

GARCH-based robust clustering of time series

Post-print version of the following publication: | Versione post-print della seguente pubblicazione:

Original Citation/Citazione:

De Giovanni, Livia; D'Urso, Pierpaolo; Massari, Riccardo. (2016). GARCH-based robust clustering of time series. FUZZY SETS AND SYSTEMS, (ISSN: 0165-0114), 305: 1-28. Doi: 10.1016/j.fss.2016.01.010.

Availability/Disponibilità:

This version is available at: [11385/164603](https://iris.luiss.it/11385/164603) since: 2016-04-26T01:38:43Z - Questa versione è disponibile alla pagina: [11385/164603](https://iris.luiss.it/11385/164603) dal: 2016-04-26T01:38:43Z

Publisher/Casa editrice:

Elsevier

Published version/Pubblicato:

DOI: <https://dx.doi.org/10.1016/j.fss.2016.01.010>

License/Licenza:

Attribution-NonCommercial-NoDerivatives 4.0 International

Availability/Termini d'uso:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. Works made available under a Creative Commons license can be used according to the terms and conditions of said license. For all terms of use and more information see the publisher's website. | I termini e le condizioni relativi al riutilizzo della presente versione della pubblicazione sono disciplinati dalla politica editoriale. Le opere messe a disposizione con licenze Creative Commons possono essere utilizzate conformemente ai termini e alle condizioni previste da tali licenze. Per l'insieme delle condizioni di utilizzo e per ulteriori informazioni si rinvia al sito web dell'editore.

This item was downloaded from IRIS Luiss (<https://iris.luiss.it/>). When citing, please refer to the published version. | Questo documento è stato scaricato da IRIS Luiss (<https://iris.luiss.it/>). Per la citazione, fare riferimento alla versione pubblicata sul sito dell'editore.

(Article begins on next page | Il contributo inizia nella pagina successiva)

Available online at www.sciencedirect.com**ScienceDirect**

Fuzzy Sets and Systems ●●● (●●●●) ●●●—●●●

FUZZY
sets and systemswww.elsevier.com/locate/fss

GARCH-based robust clustering of time series

Pierpaolo D'Urso^{a,*}, Livia De Giovanni^b, Riccardo Massari^a^a *Dipartimento di Scienze Sociali ed Economiche, Sapienza University of Rome, P.zza Aldo Moro, 5, 00185 Rome, Italy*^b *Dipartimento di Scienze Politiche, LUISS Guido Carli, Viale Romania, 32, 00197 Rome, Italy*

Received 6 February 2015; received in revised form 2 December 2015; accepted 25 January 2016

Abstract

In this paper we propose different robust fuzzy clustering models for classifying heteroskedastic (volatility) time series, following the so-called model-based approach to time series clustering and using a partitioning around medoids procedure. The proposed models are based on a GARCH parametric modeling of the time series, i.e. the unconditional volatility and the time-varying volatility GARCH representation of the time series. We first suggest a timid robustification of the fuzzy clustering. Then, we propose three robust fuzzy clustering models belonging to the so-called metric, noise and trimmed approaches, respectively. Each model neutralizes the negative effects of the outliers in the clustering process in a different manner. In particular, the first robust model, based on the metric approach, achieves its robustness with respect to outliers by taking into account a “robust” distance measure; the second, based on the noise approach, achieves its robustness by introducing a noise cluster represented by a noise prototype; the third, based on the trimmed approach, achieves its robustness by trimming away a certain fraction of outlying time series. The usefulness and effectiveness of the proposed clustering models is illustrated by means of a simulation study and two applications in finance and economics.

© 2016 Elsevier B.V. All rights reserved.

Keywords: Heteroskedastic time series; Unconditional and time-varying volatility; GARCH model; Fuzzy partitioning around medoids; Outliers; Robust metric; Noise cluster; Trimming; Volatilities daily stocks returns; International stock-market volatility daily returns

1. Introduction

In the last two decades, the literature on time series clustering has increased considerably with different methodological approaches and with a large range of applications in many different fields including finance, economics, econophysics, marketing and new frontiers of marketing—such as digital marketing and neuromarketing—genetics (microarray gene analysis), environmetrics, geology, telecommunications and biomedical sciences. See, e.g., [1–18].

From a methodological point of view, time series clustering methods can be classified into three classes [3,19]:

* Corresponding author.

E-mail addresses: pierpaolo.durso@uniroma1.it (P. D'Urso), ldegiovanni@luiss.it (L. De Giovanni), riccardo.massari@uniroma1.it (R. Massari).

<http://dx.doi.org/10.1016/j.fss.2016.01.010>

0165-0114/© 2016 Elsevier B.V. All rights reserved.

1. *Observation-based clustering*: the time series clustering methods belonging to this approach are based on the actual observed time series—considering all the observations or positions of the time series or suitable transformations of the observed time series, e.g., the so-called velocities, accelerations, polygonal representations of the observed time series [1,2,4,6,10,20–22]. This clustering approach is particularly useful with short time series. In addition, other observation based clustering methods are built upon the so-called Dynamic Time Warping (DTW), which is a well-known technique for finding an optimal alignment between two given (time-dependent) sequences under certain restrictions (see, e.g., [23]). See [19] for more details.
2. *Feature-based clustering*: in this case, the methods are based on suitable features derived for the time series. In fact, as remarked by Caiado et al. [19] “if a time series consists of a large number of observations, clustering time series based on these observations is not a desirable option because of the noise that is present, and the fact that the autocorrelation structure of the time series is ignored”. In the literature, different feature-based clustering methods have been developed to address the problem of clustering noisy raw time series data. As a broad distinction, methods can be classified as follows:
 - methods based on time domain features: i.e., autocorrelation function (ACF) [6,24–26], partial autocorrelation function (PACF) [25] and inverse autocorrelation function (IACF) [25];
 - methods based on frequency domain features: i.e., based on the periodogram and its transformations [13,26], coherence [27] and cepstral [13];
 - methods based on wavelet features: i.e., wavelet decomposition [11,14,18].
 See [19] for more details.
3. *Model-based clustering*: these methods are based on the features of the models fitted to the time series. In particular, for these methods, it is assumed that a set of time series generated from the same model have similar patterns. Time series are clustered by means of parameter estimates or by means of the residuals of the fitted models [19]. In this class, several time series clustering methods are based on the ARIMA representation of the time series (see, e.g., [16,17,28–34]). Notice that most of these methods are devoted to capturing the structure of the mean of the process hypothesized as generator of the data, whereas little attention has been put on the variance. This approach is correct when dealing with clustering based on ARIMA models and in the presence of homoskedastic variance; in fact, in this case, the variance is a function of the process parameters, so that it is implicitly considered in the clustering process [7]. Dealing with heteroskedastic time series—or volatility time series—the comparison of the dynamics of the variances is pivotal [7]. If the conditional variance follows a stochastic process, heteroskedastic time series can be represented by GARCH models. In the literature, some time series clustering methods based on GARCH models have been proposed—see, for instance, [7–9,12,15]. Finally, in the model-based time series clustering approach we can also include the methods based on functional representation (see, e.g., [35]) and on nonparametric or semiparametric (e.g. splines) (see, e.g., [36]) representation of the time series. For more details on this methodological class, see [19].

In this paper, the model-based approach is adopted. In particular, given the motivations shown in 3) we focus our attention on the model-based clustering of time series based on a GARCH parametric modeling of the time series.

The main contribution of this paper is to propose different clustering models which have the following features:

1. they are based on a “Partitioning Around Medoids” (PAM) approach;
2. they follow the fuzzy clustering approach to generate a fuzzy partition of the heteroskedastic time series;
3. they are robust to the presence of outliers, i.e. time series with anomalous patterns;
4. they are based on a GARCH parametric modeling of heteroskedastic time series, more specifically on the different components of the volatility of the GARCH representation of time series, i.e. the unconditional volatility (uv) and the time varying volatility (tvv). Notice that this approach also allows to compare time series of different lengths;
5. they are based on a suitable distance measure for heteroskedastic time series which weights differently the different components of the volatility of the GARCH representation of time series.

Given the features of the proposed models, the aim is to classify heteroskedastic time series by taking into account the different components of their volatility, in such a way that the presence of anomalous time series is detected and neutralized in the clustering process.

The paper is organized as follows. In Section 2, we briefly introduce the concepts of unconditional and time-varying volatility of GARCH models and a suitable distance measure for the volatility characteristics derived from the GARCH representation of time series. In Section 3, following a fuzzy approach different GARCH model-based robust clustering models for heteroskedastic time series are proposed. A simulation study and two applications are shown, respectively, in Sections 4 and 5. Final remarks are made in Section 6.

2. GARCH models and distance measure for heteroskedastic time series

2.1. Unconditional, minimum and time-varying volatilities

Let y_t ($t = 1, \dots, T$) be a time series, where t is the time index, modeled as the sum of a constant term μ and a zero mean heteroskedastic univariate process ε_t (heteroskedastic disturbance):

$$\begin{aligned} y_t &= \mu + \varepsilon_t \\ \varepsilon_t &= u_t \sqrt{h_t} \end{aligned} \quad (2.1)$$

where u_t is a univariate white noise process with mean 0 and variance 1. The conditional variance h_t follows a GARCH(p, q) process as [37]:

$$h_t = \gamma + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 h_{t-1} + \dots + \beta_q h_{t-q} \quad (2.2)$$

with

$$\gamma > 0, 0 \leq \alpha_i \leq 1, 0 \leq \beta_j \leq 1 \quad (i = 1, \dots, p; j = 1, \dots, q), \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$$

Consider the difference between ε_t^2 and h_t :

$$\eta_t = \varepsilon_t^2 - h_t \quad (2.3)$$

where η_t are zero mean errors, uncorrelated with past information. After simple algebra, from Equation (2.2) the squared disturbances ε_t^2 can be represented as an ARMA(p^*, q) process:

$$\varepsilon_t^2 = \gamma + \sum_{i=1}^{p^*} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 - \sum_{j=1}^q \beta_j + \eta_t \quad (2.4)$$

where $p^* = \max(p, q)$, $\alpha_i = 0$ for $i > p$, if $p^* = q$, $\beta_j = 0$ for $j > q$ if $p^* = p$.

The parameters of the GARCH(p, q) model are $(\gamma, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$. Hence, the GARCH process will be denoted also as GARCH($\gamma, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$).

Given the usual stationarity and invertibility restrictions on the roots of $(1 - (\alpha_1 + \beta_1)z - \dots - (\alpha_i + \beta_i)z^i - \dots - (\alpha_{p^*} + \beta_{p^*})z^{p^*})$ and $(1 - \beta_1 z - \dots - \beta_j z^j - \dots - \beta_q z^q)$, from (2.4) ε_t^2 can be expressed as an infinite autoregressive AR(∞) model, after recursive substitution:

$$\varepsilon_t^2 = \frac{\gamma}{1 - \sum_{j=1}^q \beta_j} + \sum_{k=1}^{\infty} \pi_k \varepsilon_{t-k}^2 + \eta_t \quad (2.5)$$

As it is well known, indicating with ϕ_i the generic AR parameter and with θ_j the generic MA parameter of an ARMA(p, q) model, the recursive formula:

$$\pi_k - \sum_{j=1}^q \theta_j \pi_{k-j} = \phi_k \quad k = 0, 1, \dots \quad (2.6)$$

provides the sequence of parameters π_k . In (2.6) $\phi_0 = 1$, $\phi_i = 0$ for $i > p$ and $\pi_k = 0$ for $k < 0$. From (2.4) and (2.6) the π_k parameters are:

$$\pi_k = (\alpha_k + \beta_k) - \sum_{j=1}^q \beta_j \pi_{k-j} \quad (2.7)$$

From (2.5) the expected volatility at time $t + 1$, given the information available at time t , can be split in a constant part and a time-varying part (which depends on the time history of the volatility) [7]:

$$E_t(\varepsilon_{t+1}^2) = \frac{\gamma}{1 - \sum_{j=1}^q \beta_j} + \sum_{k=1}^{\infty} \pi_k \varepsilon_{t-k}^2 \quad (2.8)$$

The *unconditional volatility* (uv) is then given by the unconditional expected value of ε_{t+1}^2 :

$$uv = E(\varepsilon_{t+1}^2) = \frac{\gamma}{(1 - \sum_{j=1}^q \beta_j)(1 - \sum_{k=1}^{\infty} \pi_k)} \quad (2.9)$$

The time-varying part of the volatility is an infinite weighted sum of unobservable random variables (see (2.8)). The null time-varying volatility is obtained when each of the π_k coefficient is equal to zero. Considering the metric introduced by Piccolo [28] between two ARIMA models as the Euclidean distance between the coefficients of the AR(∞) representation of the two models, the distance of a time series from the case of null time-varying volatility (hence the case of constant volatility) is given by Otranto [7], yielding the expression of *time varying volatility*:

$$tvv = \left(\sum_{k=1}^{\infty} \pi_k^2 \right)^{\frac{1}{2}} \quad (2.10)$$

2.2. A distance measure for heteroskedastic time series

The distance between each pair of time series is measured by comparing the *unconditional volatility* and the *time-varying volatility* of the time series, i.e. by separately considering the distances for the *unconditional volatility* and the *time-varying volatility* estimated parameters of the GARCH representation of the time series and using a suitable weighting system for such distance components. Thus, by considering the i -th and i' -th units, we have:

$${}_{uv,tvv}d_{ii'} = [w_1^2 {}_{uv}d_{ii'}^2 + w_2^2 {}_{tvv}d_{ii'}^2]^{\frac{1}{2}} = [w_1^2 (uv_i - uv_{i'})^2 + w_2^2 (tvv_i - tvv_{i'})^2]^{\frac{1}{2}} \quad (2.11)$$

where ${}_{uv}d_{ii'}^2 = (uv_i - uv_{i'})^2$ is the squared Euclidean distance between the unconditional volatility of units i and i' (uv_i and $uv_{i'}$ respectively); ${}_{tvv}d_{ii'}^2 = (tvv_i - tvv_{i'})^2$ is the squared Euclidean distance between the time-varying volatility of units i and i' (tvv_i and $tvv_{i'}$ respectively); $w_1, w_2 \geq 0$ are suitable weights for the unconditional and time-varying components. These weights can be fixed subjectively a priori by considering external or subjective conditions (internal weighting system) or can be computed objectively within a suitable clustering procedure (external weighting system).

