) Quantile spectral processes: asymptotic analysis and inference. Bernoulli 22(3):1770–1807

Kou G, Peng Y, Wang G (2014) Evaluation of clustering algorithms for financial risk analysis using mcdm methods. Inf Sci 275(C):1–12

Krishnapuram R, Joshi A, Yi L (1999) A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering. In: Snippet clustering, in proceedings of IEEE international conference on fuzzy systems - FUZZIEEE99, Korea, pp 1281–1286

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

Kwon SH (1998) Cluster validity index for fuzzy clustering. Electron Lett 34(22):2176–2177

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

Lafuente-Rego B, Vilar JA (2016b) Fuzzy clustering of series using quantile autocovariances. In: Douzal-Chouakria A, Vilar JA, Marteau PF (eds) Advanced analysis and learning on temporal data: first

ECML PKDD workshop, AALTD 2015, Porto, Portugal, September 11, 2015. Springer International Publishing, Cham, Revised Selected Papers, pp 49–64

Lee J, Rao S (2012) The quantile spectral density and comparison based tests for nonlinear time series. Unpublished manuscript, Department of Statistics, Texas A&M University, College Station, USA, arXiv:1112.2759v2

Li TH (2014) Quantile periodograms. J Am Stat Assoc 107(498):765–776

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

Linton O, Whang YJ (2007) The quantilogram: With an application to evaluating directional predictability. J Econom 141(1):250–282

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

Maharaj EA (1999) Comparison and classification of stationary multivariate time series. Pattern Recognit 32(7):1129–1138

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

Maharaj EA, D'Urso P (2010) A coherence-based approach for the pattern recognition of time series. Physica A 389(17):3516–3537

Maharaj EA, D'Urso P (2011) Fuzzy clustering of time series in the frequency domain. Inf Sci 181(7):1187–1211

Maharaj EA, Alonso AM, D'Urso P (2015) Clustering seasonal time series using extreme value analysis: an application to spanish temperature time series. Commun Stat 1(4):175–191

McBratney A, Moore A (1985) Application of fuzzy sets to climatic classification. Agric For Meteorol 35(1–4):165–185

Montero P, Vilar JA (2014) TSclust: An R package for time series clustering. J Stat Softw 62(1):1–43

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

Otranto E (2010) Identifying financial time series with similar dynamic conditional correlation. Comput Stat Data Anal 54(1):1–15

Peña D (2011) Outliers, influential observations, and missing data. Wiley, New York, chap 6:136–170

Peng Y, Wang G, Kou G, Shi Y (2011) An empirical study of classification algorithm evaluation for financial risk prediction. Appl Soft Comput 11(2):2906–2915

Pértega S, Vilar JA (2010) Comparing several parametric and nonparametric approaches to time series clustering: a simulation study. J Classif 27(3):333–362

Pham TD, Tran LT (1981) On the first-order bilinear time series model. J Appl Probab 18(3):617–627

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

Rani S, Sikka G (2012) Recent techniques of clustering of time series data: a survey. Int J Comput Appl 52(15):1–9

Runkler TA, Bezdek JC (1999) Alternating cluster estimation: a new tool for clustering and function approximation. IEEE Trans Fuzzy Syst 7(4):377–393

Slaets L, Claeskens G, Hubert M (2012) Phase and amplitude-based clustering for functional data. Comput Stat Data Anal 56(7):2360–2374

Tarpey T, Kinateder KK (2003) Clustering functional data. J Classif 20(1):093–114

Tsay RS (1986) Time series model specification in the presence of outliers. J Am Stat Assoc 81(393):132–141

Tsay RS (2016) Some methods for analyzing big dependent data. J Bus Econ Stat 34(4):673–688

Vilar JA, Pértega S (2004) Discriminant and cluster analysis for gaussian stationary processes: local linear fitting approach. J Nonparametr Stat 16(3–4):443–462

Vilar JM, Vilar JA, Pértega S (2009) Classifying time series data: a nonparametric approach. J Classif 26(1):3–28

Vilar JA, Alonso AM, Vilar JM (2010) Non-linear time series clustering based on non-parametric forecast densities. Comput Stat Data Anal 54(11):2850–2865

Vilar JA, Lafuente-Rego B, D'Urso P (2018) Quantile autocovariances: a powerful tool for hard and soft partitional clustering of time series. Fuzzy Sets Syst 340:38–72

Wedel M, Kamakura WA (1998) Market segmentation: conceptual and methodological foundations. Kluwer Academic Press, Boston

Wu KL, Yang MS (2002) Alternative c-means clustering algorithms. Pattern Recognit 35(10):2267–2278

Xie XL, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13(8):841–847

Xiong Y, Yeung DY (2004) Time series clustering with ARMA mixtures. Pattern Recognit 37(8):1675–1689

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

Affiliations

B. Lafuente-Rego[1] · P. D'Urso[2] · J. A. Vilar[1]

P. D'Urso
pierpaolo.durso@uniroma1.it

J. A. Vilar
jose.vilarf@udc.es

[1]   Research Group on Modeling, Optimization and Statistical Inference (MODES), Department of Mathematics, Computer Science Faculty, University of A Coruña, 15071 A Coruña, Spain

[2]   Dipartimento di Scienze Sociali ed Economiche, Sapienza University of Rome, Pza. Aldo Moro, 5, 00185 Rome, Italy