We assume the following conditions: $w_1 + w_2 = 1$ (normalization condition) and $w_1, w_2 \geq 0$. We can set $w_1 = w$, $w_2 = (1 - w)$.

To select the weights, an objective criterion has been adopted. Therefore, the weight values are not fixed a priori but are computed via a minimization algorithm. Thus we obtain weights that minimize the loss function with respect to the optimal values of w_1, w_2 .

The distance measure (2.11) has the following features:

- i) ${}_{uv,tvv}d_{ii'}$ is a metric. The properties of identity, non-negativity, symmetry are easily verified. Also the triangular inequality property is satisfied, as it can be verified after some algebra.
- ii) ${}_{uv,tvv}d_{ii'}$ is computationally easy and theoretically intuitive.

The weights w_1, w_2 are intrinsically associated to the characteristics of the time series captured by the unconditional and time-varying volatility; thus the weights allows to tune properly the influence of the two components of the volatility of the time series when calculating the distance.

Series with similar unconditional volatility can have different dynamics, characterized by different time-varying volatilities; conversely series with similar time-varying volatility can have different unconditional volatilities.

Through the normalization condition $w_1 + w_2 = 1$, we can easily assess, in a comparative fashion, the contribution of the components in the computation of ${}_{uv, tvv}d_{ii'}$.

The measure ${}_{uv, tvv}d_{ii'}$ is utilized for comparisons within a set of data rather than for examining a single pair of data. Thus, for a given dataset, the weighting system is optimal only for the dataset involved.

If two series have the same unconditional volatility and the same time-varying volatility, it is not possible to conclude that they are generated by the same GARCH process. In fact the equality of the two unconditional volatilities and the two time-varying volatilities is relative to two nonlinear combinations of the parameters of the GARCH representation. The case of equality of the two data generating processes is obtained if and only if the constant γ and the parameters α_i and β_j ($i = 1, \dots, p$; $j = 1, \dots, q$) are the same for the two GARCH models. This is a stronger requirement with respect to the equality of the unconditional volatilities and the time-varying volatilities. It is obvious that the equality of the parameters of two time series implies equal unconditional and time-varying volatilities (but not vice versa).

In the paper robust fuzzy clustering models for heteroskedastic time series based on the estimated parameters of the GARCH representation are introduced. Specifically, the proposed models focus on the contribution of the parameters of the GARCH representation to the unconditional and time-varying volatility.

Other clustering methods based on the volatility related to the GARCH representation of time series have been proposed in the literature.

Caiado and Crato [5] have introduced a dissimilarity measure between two volatility time series based on the Euclidean distance among the estimated parameters γ , α_i and β_j ($i = 1, \dots, p$; $j = 1, \dots, q$) of the GARCH representation of the time series weighted with their estimated covariances. Then a hierarchical agglomerative clustering method augmented with the dissimilarity measure proposed is used.

Otranto [7] has proposed a three level clustering of time series. The first level groups time series on the basis of the unconditional volatility. The second level subgroups time series with similar time-varying volatility within the groups characterized by similar unconditional volatility. Finally, a more accurate classification is obtained distinguishing, within the groups with equal unconditional and time-varying volatilities, the time series with equal parameters of the GARCH representation. The similarity of the time series in each of the three levels is determined using the result (p -value) of a Wald statistical test on the unconditional volatility, on the time-varying volatility, on the estimated GARCH parameters, respectively. The threshold on the p -values allows the automatic detection of the number of clusters.

D'Urso et al. [15] have introduced two fuzzy clustering models of volatility time series in the framework of a partitioning around medoids approach. One is based on the Caiado and Crato dissimilarity between the estimated parameters of the GARCH representation of the volatility time series [5] and the other on the autoregressive distance measure [28] applied by [7] to GARCH processes generating the volatility time series.

With respect to [5] and to [7,12] the novelty of our proposed models are: i) the use of a distance based on the components of the volatility (unconditional and time-varying) objectively weighted; ii) the robustness to the presence of outliers of the clustering algorithms in which the distance works; iii) the fuzziness of the partition.

With respect to [15] the novelty of the proposed models are: i) the use of a distance based on the components of the volatility (unconditional and time-varying) objectively weighted; ii) the robustness to the presence of outliers of the clustering algorithms in which the distance works.

The $N \times T$ data matrix \mathbf{Y} consists of N time series observed in T times periods:

$$\mathbf{Y} = \begin{pmatrix} y_{11} & \dots & y_{1t} & \dots & y_{1T} \\ \vdots & & \vdots & & \vdots \\ y_{i1} & \dots & y_{it} & \dots & y_{iT} \\ \vdots & & \vdots & & \vdots \\ y_{N1} & \dots & y_{Nt} & \dots & y_{NT} \end{pmatrix} \quad (2.12)$$

In the GARCH representation it turns into:

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} \\ \vdots & \vdots \\ v_{i1} & v_{i2} \\ \vdots & \vdots \\ v_{N1} & v_{N2} \end{pmatrix} = \begin{pmatrix} \tilde{v}'_1 \\ \vdots \\ \tilde{v}'_i \\ \vdots \\ \tilde{v}'_N \end{pmatrix} \quad (2.13)$$

where $v_{i1} = uv_i$ and $v_{i2} = tvv_i$.

Notice that it is not strictly necessary that the time series are of equal length. However, the fact that the time series could have different length does not affect overall results, since the analysis is based on their GARCH representations. Hence, to avoid a more cumbersome notation, we stick to the case in which all time series have the same length. Finally, we observe that the orders p and q of the GARCH process are not necessarily equal.

3. Robust fuzzy clustering models of time series

In this section we propose different robust clustering models for time series based on a GARCH modeling of the time series and a fuzzy partitioning around medoids (fuzzy PAM) approach. The motivations connected both to the fuzzy and PAM approaches are shown in the following sections.

3.1. Fuzzy motivation

We adopt a fuzzy approach in order to take into account the intrinsic (non-stochastic) uncertainty derived from the clustering of such complex data as time series, and to capture the switching or drifting nature of some time series in the clustering process. Fuzzy clustering allows one time series to be allocated to two or more cluster, with a membership degree that represents the (non-stochastic) uncertainty related to the assignment of the time series to each cluster. In particular, our clustering models inherit the several advantages of the fuzzy approach to cluster analysis [2,4,11,13,14,16,17,22,38,39]. In particular, as remarked by Hwang et al. [40]:

1. the fuzzy clustering algorithm is attractive because it is compatible with distribution-free procedure;
2. due to the difficulty of identifying a clear boundary between clusters in real world problems, the partial classification of fuzzy clustering appears more attractive than the deterministic classification of nonoverlapping clustering methods such as k -means [41,42];
3. the fuzzy clustering is computationally more efficient because dramatic changes in the value of cluster membership are less likely to occur in estimation procedures [41] and it has been shown to be less afflicted by local optima problems [43];
4. the memberships for a given set of respondents indicate whether there is a second-best cluster almost as good as the best cluster—a result which traditional clustering methods cannot uncover [44];
5. the fuzzy approach is preferable to the probabilistic approach—e.g., finite mixture approach—, because, generally, the latter approach assumes stringent distributional assumptions on data within unknown clusters. In the fuzzy clustering approach, no specific form of distributions for observed data (within each cluster) needs to be assumed a priori for the proposed method. In addition, the fuzzy approach is less afflicted by the problem of convergence compared to the finite mixture approach [40].

In addition, we have the following motivations justifying the fuzzy approach for the clustering of time series [22,45]:

6. Greater sensitivity in capturing the details characterizing the time series. In many cases, since the dynamics of the time series are drifting or switching, the standard (non-fuzzy) clustering approaches are likely to miss this underlying structure. The switches, which are usually vague, can be naturally treated by means of fuzzy clustering.
7. Greater adaptivity in defining the prototype time series. This can be better appreciated when the observed time patterns 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. PAM motivation

The task of clustering is useful to suitably summarize the information in a set of time series. Instead of considering all time series one can analyze prototypical time series, i.e., time series that retain the main features of similar time series classified in the same group. To this end, we adopt the so called “Partitioning Around Medoid” (PAM) approach [46] in a fuzzy framework, i.e. the Fuzzy C -Medoids (FCMd) approach [47,48]. With FCMd, the prototypes of each group, henceforth “medoid time series”, are time series actually observed and not “virtual” time series, like the “centroids” derived with a Fuzzy C -Means (FCM) approach [49]. The possibility of obtaining non-fictitious representative time series in the clusters is very appealing and useful in a wide range of applications. This is very important for the interpretation of the selected clusters. In fact, as remarked by Kaufman and Rousseeuw [46] “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”. Notice that, since the medoids are observed time series, by using the PAM approach, the constraints on the GARCH coefficients of the prototypical time series are always satisfied (provided that the observed time series are stationary and invertible), which is not necessarily guaranteed when using the k -means approach.

3.3. GARCH-based fuzzy clustering models

In the literature on fuzzy clustering, a very interesting line of research has focused on how to use fuzzy clustering in the presence of anomalous data [50–56].

To this purpose, in the present section, we first propose the GARCH-based Fuzzy C -Medoids Clustering model (GARCH-FCMdC model) (see Section 3.3.1).

This model, inheriting the features of the non-fuzzy clustering version (see, [36,57]), provides only a timid robustification of the fuzzy clustering. The model alleviates the negative effects of the presence of outliers in the dataset but does not solve the problem. For this reason, we then propose the following robust models which represent different types of robustification of the GARCH-FCMdC model:

- GARCH-based Exponential Fuzzy C -Medoids Clustering model (GARCH-E-FCMdC model) (section 3.3.2),
- GARCH-based Fuzzy C -Medoids clustering with Noise Cluster model (GARCH-NC-FCMdC model) (section 3.3.3),
- GARCH-based Trimmed Fuzzy C -Medoids Clustering model (GARCH-Tr-FCMdC model) (section 3.3.4).

All the above clustering models include in the objective function the distance measure (2.11).

Each model neutralizes in a different manner the negative effects of the outliers in the clustering process. In particular, the GARCH-E-FCMdC model achieves its robustness with respect to outliers by taking into account a “robust” distance measure; the GARCH-NC-FCMdC model achieves its robustness by introducing a noise cluster represented by a noise prototype; the GARCH-Tr-FCMdC model achieves its robustness by trimming away a certain fraction of anomalous time series.

3.3.1. GARCH-based Fuzzy C -Medoids Clustering model (GARCH-FCMdC)

By considering the data matrix \mathbf{Y} and the distance measure (2.11), in which the weights are objectively computed during the clustering process, we can classify units with fuzzy information within a fuzzy framework, by means of the GARCH Fuzzy C -Medoids Clustering model (GARCH-FCMdC), characterized as follows:

$$\begin{aligned} \min : \mathcal{F}_{\text{GARCH-FCMdC}}(\mathbf{U}, \mathbf{V}, w_1, w_2) &\equiv \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m u_{ic}^{m,uv} d_{ic}^2 \\ &= \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [w_1^2 d_{ic}^2 + w_2^2 d_{ic}^2] \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [w_1^2 (uv_i - uv_c)^2 + w_2^2 (tvv_i - tvv_c)^2] \\
 &\sum_{c=1}^C u_{ic} = 1; w_1 + w_2 = 1; w_1, w_2 \geq 0
 \end{aligned} \tag{3.1}$$

where $m > 1$ is a weighting exponent that controls the fuzziness of the obtained partition; $\tilde{v}'_c = (v_{c1}, v_{c2})$ (see (2.13)) is the medoid for cluster c ; u_{ic} indicates the membership degree of the i -th unit in the c -th cluster; \mathbf{U} is the $N \times C$ matrix of the membership degrees.

By solving the constrained quadratic minimization problem (3.1) via the Lagrangian multiplier method we obtain the optimal solution u_{ic} and w_1, w_2 . We consider the following Lagrangian function:

$$L_m(u_{ic}, \lambda, w_1, w_2) = \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [w_1^2 uv d_{ic}^2 + w_2^2 tvv d_{ic}^2] - \lambda \left(\sum_{c=1}^C u_{ic} - 1 \right) \tag{3.2}$$

We take the first partial derivatives with respect to u_{ic} and λ and set them to 0:

$$\frac{\partial L_m(u_{ic}, \lambda, w_1, w_2)}{\partial u_{ic}} = 0 \Leftrightarrow m u_{ic}^{m-1} [w_1^2 uv d_{ic}^2 + w_2^2 tvv d_{ic}^2] - \lambda = 0 \tag{3.3}$$

$$\frac{\partial L_m(u_{ic}, \lambda, w_1, w_2)}{\partial \lambda} = 0 \Leftrightarrow \sum_{c=1}^C u_{ic} - 1 = 0 \tag{3.4}$$

From (3.3) we have:

$$u_{ic} = \left(\frac{\lambda}{m} \right)^{\frac{1}{m-1}} \left[\frac{1}{w_1^2 uv d_{ic}^2 + w_2^2 tvv d_{ic}^2} \right]^{\frac{1}{m-1}} \tag{3.5}$$

and, upon substituting (3.5) in (3.4),

$$\left(\frac{\lambda}{m} \right)^{\frac{1}{m-1}} \sum_{c=1}^C \left[\frac{1}{w_1^2 uv d_{ic}^2 + w_2^2 tvv d_{ic}^2} \right]^{\frac{1}{m-1}} = 1 \tag{3.6}$$

Therefore

$$\left(\frac{\lambda}{m} \right)^{\frac{1}{m-1}} = \frac{1}{\sum_{c=1}^C \left[\frac{1}{w_1^2 uv d_{ic}^2 + w_2^2 tvv d_{ic}^2} \right]^{\frac{1}{m-1}}} \tag{3.7}$$

Finally, substituting (3.7) into (3.5) yields to:

$$\begin{aligned}
 u_{ic} &= \frac{1}{\sum_{c'=1}^C \left[\frac{w_1^2 uv d_{ic}^2 + w_2^2 tvv d_{ic}^2}{w_1^2 uv d_{ic'}^2 + w_2^2 tvv d_{ic'}^2} \right]^{\frac{1}{m-1}}} \\
 &= \frac{1}{\sum_{c'=1}^C \left[\frac{w_1^2 (uv_i - uv_c)^2 + w_2^2 (tvv_i - tvv_c)^2}{w_1^2 (uv_i - uv_{c'})^2 + w_2^2 (tvv_i - tvv_{c'})^2} \right]^{\frac{1}{m-1}}}.
 \end{aligned} \tag{3.8}$$

By taking the partial derivative of (3.2) with respect to w_1 , noting that $w_2 = 1 - w_1$, it follows:

$$\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [w_1 uv d_{ic}^2 - (1 - w_1) tvv d_{ic}^2] = 0$$

from which we obtain:

$$\begin{aligned}
 w_1 &= \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m u_{ic}^{tv} d_{ic}^2}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (u_{ic} d_{ic}^2 + u_{ic}^{tv} d_{ic}^2)} \\
 &= \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (u_{ic}^{tv} - u_{ic})^2}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [(u_{ic} - u_{ic}^{tv})^2 + (u_{ic}^{tv} - u_{ic})^2]}
 \end{aligned} \tag{3.9}$$

In a similar way, we can derive w_2 :

$$\begin{aligned}
 w_2 &= \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m u_{ic}^{uv} d_{ic}^2}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (u_{ic} d_{ic}^2 + u_{ic}^{uv} d_{ic}^2)} \\
 &= \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (u_{ic} - u_{ic}^{uv})^2}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [(u_{ic} - u_{ic}^{uv})^2 + (u_{ic}^{uv} - u_{ic})^2]}
 \end{aligned} \tag{3.10}$$

The medoids are obtained by solving the following minimization problem:

$$q = \arg \min_{1 \leq i' \leq N} \sum_{i=1}^N u_{ic}^m [w_1^2 u_{ic} d_{ii'}^2 + w_2^2 u_{ic}^{tv} d_{ii'}^2] = \arg \min_{1 \leq i' \leq N} g(\tilde{v}_i) \tag{3.11}$$

where

$$g(\tilde{v}_i) = \sum_{i=1}^N u_{ic}^m [w_1^2 u_{ic} d_{ii'}^2 + (1 - w_1)^2 u_{ic}^{tv} d_{ii'}^2].$$

For $c = 1, \dots, C$, if the value of q is smaller than $g(\tilde{v}_c)$ the new medoid for cluster c is \tilde{v}_q . In the following, we show the algorithm for the GARCH-FCMdC model.

Algorithm (GARCH-FCMdC model).

Step 0: Fix the power of the membership degrees (m), the number of clusters (C) and the maximum number of iterations ($maxiter$). Generate randomly the membership degree matrix $\mathcal{U}^{(0)}$ subject to conditions in (3.1), and pick C medoids in the data matrix. Set $\tilde{\mathcal{H}}$ the submatrix containing the C medoids, \tilde{v}_c being the medoid for cluster c .

Step 1: Compute $w_1^{(s-1)}$ according to (3.9) using $\mathcal{U}^{(s-1)}$ and $\tilde{\mathcal{H}}^{(s-1)}$, where $s \geq 1$ denotes the iteration number.

Step 2: Update the medoid matrix $\tilde{\mathcal{H}}^{(s)}$: for $c = 1, \dots, C$ set

$$q = \arg \min_{1 \leq i' \leq N} \sum_{i=1}^N u_{ic}^m [w_1^2 u_{ic} d_{ii'}^2 + (1 - w_1)^2 u_{ic}^{tv} d_{ii'}^2]$$

and $\tilde{v}_c^{(s)} = \tilde{v}_q$. Update $\mathcal{U}^{(s)}$ as in (3.8), using $\tilde{\mathcal{H}}^{(s)}$ and $w_1^{(s-1)}$ (for the unit i'' that is medoid for cluster l set $u_{i''l} = 1$; 0 otherwise).

Step 3: If $\tilde{\mathcal{H}}^{(s-1)} \equiv \tilde{\mathcal{H}}^{(s)}$, or iteration number $s = maxiter$ the algorithm has converged, otherwise go to step 1.

3.3.2. GARCH-based Exponential Fuzzy C-Medoids Clustering model (GARCH-E-FCMdC)

In this section, we introduce a robust fuzzy clustering model belonging to the metric approach, the GARCH-based Exponential Fuzzy C-Medoids Clustering model (GARCH-E-FCMdC). This model achieves its robustness with respect to outliers by taking into account the following ‘‘robust’’ distance measure [52,58]:

$$\begin{aligned}
 u_{ic}^{exp} d_{ii'} &\equiv [1 - \exp\{-\beta u_{ic} d_{ii'}^2\}]^{\frac{1}{2}} \\
 &\equiv [1 - \exp\{-\beta (w_1^2 u_{ic} d_{ii'}^2 + w_2^2 u_{ic}^{tv} d_{ii'}^2)\}]^{\frac{1}{2}} \\
 &= [1 - \exp\{-\beta [w_1^2 (u_{ic} - u_{ic}^{tv})^2 + w_2^2 (u_{ic}^{tv} - u_{ic})^2]\}]^{\frac{1}{2}}
 \end{aligned} \tag{3.12}$$

where β is a suitable parameter (positive constant) determined according to the variability of the data.

In particular, following [58], it is easy to prove that (3.12) is a metric.

Notice that, within a fuzzy k -means clustering framework, Wu and Yang [52] describe a theoretical foundation to asseverate that k -means clustering models based on the “exponential distance” are more robust than the models based on the Euclidean distance.

By considering the “exponential distance”, we have the following GARCH-E-FCMdC model:

$$\begin{aligned} \min : \mathcal{F}_{\text{GARCH-E-FCMdC}}(\mathbf{U}, \mathbf{V}, w_1, w_2) &\equiv \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (u_{v, tvv}^{exp} d_{ic}^2) \\ &= \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [1 - \exp\{-\beta[w_1^2(uv_i - uv_c)^2 + w_2^2(tvvi - tvvc)^2]\}] \end{aligned} \tag{3.13}$$

The model (3.13) represents an extension of Wu and Yang’s model [52] to heteroskedastic time series modeled via GARCH with medoid prototypes and a suitable (squared) distance measure for GARCH modeled time series. We obtain the optimal solutions u_{ic} and w_1, w_2 by solving the constrained quadratic minimization problem (3.13). In particular, from the following Lagrangian function (under the constraints in (3.1) for the membership degrees):

$$L_m(u_{ic}, \lambda, w_1, w_2) = \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (u_{v, tvv}^{exp} d_{ic}^2) - \lambda \left(\sum_{c=1}^C u_{ic} - 1 \right) \tag{3.14}$$

By taking the partial derivatives of (3.14) with respect to u_{ic} and λ and by setting them to 0:

$$\begin{aligned} \frac{\partial L_m(u_{ic}, \lambda, w_1, w_2)}{\partial u_{ic}} = 0 &\Leftrightarrow m u_{ic}^{m-1} [u_{v, tvv}^{exp} d_{ic}^2] - \lambda = 0 \\ \frac{\partial L_m(u_{ic}, \lambda, w_1, w_2)}{\partial \lambda} = 0 &\Leftrightarrow \sum_{c=1}^C u_{ic} - 1 = 0 \end{aligned}$$

from which we obtain, with algebraic passages similar to those illustrated in (3.5)–(3.7):

$$\begin{aligned} u_{ic} &= \frac{1}{\sum_{c'=1}^C \left[\frac{u_{v, tvv}^{exp} d_{ic'}^2}{u_{v, tvv}^{exp} d_{ic}^2} \right]^{\frac{1}{m-1}}} \\ &= \frac{1}{\sum_{c'=1}^C \left[\frac{1 - \exp\{-\beta[w_1^2(uv_i - uv_c)^2 + w_2^2(tvvi - tvvc)^2]\}}{1 - \exp\{-\beta[w_1^2(uv_i - uv_{c'})^2 + w_2^2(tvvi - tvvc')^2]\}} \right]^{\frac{1}{m-1}}}. \end{aligned} \tag{3.15}$$

By taking the partial derivatives of (3.14) with respect to w_1 it follows:

$$\begin{aligned} \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m \left[\exp\{-\beta[w_1^2(uv_i - uv_c)^2 + w_2^2(tvvi - tvvc)^2]\} \right] \\ \cdot \beta \left[w_1 (u_{v, tvv} d_{ic}^2 + tvv d_{ic}^2) - tvv d_{ic}^2 \right] = 0 \end{aligned}$$

from which we obtain:

$$\begin{aligned} w_1 &= \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m tvv d_{ic}^2 (1 - u_{v, tvv}^{exp} d_{ic}^2)}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (u_{v, tvv} d_{ic}^2 + tvv d_{ic}^2) (1 - u_{v, tvv}^{exp} d_{ic}^2)} \\ &= \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (tvvi - tvvc)^2 \exp\{-\beta[w_1^2(uv_i - uv_c)^2 + w_2^2(tvvi - tvvc)^2]\}}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [(uv_i - uv_c)^2 + (tvvi - tvvc)^2] \exp\{-\beta[w_1^2(uv_i - uv_c)^2 + w_2^2(tvvi - tvvc)^2]\}} \end{aligned} \tag{3.16}$$

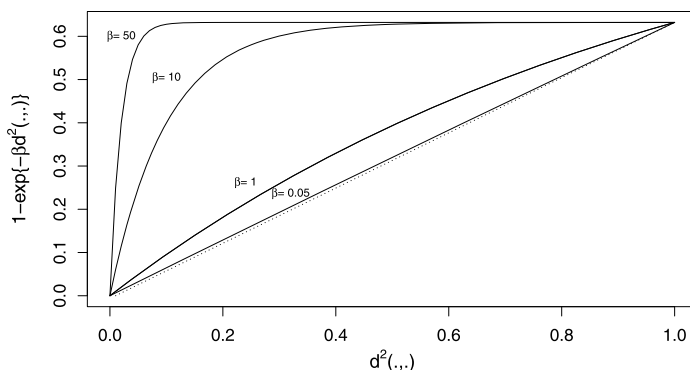


Fig. 3.1. Effect of the parameter β on the squared distance (3.12).

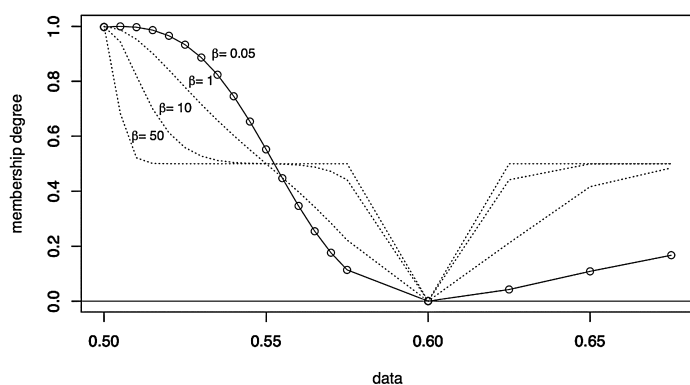


Fig. 3.2. Effect of the parameter β on the membership degrees (3.15).

In a similar way, we can derive w_2 :

$$\begin{aligned}
 w_2 &= \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m w_{ic}^2 (1 - \frac{exp}{uv, tvv} d_{ic}^2)}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (uv d_{ic}^2 + tvv d_{ic}^2) (1 - \frac{exp}{uv, tvv} d_{ic}^2)} \\
 &= \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m (uv_i - uv_c)^2 exp\{-\beta[w_1^2 (uv_i - uv_c)^2 + w_2^2 (tvv_i - tvv_c)^2]\}}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m [(uv_i - uv_c)^2 + (tvv_i - tvv_c)^2] exp\{-\beta[w_1^2 (uv_i - uv_c)^2 + w_2^2 (tvv_i - tvv_c)^2]\}}
 \end{aligned} \tag{3.17}$$

Notice that (3.16) and (3.17) can be solved only using an iterative method.

We observe that $\frac{exp}{uv, tvv} d_{ii'}^2$ is a monotone increasing function of $uv, tvv d_{ii'}$ which assigns a small weight to outliers.

The value of β , computed as the inverse of the variance in the data [52], appropriately affects the distance (3.12) and the membership degree (3.15) in terms of robustness to outliers.

The effect of the parameter β on the squared distance (3.12) is illustrated in Fig. 3.1.

In the presence of low variability of the data, increasing distances receive a weight lower than in the case of high variability.

The effect of the parameter β on the membership degree (3.15) is illustrated in Fig. 3.2.

Fig. 3.2 shows different membership curves for different parameters β obtained with the GARCH-E-FCMdc in the case of two clusters with centers in 0.5 and 0.6 (the same values of uv and tvv have been considered). The curve with circle points represents the GARCH-FCMdc membership curve. If β is very small (high variability in the data) the GARCH-E-FCMdc membership curve is very close to the GARCH-FCMdc membership curve which well represents fuzzy boundaries; if β is very large (low variability in the data) the GARCH-E-FCMdc membership curve (shaped

as a step function) is very different from to the GARCH-FCMdc membership curve as it assigns membership 0.5 to data that are slightly far from the centers, well representing the characteristic of separation between the clusters.

To solve the model (3.13), we can use the following algorithm.

Algorithm (GARCH-E-FCMdc model).

Step 0: Fix the power of the membership degrees (m), the number of clusters (C), the value of β and the maximum number of iterations ($maxiter$). Generate randomly the membership degree matrix $\mathbf{U}^{(0)}$ subject to conditions in (3.1), and pick C medoids in the data matrix. Set \mathcal{H} the submatrix containing the C medoids, \tilde{v}_c being the medoid for cluster c .

Step 1: Compute $w_1^{(s-1)}$ according to (3.9) using $\mathbf{U}^{(s-1)}$ and $\mathcal{H}^{(s-1)}$, where $s \geq 1$ denotes the iteration number.

Step 2: Update the medoid matrix $\mathcal{H}^{(s)}$: for $c = 1, \dots, C$ set

$$q = \arg \min_{1 \leq i' \leq N} \sum_{i=1}^N u_{ic}^m (\exp_{uv, tvv} d_{ii'}^2)$$

and $\tilde{v}_c^{(s)} = \tilde{v}_q$. Update $\mathbf{U}^{(s)}$ as in (3.8), using $\mathcal{H}^{(s)}$ and $w_1^{(s-1)}$ (for the unit i'' that is medoid for cluster l set $u_{i''l} = 1$; 0 otherwise).

Step 3: If $\mathcal{H}^{(s-1)} \equiv \mathcal{H}^{(s)}$, or iteration number $s = maxiter$ the algorithm has converged, otherwise go to step 1.

3.3.3. GARCH-based Fuzzy C-Medoids Clustering with Noise Cluster model (GARCH-NC-FCMdc)

In this section, we propose the Fuzzy C-Medoids clustering model with Noise Cluster (GARCH-NC-FCMdc). This model achieves its robustness with respect to outliers by introducing a noise cluster represented by a noise prototype, i.e. a noise medoid, which is always at the same distance from all units. By following [50], let there be $C - 1$ good clusters and let the C -th cluster be the noise cluster. Let \tilde{v}_C be the noise prototype (i.e. noise medoid). It is assumed that the distance measure $_{uv, tvv} d_{iC}$ of unit i from medoid \tilde{v}_C is equal to δ , $i = 1 \dots, N$.

Then the GARCH-NC-FCMdc model can be formalized as follows:

$$\begin{aligned} \min : & \mathcal{F}_{\text{GARCH-NC-FCMdc}}(\mathbf{U}, \mathbf{V}, w_1, w_2) \\ \equiv & \sum_{i=1}^N \sum_{c=1}^{C-1} u_{ic}^m \exp_{uv, tvv} d_{ic}^2 + \sum_{i=1}^N u_{iC}^m \delta^2 \\ = & \sum_{i=1}^N \sum_{c=1}^{C-1} u_{ic}^m [w_1^2 \exp_{uv} d_{ic}^2 + w_2^2 \exp_{tvv} d_{ic}^2] + \sum_{i=1}^N u_{iC}^m \delta^2 \\ = & \sum_{i=1}^N \sum_{c=1}^{C-1} u_{ic}^m [w_1^2 (uv_i - uv_c)^2 + w_2^2 (tvv_i - tvv_c)^2] + \sum_{i=1}^N u_{iC}^m \delta^2 \end{aligned} \tag{3.18}$$

where $u_{iC} = 1 - \sum_{c=1}^{C-1} u_{ic}$.

Notice that the model (3.18) represents an extension of model proposed by Davé [50] to GARCH data with medoid prototypes.

The distance from the noise cluster depends on the average distance among units $\delta^2 = \rho(N(C - 1))^{-1} \times \sum_{i=1}^N \sum_{c=1}^{C-1} u_{ic}^m \exp_{uv, tvv} d_{ic}^2$. In any case, the results do not seem very sensitive to the value of the multiplier ρ [50]. Due to the presence of δ , units that are close to good clusters are correctly classified in a good cluster while the noise units that are away from good clusters are classified in the noise cluster.

By considering the Lagrangian function (under the constraints in (3.1) for the membership degrees):

$$L_m(u_{ic}, \lambda, w_1, w_2) = \sum_{i=1}^N \sum_{c=1}^{C-1} u_{ic}^m [w_1^2 \exp_{uv} d_{ic}^2 + w_2^2 \exp_{tvv} d_{ic}^2] + \sum_{i=1}^N u_{iC}^m \delta^2 - \lambda \left(\sum_{c=1}^C u_{ic} - 1 \right)$$

and developing the derivatives, we obtain the following optimal solutions:

$$\begin{aligned}
 u_{ic} &= \frac{1}{\sum_{c'=1}^{C-1} \left[\frac{w_1^2 u v d_{ic}^2 + w_2^2 tv v d_{ic}^2}{w_1^2 u v d_{ic'}^2 + w_2^2 tv v d_{ic'}^2} \right]^{\frac{1}{m-1}} + \left[\frac{w_1^2 u v d_{ic}^2 + w_2^2 tv v d_{ic}^2}{\delta^2} \right]^{\frac{1}{m-1}}} \\
 &= \frac{1}{\sum_{c'=1}^{C-1} \left[\frac{w_1^2 (u v_i - u v_c)^2 + w_2^2 (t v v_i - t v v_c)^2}{w_1^2 (u v_i - u v_{c'})^2 + w_2^2 (t v v_i - t v v_{c'})^2} \right]^{\frac{1}{m-1}} + \left[\frac{w_1^2 (u v_i - u v_c)^2 + w_2^2 (t v v_i - t v v_c)^2}{\delta^2} \right]^{\frac{1}{m-1}}}
 \end{aligned}$$

for $c = 1, \dots, C - 1$, while the membership degrees for the noise cluster are:

$$\begin{aligned}
 u_{iC} &= \frac{\left[\frac{1}{\delta^2} \right]^{\frac{1}{m-1}}}{\sum_{c'=1}^{C-1} \left[\frac{1}{w_1^2 u v d_{ic'}^2 + w_2^2 tv v d_{ic'}^2} \right]^{\frac{1}{m-1}} + \left[\frac{1}{\delta^2} \right]^{\frac{1}{m-1}}} \\
 &= \frac{\left[\frac{1}{\delta^2} \right]^{\frac{1}{m-1}}}{\sum_{c'=1}^{C-1} \left[\frac{1}{w_1^2 (u v_i - u v_{c'})^2 + w_2^2 (t v v_i - t v v_{c'})^2} \right]^{\frac{1}{m-1}} + \left[\frac{1}{\delta^2} \right]^{\frac{1}{m-1}}}
 \end{aligned}$$

The value of w_1 and w_2 is obtained as in (3.9)–(3.10), where the inner summation ranges over the $C - 1$ good clusters.

Algorithm (GARCH-NC-FCMdC model).

Step 0: Fix the power of the membership degrees (m), the number of clusters (C), a value for ρ and the maximum number of iterations ($maxiter$). Generate randomly the membership degree matrix $\mathcal{U}^{(0)}$ subject to conditions in (3.1), and pick $C - 1$ medoids in the data matrix. Set \mathcal{H} the submatrix containing the $C - 1$ medoids, \tilde{v}_c being the medoid for cluster c .

Step 1: Compute $w_1^{(s-1)}$ according to (3.9) using $\mathcal{U}^{(s-1)}$ and $\mathcal{H}^{(s-1)}$ and $\delta^{2(t-1)}$, where $s \geq 1$ denotes the iteration number.

Step 2: Update the medoid matrix $\mathcal{H}^{(s)}$: for $c = 1, \dots, C - 1$ set

$$q = \arg \min_{1 \leq i' \leq N} \sum_{i=1}^N u_{ic}^m [w_1^2 u v d_{ii'}^2 + (1 - w_1)^2 t v v d_{ii'}^2]$$

and $\tilde{v}_c^{(s)} = \tilde{v}_q$. Update $\mathcal{U}^{(s)}$ as in (3.8), using $\mathcal{H}^{(s)}$ and $w_1^{(s-1)}$ (for the unit i'' that is medoid for cluster l set $u_{i''l} = 1$; 0 otherwise).

Step 3: If $\mathcal{H}^{(s-1)} \equiv \mathcal{H}^{(s)}$, or iteration number $s = maxiter$ the algorithm has converged, otherwise go to step 1.

3.3.4. GARCH-based Trimmed Fuzzy C-Medoids Clustering model (GARCH-Tr-FCMdC)

To overcome the outlier problem in the fuzzy clustering, it is also possible to adopt a trimming-based robust version of the GARCH-FCMdC model, the Trimmed Fuzzy C-Medoids Clustering model (GARCH-Tr-FCMdC). Following [48], we define an objective function for a robust version of the GARCH-FCMdC model by considering the Least Trimmed Squares approach and using (3.8)–(3.10).

This model achieves its robustness with respect to outliers by trimming away a certain fraction of the data units and requires the specification of the “trimming ratio”, τ , which is the fraction of the data units that has to be trimmed [48].

The objective function is obtained by substituting in the objective function (3.1) the expression of the optimal u_{ic} (3.8):

$$\begin{aligned}
 \min : & \mathcal{F}_{\text{GARCH-Tr-FCMdC}}(\text{opt}) \\
 \equiv & \sum_{i=1}^N \left(\sum_{c=1}^C [w_1^2 u v d_{ic}^2 + w_2^2 t v v d_{ic}^2]^{\frac{1}{1-m}} \right)^{1-m} = \sum_{i=1}^N \text{harm}_i
 \end{aligned}$$

where

$$\begin{aligned} harm_i &= \left(\sum_{c=1}^C [w_{1uv}^2 d_{ic}^2 + w_{2tvv}^2 d_{ic}^2]^{1-m} \right)^{1-m} \\ &= \left(\sum_{c=1}^C [w_1^2 (uv_i - uv_c)^2 + w_2^2 (tvv_i - tvv_c)^2]^{1-m} \right)^{1-m} \end{aligned} \tag{3.19}$$

is C^{-1} times the harmonic mean of $[w_{1uv}^2 d_{ic}^2 + w_{2tvv}^2 d_{ic}^2]^{1-m}$ when $m = 2$.

The objective function of the GARCH-Tr-FCM_{dC} model, when τN ($0 \leq \tau \leq 1$) data units are trimmed, corresponding to those with the highest values of $harm_i$ (the farthest from the medoids), is obtained by modifying the objective function in (3.1) setting to $N' = N - \tau N$ the limit of the range of the outer summation:

$$\begin{aligned} \min : \mathcal{F}_{\text{GARCH-FCM}_{dC}}(\mathbf{U}, \mathbf{V}, w_1, w_2) &\equiv \sum_{i=1}^{N'} \sum_{c=1}^C u_{ic}^m [w_{1uv}^2 d_{ic}^2 + w_{2tvv}^2 d_{ic}^2]^2 \\ &= \sum_{i=1}^{N'} \sum_{c=1}^C u_{ic}^m [w_1^2 d_{ic}^2 + w_2^2 d_{ic}^2] \\ &= \sum_{i=1}^{N'} \sum_{c=1}^C u_{ic}^m [w_1^2 (uv_i - uv_c)^2 + w_2^2 (tvv_i - tvv_c)^2] \end{aligned} \tag{3.20}$$

and is minimized with respect to the related parameters.

The values of w_1 and w_2 are obtained as in (3.9)–(3.10).

Notice that GARCH-Tr-FCM_{dC} model represents an extension of the Krishnapuram et al.'s model [48] to GARCH data with medoid prototypes.

Algorithm (GARCH-Tr-FCM_{dC} model).

Step 0: Fix the power of the membership degrees (m), the number of clusters (C), a value for τ and the maximum number of iterations ($maxiter$). Generate randomly the membership degree matrix $\mathbf{U}^{(0)}$ subject to conditions in (3.1), and pick C medoids in the data matrix. Set $\tilde{\mathcal{H}}$ the submatrix containing the C medoids, \tilde{v}_c being the medoid for cluster c .

Step 1: Compute $w_1^{(s-1)}$ according to (3.9) using $\mathbf{U}^{(s-1)}$ and $\tilde{\mathcal{H}}^{(s-1)}$, where $s \geq 1$ denotes the iteration number.

Step 2: Compute the harmonic distance $harm_i$ (3.19) for $i = 1, \dots, N$ and trim the τN ($0 \leq \tau \leq 1$) units corresponding to the highest values of $harm_i$.

Step 3: Compute the membership degrees matrix $\mathbf{U}_{N'}^{(s)}$, as in (3.8) for the $N' = N - \tau N$ data units. Update the medoid matrix $\tilde{\mathcal{H}}^{(s)}$: for $c = 1, \dots, C$ set

$$q = \arg \min_{1 \leq i' \leq N'} \sum_{i=1}^{N'} u_{ic}^m [w_1^2 d_{ii'}^2 + (1 - w_1)^2 d_{ii'}^2]$$

and $\tilde{v}_c^{(s)} = \tilde{v}_q$. Update $\mathbf{U}^{(s)}$ as in (3.8), using $\tilde{\mathcal{H}}^{(s)}$ and $w_1^{(s-1)}$ (for the unit i'' that is medoid for cluster l set $u_{i''l} = 1$; 0 otherwise).

Step 4: If $\tilde{\mathcal{H}}^{(s-1)} \equiv \tilde{\mathcal{H}}^{(s)}$, or iteration number $s = maxiter$ the algorithm has converged, otherwise go to step 1.

3.3.5. Some remarks

To choose the optimal number of clusters for the fuzzy classification, we considered the Xie–Beni index [59]. For the robust procedures considered it is also necessary to detect the optimal value of the parameters that control for the presence of anomalous time series, i.e. δ for GARCH-NC-FCM_{dC} and τ for GARCH-Tr-FCM_{dC}. The query for the optimal values of C and of the parameters is conducted over a grid of values. For C fixed, we choose the optimal value of the parameters for each method, thus obtaining a set of optimal values of the parameter for each value of C . Then, we select the number of clusters that minimize the value of the Xie–Beni index.

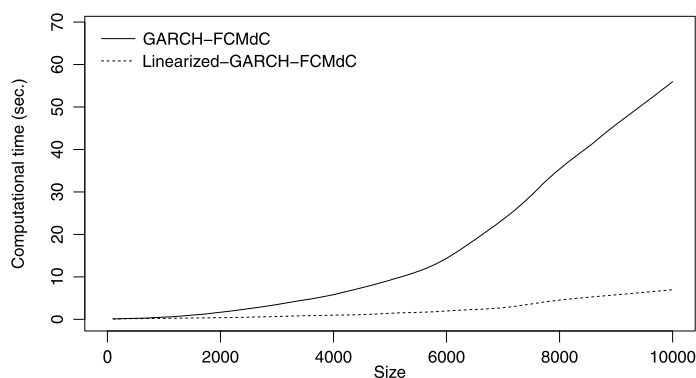


Fig. 3.3. Computation time for GARCH-FCMdc and Linearized-GARCH-FCMdc.

The optimal value of δ for the GARCH-NC-FCMdc method is obtained by adopting the strategy proposed by [16]. First, we executed the fuzzy clustering with noise cluster method with increasing values of δ . Then, by analyzing the distribution of the percentage of objects assigned to the noise cluster, we determined the optimal value of δ in correspondence of an abrupt change of the “slope” of the distribution.

The optimal values of τ for the trimmed-based proposed procedure is determined according to the Unsupervised Fuzzy Trimmed Algorithm presented by [51]. We gradually increase τ and compute the objective function for each value of τ . We expect that as τ increases, the outlier time series are eliminated and the value objective function decreases. The optimal value of τ is obtained in correspondence with the largest change, in absolute value, of the objective function.

Computational complexity and scalability are two important issues in objective function-based clustering. As the complexity is always linear with respect to the number of variables and of clusters, we consider the complexity with respect to the number of units N for each iteration of the algorithm. The complexity of all the proposed robust algorithms is $O(N^2)$. In fact these algorithms operate on each data unit twice (nested cycle) as the prototype is the medoid. To lower the complexity of the proposed robust algorithms the algorithms can be modified so that only a subset of data units is examined while updating the medoid for cluster c (Linearization). The subset is the set of data units that correspond to the top highest membership values in cluster c (under a proper threshold). The complexity of all the robust algorithms is reduced to $O(N)$ with the exception of the GARCH-Tr-FCMdc algorithm that is reduced to $O(N \log N)$ [48]. The complexity of this model is in fact dominated by the complexity of the ordering step that is $O(N \log N)$. The GARCH-FCMdc algorithm and the Linearized-GARCH-FCMdc algorithm have been run on simulated datasets considering $C = 2$ clusters, the value $m = 1.5$ of the power of the fuzziness degree, two variables, one hundred values of N (100 to 10000 step 100). The results are shown in Fig. 3.3.

We observe that the complexity of the GARCH-FCMdc grows as $O(N^2)$ whilst the complexity of the Linearized-GARCH-FCMdc as $O(N)$; the time required for $N = 10000$ is 100 (GARCH-FCMdc) and 10 (Linearized-GARCH-FCMdc) times the time required for $N = 1000$, respectively.

The value of the fuzziness exponent has been set to $m = 1.5$ according to suggestions by [60].

4. Simulation study

In this section, we illustrate the results of a simulation study conducted to evaluate the performance and the clustering accuracy of the proposed robust clustering models¹ GARCH-FCMdc, GARCH-E-FCMdc, GARCH-NC-FCMdc, GARCH-Tr-FCMdc. Datasets according to two different scenarios have been simulated. In each dataset $N = 100$ time series of length $T = 1000$ are generated. For each scenario, two separated clusters of time series generated, respectively, from two different GARCH processes have been simulated. In the first Scenario the first generating process is GARCH(0.40, 0.30, 0.20) resulting in $uv = 0.67$ and $tvv = 0.31$, the second generating process is GARCH(0.40, 0.60, 0.20) resulting in $uv = 1.00$ and $tvv = 0.61$; in the second Scenario the first generating process

¹ Source code of the algorithms is available from the authors upon request.

Table 4.1
Data and outlier generation in the simulation experiment.

		$GARCH(\gamma, \alpha, \beta)$	uv	tvv
Scenario 1	cluster 1 ($i = 1, \dots, N/2$)	$GARCH(0.40, 0.30, 0.20)$	0.67	0.31
	cluster 2 ($i = N/2 + 1, \dots, N$)	$GARCH(0.40, 0.60, 0.20)$	1.00	0.61
	outliers	$\gamma N(0.02, 0.005)$ $\alpha N(0.85, 0.005)$ $\beta U(0.10, 1 - \alpha)$	0.10	0.86
Scenario 2	cluster 1 ($i = 1, \dots, N/2$)	$GARCH(0.10, 0.35, 0.10)$	0.16	0.35
	cluster 2 ($i = N/2 + 1, \dots, N$)	$GARCH(0.30, 0.10, 0.10)$	0.37	0.10
	outliers	$\gamma N(0.14, 0.001)$ $\alpha N(0.85, 0.001)$ $\beta U(0.05, 1 - \alpha)$	0.69	0.85

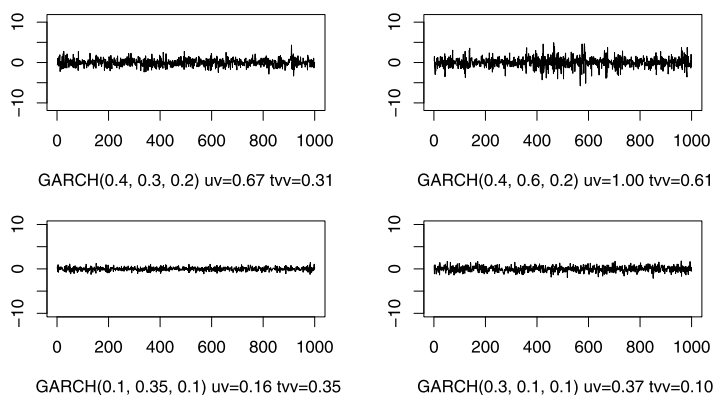


Fig. 4.1. Simulation Scenarios (Scenario 1 top, Scenario 2 bottom).

is $GARCH(0.10, 0.35, 0.10)$ resulting in $uv = 0.16$ and $tvv = 0.35$, the second generating process is $GARCH(0.30, 0.10, 0.10)$ resulting in $uv = 0.37$ and $tvv = 0.10$.

In order to evaluate the robustness of the proposed models in the presence of outliers, $0.1N$ and $0.2N$ outliers were added to the 100 time series. In the first Scenario the value of the parameter γ of the outliers is generated from a Gaussian distribution $N(0.02, 0.005)$, the value of the parameter α from a Gaussian distribution $N(0.85, 0.005)$, the value of the parameter β from a Uniform distribution $U(0.10, 1 - \alpha)$ (in order to guarantee $\alpha + \beta < 1$), resulting in $uv = 0.10$ and $tvv = 0.85$. The resulting either 100 or 120 time series of the first Scenario are characterized by the variability (measured by the sample variance of the simulated values) of the unconditional volatility greater than the variability of the time-varying volatility. In the second Scenario the value of the parameter γ of the outliers is generated from a Gaussian distribution $N(0.14, 0.001)$, the value of the parameter α from a Gaussian distribution $N(0.85, 0.001)$, the value of the parameter β from a Uniform distribution $U(0.05, 1 - \alpha)$ (in order to guarantee $\alpha + \beta < 1$), resulting in $uv = 0.69$ and $tvv = 0.85$. The resulting either 100 or 120 time series of the second scenario are characterized by the variability of the time-varying volatility greater than the variability of the time unconditional volatility.

Data generation is summarized in Table 4.1, while the simulated scenarios are depicted in Fig. 4.1.

The performances of each model have been evaluated with respect to misclassification and to prototype detection.

To evaluate the robustness of a clustering model with respect to misclassification in the presence of outliers, a criterion for comparing clustering partitions is needed. A review and analysis of these criteria in the non-fuzzy framework can be found in [61]. The Fuzzy Rand index (*fri*) [62,63] has been used to compare the hard partition in two clusters with the fuzzy partition obtained as output of the robust models. The Fuzzy Rand index is a fuzzy extension of the Rand index based on agreements and disagreements in the two partitions. The index is averaged over 100 simulation runs.

The Xie–Beni index of cluster-validity (*xb*) is also presented.

Table 4.2
Performance parameters of the robust models.

	%outlier	Scenario 1				Scenario 2			
		w_1	xb	fri	md	w_1	xb	fri	md
GARCH-FCMdc	0%	0.35	0.06	0.96	1.05	0.60	0.03	0.98	1.04
	10%	0.40	0.25	0.94	1.12	0.52	0.44	0.95	1.17
	20%	0.40	0.41	0.93	1.25	0.51	0.69	0.92	1.22
GARCH-E-FCMdc	0%	0.36	0.06	0.96	1.05	0.62	0.03	0.99	1.04
	10%	0.37	0.19	0.95	1.05	0.62	0.25	0.98	1.09
	20%	0.39	0.32	0.94	1.09	0.56	0.51	0.97	1.14
GARCH-NC-FCMdc	0%	0.35	0.06	0.96	1.05	0.56	0.03	0.99	1.04
	10%	0.33	0.07	0.95	1.05	0.56	0.03	0.98	1.05
	20%	0.34	0.08	0.95	1.05	0.56	0.08	0.98	1.05
GARCH-Tr-FCMdc	0%	0.32	0.06	0.96	1.05	0.68	0.03	0.99	1.04
	10%	0.35	0.06	0.96	1.05	0.61	0.04	0.99	1.05
	20%	0.32	0.06	0.96	1.05	0.63	0.12	0.98	1.05

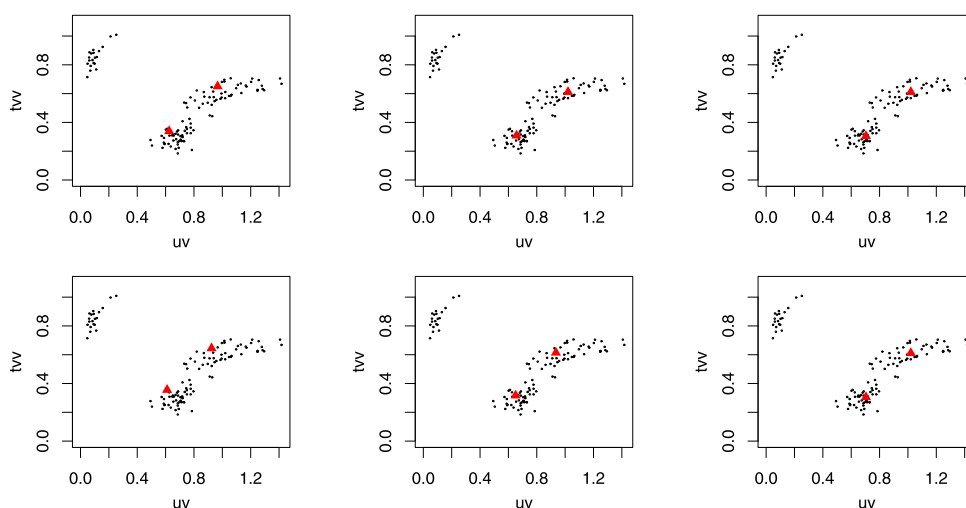


Fig. 4.2. Simulated data (dots) and medoids (triangles)—first Scenario. Top 10% outliers, bottom 20% outlier; left to right GARCH-FCMdc, GARCH-E-FCMdc, GARCH-NC-FCMdc/GARCH-Tr-FCMdc (same medoids). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

To evaluate the robustness of a clustering model with respect to prototype detection in the presence of outliers, an index md has been introduced which compares the medoids obtained in the presence of outliers with the expectation of the generating random variable. The index for the case of n outliers is the following:

$$md[(0\tilde{v}_1, 0\tilde{v}_2), (n\tilde{v}_1, n\tilde{v}_2)] = \frac{d(0\tilde{v}_1, n\tilde{v}_1) + d(0\tilde{v}_2, n\tilde{v}_2) + d(0\tilde{v}_1, n\tilde{v}_2) + d(0\tilde{v}_2, n\tilde{v}_1)}{2d(0\tilde{v}_1, 0\tilde{v}_2)}$$

where $n\tilde{v}_c$ denotes the medoid of cluster c in the case of n outliers and $0\tilde{v}_c$ the medoid obtained considering the expected values of the generative random variables and $d(., .)$ the Euclidean distance. It holds $md \geq 1$ due to the triangular inequality.

The index is averaged over the 100 simulation runs.

The results are presented in Table 4.2. For each scenario, the best results are in bold font. In Figs. 4.2–4.3 the optimal medoids for the two scenarios are represented by red triangles.

Table 4.2 shows the high performances of the proposed models either in terms of classification correctness (fri close to 1) or in terms of robustness in the presence of an increasing number of outliers (md close to 1). The model GARCH-FCMdc is less robust in the presence of outliers either considering misclassification or prototype detection

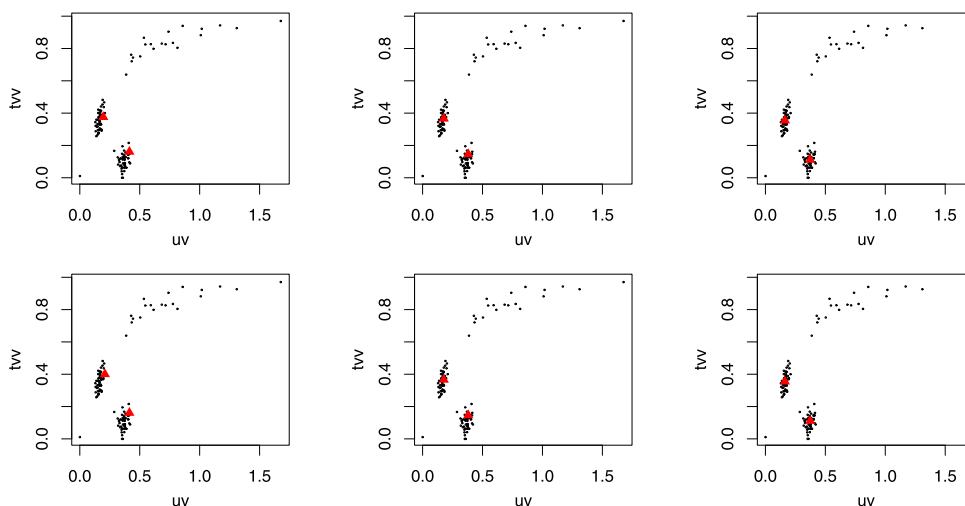


Fig. 4.3. Simulated data (dots) and medoids (triangles)—second Scenario. Top 10% outliers, bottom 20% outlier; left to right GARCH-FCMdC, GARCH-E-FCMdC, GARCH-NC-FCMdC/GARCH-Tr-FCMdC (same medoids). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

(higher value of md than the other models, graphically represented in Figs. 4.2 and 4.3). In scenario 1 for all the models the value of w_1 is smaller than 0.5 due to the greater variability of the unconditional volatility uv ; in scenario 2 for all the models the value of w_1 is greater than 0.5 due to the greater variability of the time varying volatility tvv (see section 2).

5. Applications

5.1. International stock-market volatility daily index returns

The first application is based on the volatility daily index returns of the major international stock exchanges. The data includes 30 time series with unequal length, since the period of observation of time series starts from January 2000 until December 2010, but some of time series data were available only for a shorter period. As observed above (see Sections 1 and 2.2), the classification on time series that have unequal length is possible because we use a parametric approach based on representation the time series in GARCH process. All data are available on the finance section of Yahoo website.² The goal of the application is to show the usefulness of the fuzzy clustering approach in identifying groups of countries whose daily index returns share a similar volatility pattern. The time series of daily index returns are shown in Fig. 5.1.

In order to choose the best GARCH(p, q) models for the time series of daily index returns we consider the Box and Jenkins modeling procedure. The final results of this procedure are shown in Table 5.1. The best model selected for all time series is the GARCH(1, 1). All estimated coefficients are significant. In Table 5.1 we show the models' coefficients, their standard error (in brackets) and some diagnostic tests with the related p -value. Q and Q^2 are the values of the Ljung–Box test up to twenty lags for serial correlation in the residuals and the LM is the Lagrange multiplier test for ARCH (AutoRegressive Conditional Heteroskedasticity) effects in the residuals. We then apply the clustering algorithms which uses the truncated $AR(\infty)$ representation described in Section 2. First we represent the GARCH(1,1) process in the ARMA(1,1) process; then the time series transformed in the ARMA process are fitted with truncated $AR(\infty)$ models and those time series fitted with AR models retain their coefficients.

The indices are represented in Fig. 5.2 according to the values of uv , tvv .

We apply our four robust clustering models—i.e. the GARCH-FCMdC model, the GARCH-E-FCMdC model, the GARCH-NC-FCMdC model, the GARCH-Tr-FCMdC model compared with the crisp version of each model. Data are standardized dividing each variable by its maximum value.

² <http://finance.yahoo.com>.

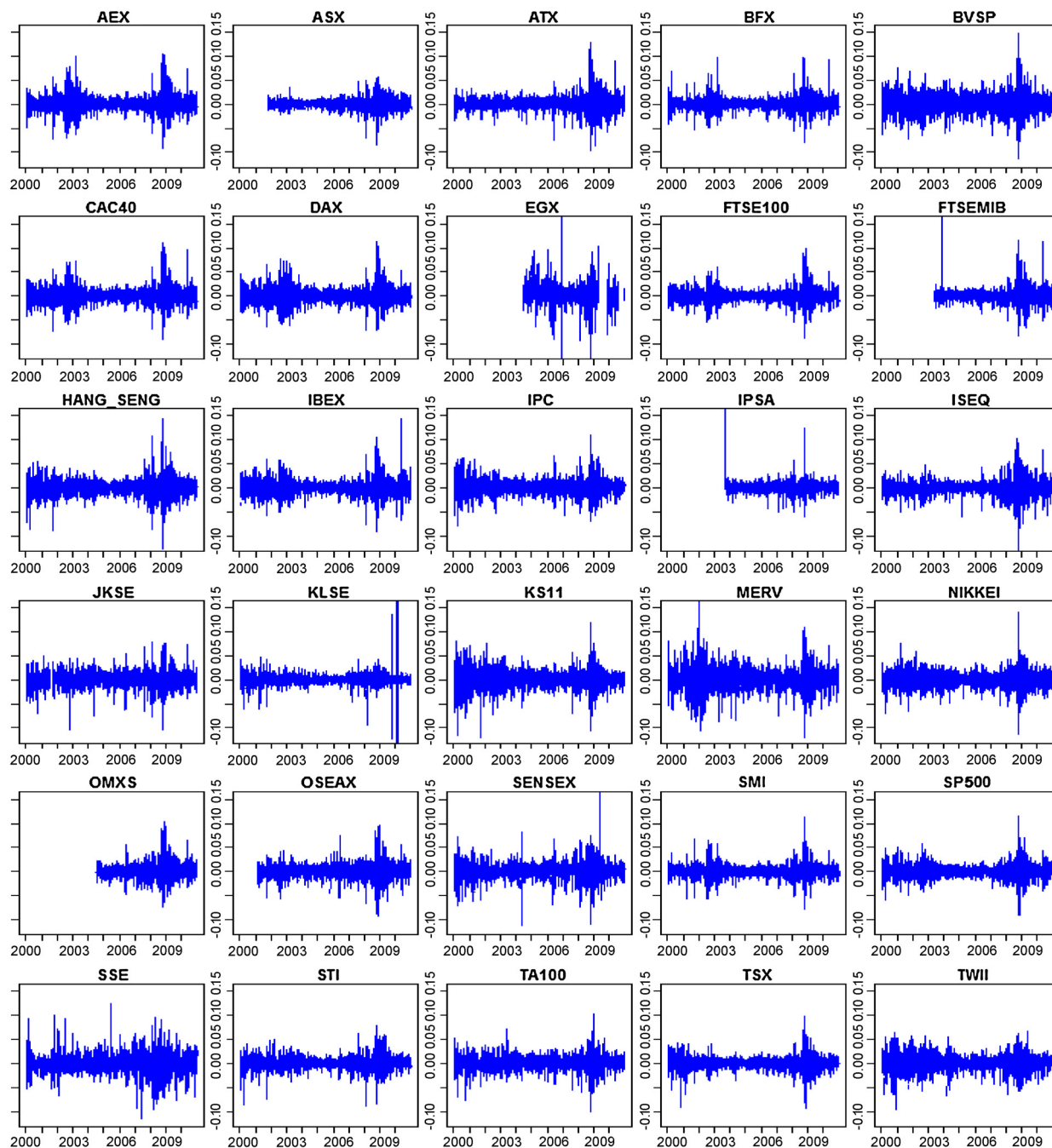


Fig. 5.1. Daily index returns of 30 international stock exchanges.

The membership degrees are presented in [Table 5.2](#).

The best value of cluster validity for all the models is obtained with a partition in 2 clusters, with lowest values shown by the three robust models. The partitions based on the robust fuzzy clustering for the models GARCH-E-FCM_dC, GARCH-NC-FCM_dC, GARCH-Tr-FCM_dC are the following (notice that units 8 and 14 are present only in the partition obtained with GARCH-E-FCM_dC):

$$(1, 3, 4, 9, 10, 12, 14, 15, 16, 17, 22, 23, 24, 27)(2, 5, 6, 7, 8, 9, 11, 13, 18, 19, 20, 21, 25, 26, 28, 29, 30)$$

Table 5.1
Estimated coefficients of GARCH(1,1) processes for the daily index returns.

Indices	ARCH $\hat{\alpha}$	GARCH $\hat{\beta}$	$Q(20)$	p -value	$Q^2(20)$	p -value	LM	p -value
1 AEX (NL)	0.1081 (0.0141)	0.8858 (0.0129)	22.4556	0.3163	22.6634	0.3056	18.0606	0.1159
2 S&P/ASX 200 (AUS)	0.0869 (0.0164)	0.9096 (0.0164)	21.1871	0.3861	18.7262	0.5396	13.7442	0.3173
3 ATX (AT)	0.1284 (0.0211)	0.8578 (0.0249)	28.355	0.1188	16.6635	0.6747	11.1938	0.5123
4 BFX (BE)	0.1444 (0.0216)	0.8466 (0.0208)	22.4823	0.3149	16.0483	0.7136	8.9599	0.7063
5 BVSP (BR)	0.0689 (0.0176)	0.9088 (0.0159)	25.0868	0.1981	17.4879	0.621	10.4213	0.579
6 CAC40 (F)	0.0939 (0.0122)	0.9000 (0.0146)	26.7516	0.1423	12.7494	0.8878	9.9122	0.6236
7 DAX (D)	0.0927 (0.0123)	0.8990 (0.0120)	22.472	0.3154	26.3797	0.1536	17.3854	0.1244
8 EGX30 (ET)	0.0260 (0.0240)	0.9655 (0.0378)	25.0183	0.1528	20.174	0.447	11.2824	0.5048
9 FTSE100 (GB)	0.1053 (0.0135)	0.8898 (0.0131)	22.442	0.317	29.8209	0.1105	17.9871	0.1201
10 FTSEMIB (I)	0.1079 (0.0209)	0.8845 (0.0195)	14.5796	0.7999	25.5811	0.1801	16.9896	0.1308
11 HANG SENG (HK)	0.0664 (0.0095)	0.9286 (0.0091)	23.5587	0.2621	24.3376	0.2279	17.9232	0.1259
12 IBEX 35 (E)	0.1009 (0.0190)	0.8845 (0.0173)	17.2738	0.6351	28.1289	0.1277	17.2675	0.1301
13 IPC (MEX)	0.0789 (0.0157)	0.9089 (0.0187)	30.3476	0.0998	9.0994	0.9817	7.2364	0.8415
14 IPSA (RCH)	0.1763 (0.0304)	0.7962 (0.0291)	31.8239	0.0812	18.7324	0.5392	12.9544	0.3723
15 ISEQ (IRL)	0.1002 (0.0176)	0.8908 (0.0186)	28.6551	0.1106	14.695	0.7935	9.2692	0.6797
16 JKSE (RI)	0.1393 (0.0315)	0.8007 (0.0470)	32.8754	0.0779	18.4597	0.5571	9.0147	0.7016
17 KLSE (MAL)	0.1298 (0.0357)	0.8677 (0.0361)	30.1713	0.0954	26.8163	0.1514	2.2909	0.9988
18 KS11 (ROK)	0.0806 (0.0157)	0.9151 (0.0146)	27.8968	0.1118	15.2465	0.7621	10.1004	0.6071
19 MERV (RA)	0.0969 (0.0145)	0.8764 (0.0174)	29.7804	0.1034	14.2712	0.8164	10.6137	0.5622
20 NIKKEI (J)	0.0948 (0.0132)	0.8947 (0.0126)	18.9275	0.5265	27.5483	0.1205	17.4932	0.1319
21 OMXS30 (S)	0.1240 (0.0204)	0.8487 (0.0243)	16.1155	0.7094	19.9271	0.4624	11.5688	0.4808
22 OSEAX (N)	0.0675 (0.0207)	0.9244 (0.0237)	28.8912	0.0956	23.7344	0.2541	13.9576	0.3042
23 SENSEX (IND)	0.1403 (0.0313)	0.8432 (0.0250)	32.2054	0.071	14.5702	0.8004	12.2251	0.4277
24 SMI (CH)	0.1176 (0.0145)	0.8684 (0.0143)	22.5214	0.3128	29.6336	0.1098	18.9469	0.1019
25 S&P 500 (USA)	0.0801 (0.0108)	0.9124 (0.0108)	13.5258	0.8537	23.1422	0.2818	18.4981	0.1093
26 SSE (CN)	0.0718 (0.0175)	0.9210 (0.0181)	28.1062	0.1027	13.0897	0.8735	5.9461	0.9187
27 STI (SGP)	0.1032 (0.0198)	0.8953 (0.0171)	28.26	0.1033	20.9776	0.3984	13.437	0.3381
28 TA100 (IL)	0.0921 (0.0183)	0.8791 (0.0223)	29.7282	0.1005	16.6511	0.6754	12.8321	0.3813
29 S&P/TSX (CND)	0.0658 (0.0133)	0.9296 (0.0142)	12.5148	0.8972	14.0501	0.8279	11.6758	0.4721
30 TSEC (TW)	0.0672 (0.0129)	0.9281 (0.0130)	12.5148	0.8972	29.9133	0.1027	17.2473	0.1283

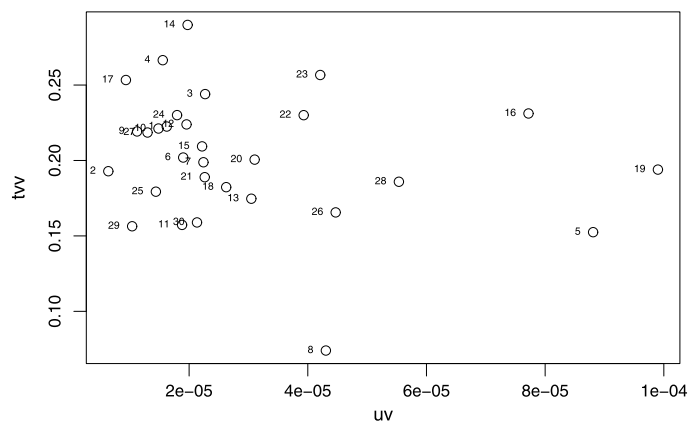


Fig. 5.2. Unconditional volatility and time-varying volatility of the indices.

The partition based on the GARCH-FCMdC model is:

(1, 2, 3, 4, 6, 7, 9, 10, 12, 14, 15, 16, 17, 20, 21, 22, 23, 24, 27)(5, 8, 11, 13, 18, 19, 25, 26, 28, 29, 30)

Some comments follow.

Table 5.2
Membership degrees.

		GARCH-FCMdc			GARCH-NC-FCMdc				GARCH-E-FCMdc			GARCH-Tr-FCMdc		
		membership degrees	crisp		membership degrees	crisp		membership degrees	crisp		membership degrees	crisp		
1	AEX	1.000	0.000	1	0.998	0.001	0.001	1	0.999	0.001	1	0.999	0.001	1
2	S&P/ASX 200	0.850	0.150	1	0.008	0.990	0.003	1	0.070	0.930	1	0.059	0.941	2
3	ATX	0.996	0.004	1	0.990	0.003	0.007	1	0.997	0.003	1	0.998	0.002	1
4	BFX	0.975	0.025	1	0.735	0.026	0.239	1	0.943	0.057	1	0.976	0.024	1
5	BVSP	0.057	0.943	2	0.021	0.832	0.147	3	0.020	0.980	2	0.009	0.991	2
6	CAC40	0.958	0.042	1	0.189	0.789	0.022	1	0.474	0.526	1	0.322	0.678	1
7	DAX	0.892	0.108	1	0.072	0.915	0.013	1	0.276	0.724	1	0.265	0.735	2
8	EGX30	0.216	0.784	2	0.009	0.038	0.954	3	0.384	0.616	2	outlier		
9	FTSE100	0.999	0.001	1	0.990	0.007	0.003	1	0.995	0.005	1	0.997	0.003	1
10	FTSEMIB	1.000	0.000	1	0.996	0.003	0.001	1	0.998	0.002	1	0.999	0.001	1
11	HANG SENG	0.053	0.947	2	0.013	0.920	0.067	2	0.007	0.993	2	0.003	0.997	2
12	IBEX 35	1.000	0.000	1	0.999	0.000	0.000	1	1.000	0.000	1	1.000	0.000	1
13	IPC	0.021	0.979	2	0.000	0.999	0.001	2	0.000	1.000	2	0.000	1.000	2
14	IPSA	0.934	0.066	1	0.289	0.028	0.683	3	0.818	0.182	1	outlier		
15	ISEQ	0.993	0.007	1	0.725	0.250	0.025	1	0.876	0.124	1	0.887	0.113	1
16	JKSE	0.584	0.416	1	0.998	0.001	0.001	3	0.997	0.003	1	0.999	0.001	1
17	KLSE	0.989	0.011	1	0.938	0.011	0.051	1	0.984	0.016	1	0.992	0.008	1
18	KS11	0.159	0.841	2	0.000	1.000	0.000	1	0.001	0.999	2	0.000	1.000	2
19	MERV	0.203	0.797	2	0.030	0.958	0.012	3	0.127	0.873	2	0.101	0.899	2
20	NIKKEI	0.805	0.195	1	0.126	0.857	0.018	1	0.377	0.623	1	0.271	0.729	2
21	OMXS30	0.510	0.490	2	0.001	0.999	0.000	1	0.018	0.982	2	0.015	0.985	2
22	OSEAX	0.979	0.021	1	1.000	0.000	0.000	1	1.000	0.000	1	1.000	0.000	1
23	SENSEX	0.954	0.046	1	0.898	0.015	0.087	1	0.975	0.025	1	0.989	0.011	1
24	SMI	1.000	0.000	1	1.000	0.000	0.000	1	1.000	0.000	1	1.000	0.000	1
25	S&P 500	0.366	0.634	2	0.000	1.000	0.000	1	0.000	1.000	2	0.000	1.000	2
26	SSE	0.000	1.000	2	0.005	0.980	0.015	2	0.001	0.999	2	0.000	1.000	2
27	STI	1.000	0.000	1	0.987	0.010	0.003	1	0.994	0.006	1	0.995	0.005	1
28	TA100	0.029	0.971	2	0.000	1.000	0.000	2	0.007	0.993	2	0.005	0.995	2
29	S&P/TSX	0.131	0.869	2	0.014	0.909	0.077	2	0.008	0.992	2	0.004	0.996	2
30	TSEC	0.039	0.961	2	0.011	0.937	0.052	2	0.005	0.995	2	0.002	0.998	2

The analysis of the crisp classification—i.e., the classification obtained when each time series is assigned to a cluster according to the maximal membership degree—for all the models shows worse performances with respect to the cluster validity indices.

The obtained value of w_1 is 0.29 for GARCH-FCMdc, 0.10 for GARCH-E-FCMdc, GARCH-NC-FCMdc, GARCH-Tr-FCMdc showing higher variability for the unconditional volatility component.

Due to the values of w_1 , the clusters are identified mainly on the basis of the time-varying volatility. In particular, cluster 1 includes time series with low values of uv and high values of tvv .

In the GARCH-NC-FCMdc model indices 8 (EGX30) and 14 (IPSA) have the highest membership to the noise cluster. In the GARCH-Tr-FCMdc model the indices trimmed are 8 and 14. Then, GARCH-NC-FCMdc and GARCH-Tr-FCMdc identify the Egyptian index EGX30 and the Chilean index IPSA as outliers. The outlier EGX30 shows a strong membership to the noise cluster, while the membership degree of the outlier index IPSA is lighter. Index EGX30 is an outlier as it has very low value for the time-varying volatility, while IPSA has a high value for time-varying volatility (Fig. 5.2).

Indexes 8 and 14 are assigned to clusters 2 and 1 in the models GARCH-FCMdc and GARCH-E-FCMdc, but in the model GARCH-E-FCMdc with a lower membership, thus not affecting the value of the cluster validity index.

In the robust model GARCH-E-FCMdc the medoids are (24, 13) in GARCH-NC-FCMdc, GARCH-Tr-FCMdc (24, 18); in GARCH-FCMdc (12, 26). As for GARCH-FCMdc, the medoids are influenced by the presence of index 8; in fact index 26 has a lower value of tvv with respect to indices 13 or 18. As a consequence indices 2, 6, 7, 20 and 21 move to the other cluster for the reason that they are too far from index 26. Moreover, beside index 8, also indices 16 and 19 have low membership to the related cluster as the higher value of w_1 leads to consider also the high value of uv as anomalous.

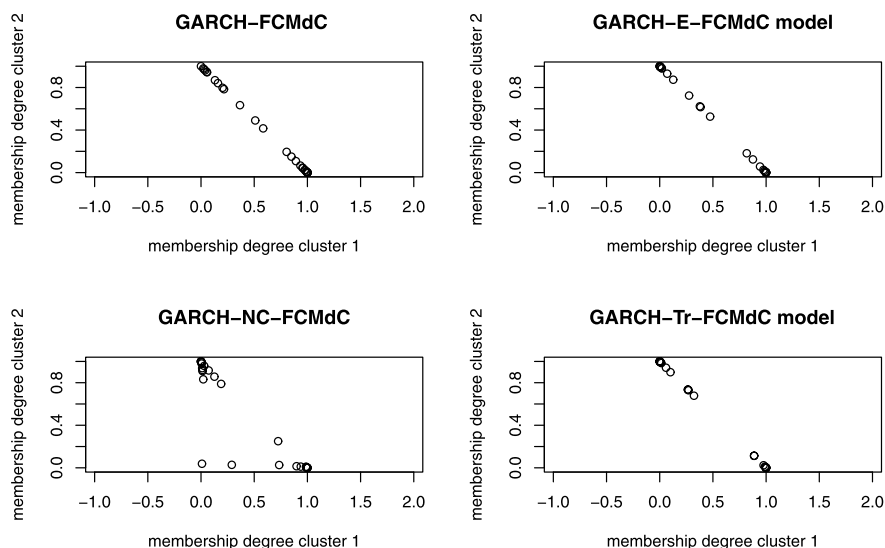


Fig. 5.3. Membership degrees.

With regards to the crisp classification, the crisp version of GARCH-NC-FCMdC assigns to cluster 3 also indices 5, 16 and 19 beside 8 and 14, due to the higher value of w_1 (higher weight to uv).

Separation and compactness of the obtained partitions are illustrated in Fig. 5.3. The analysis of the membership degrees shows that some indices present uncertain classification in the model GARCH-FCMdC, specifically indices 8, 14, 16, 21 and 25, and in the model GARCH-E-FCMdC, specifically indices 6, 7, 8 and 14. Indexes 8 and 14 are in the noise cluster in the model GARCH-NC-FCMdC and are trimmed in the model GARCH-Tr-FCMdC; then the clusters obtained by the highest membership degree are clearly identified.

5.2. Volatility daily stocks returns in FTSEMIB index

The second application is based on daily returns of stocks that form the FTSEMIB index. It is the most significant index in the Italian Stock Exchange. FTSEMIB index contains stocks of 40 major Italian and foreign companies listed on the Italian Stock Exchange. The period of observation of time series starts from January 2000 until December 2010, even if some time series are observed for a shorter period (see Section 5.1). As for the previous application, data are available on the finance section of the Yahoo website.³ The time series of daily stocks returns are shown in Fig. 5.4.

Following the same procedure as in section 5.1, we obtain the results shown in Table 5.3. The best model selected for all index is the GARCH(1,1). All estimated coefficients are significant.

The uv and tvv values of the time series are represented in Fig. 5.5.

We proceed to apply the clustering algorithms in a similar way as in section 5.1. The membership degrees are presented in Table 5.4.

The best value of cluster validity for all the models is obtained with a partition in 2 clusters. The partitions based on the robust fuzzy clustering for the models GARCH-E-FCMdC, GARCH-NC-FCMdC, GARCH-Tr-FCMdC are the following (units 13 appear only in the partition obtained with GARCH-E-FCMdC):

(1, 2, 3, 4, 5, 7, 8, 9, 10, 14, 15, 16, 18, 19, 21, 23, 25, 26, 27, 28, 29, 30, 33, 37)

(6, 11, 12, 13, 17, 20, 22, 24, 31, 32, 34, 35, 36, 38, 39, 40)

The partition based on the GARCH-FCMdC model is:

³ <http://finance.yahoo.com>.

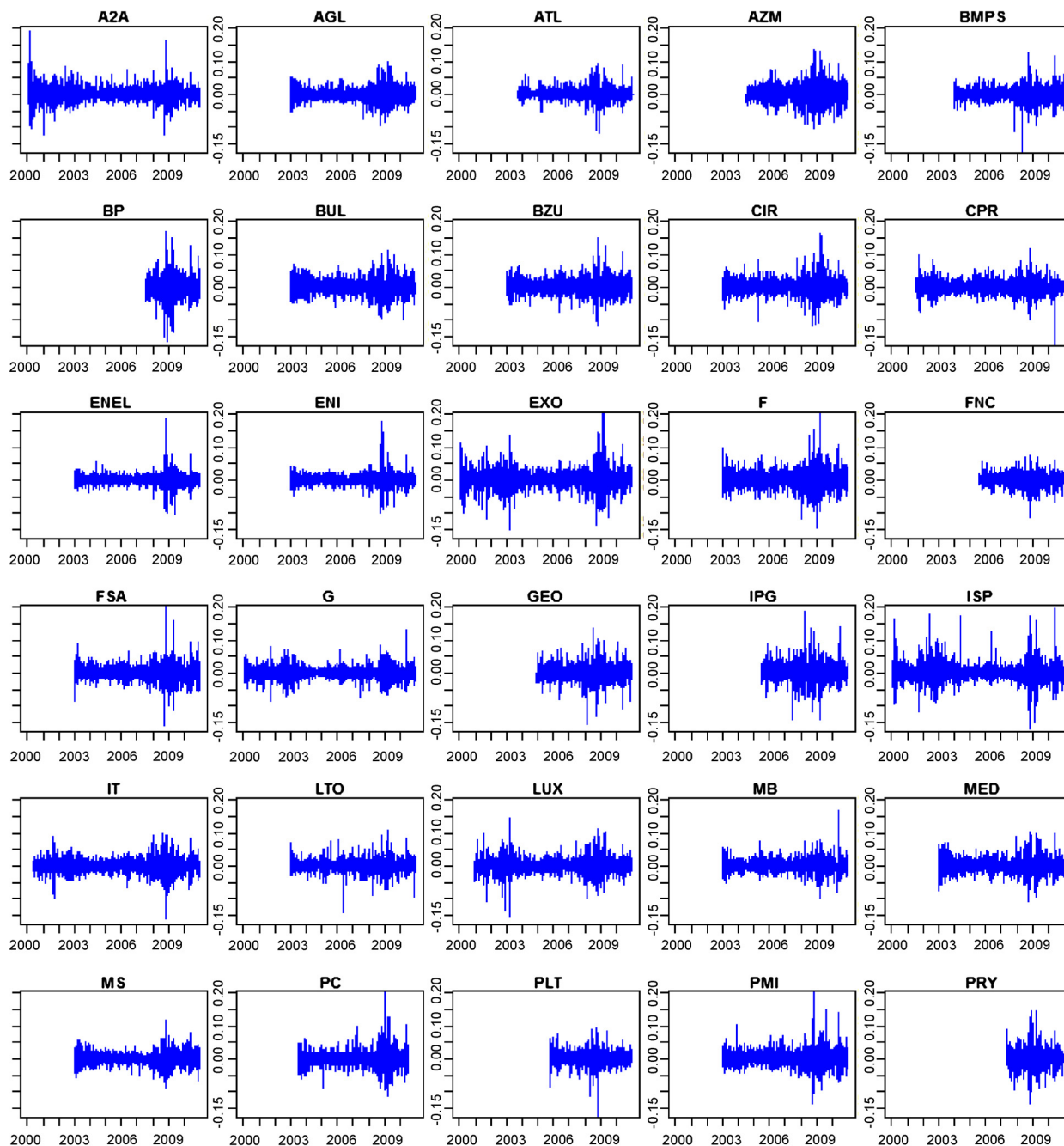


Fig. 5.4. Daily stock returns in FTSEMIB index.

(1, 2, 3, 4, 5, 7, 8, 9, 10, 14, 15, 16, 18, 19, 21, 23, 25, 26, 27, 28, 29, 30, 33, 36, 37, 38)

(6, 11, 12, 13, 17, 20, 22, 24, 31, 32, 34, 35, 39, 40)

Some comments follow.

The value of w_1 is 0.19 for GARCH-FCM_dC, 0.10 for GARCH-E-FCM_dC, GARCH-NC-FCM_dC, GARCH-Tr-FCM_dC showing higher variability for the unconditional volatility component. As in the previous application, the clusters are identified mainly on the basis of the time-varying volatility. In particular, cluster 1 includes stocks with low values of tvv .

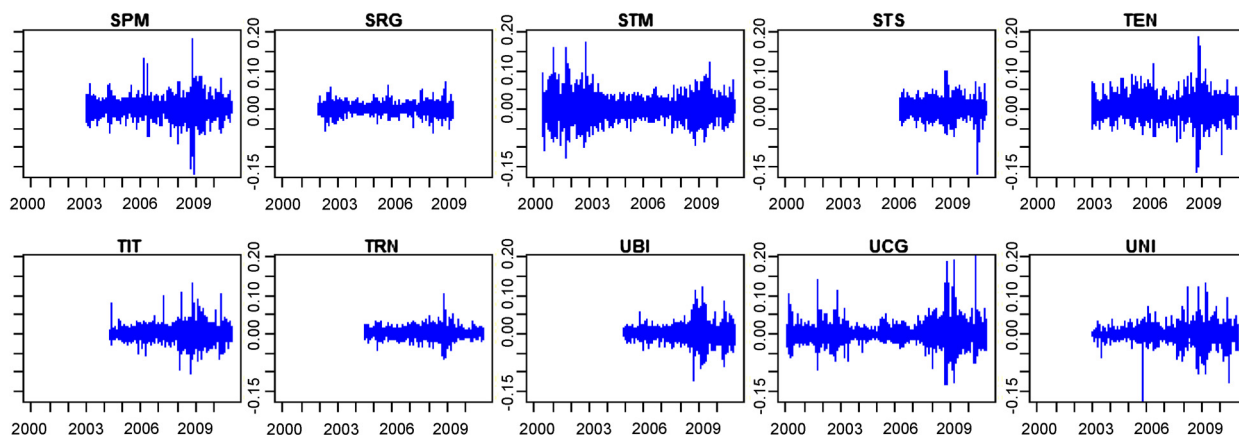


Fig. 5.4. (continued)

Table 5.3
Estimated coefficients of GARCH(1,1) processes for the volatilities daily returns of stocks that make up the FTSEMIB index.

Indices		ARCH $\hat{\alpha}$	GARCH $\hat{\beta}$	$Q(20)$	p -value	$Q^2(20)$	p -value	LM	p -value
1	A2A	0.0732 (0.0166)	0.9129 (0.0194)	15.4036	0.7528	11.4715	0.933	7.5247	0.821
2	AGL	0.0554 (0.0175)	0.9384 (0.0188)	16.3899	0.6921	18.4857	0.5554	11.8067	0.4613
3	ATL	0.0767 (0.0456)	0.9107 (0.0537)	27.3807	0.1139	19.1045	0.515	13.0634	0.3644
4	AZM	0.0646 (0.0175)	0.9287 (0.0208)	24.2525	0.2315	14.7687	0.6323	11.8495	0.4578
5	BMPS	0.0633 (0.0494)	0.9330 (0.0581)	19.1115	0.5145	2.5685	0.9999	1.2496	0.9999
6	BP	0.0946 (0.0392)	0.8894 (0.0523)	14.5424	0.8019	21.3779	0.3751	16.1554	0.1842
7	BUL	0.0288 (0.0076)	0.9670 (0.0091)	23.2198	0.2781	17.8185	0.5993	9.4886	0.6607
8	BZU	0.0739 (0.0240)	0.9109 (0.0378)	25.0183	0.1528	20.174	0.447	11.2824	0.5048
9	CIR	0.0773 (0.0264)	0.8974 (0.0360)	15.0904	0.7712	14.4385	0.8075	9.5823	0.6525
10	CPR	0.0609 (0.0266)	0.9250 (0.0358)	24.6161	0.2055	15.9732	0.7182	5.9035	0.9208
11	ENEL	0.1032 (0.0264)	0.8766 (0.0309)	26.5389	0.1487	15.7944	0.7292	11.9232	0.4518
12	ENI	0.1006 (0.0177)	0.8790 (0.0199)	9.9481	0.9691	12.4756	0.8987	7.6896	0.8088
13	EXO	0.3724 (0.0911)	0.5360 (0.0812)	21.3099	0.379	13.1397	0.8713	5.4366	0.9417
14	F	0.0710 (0.0178)	0.9162 (0.0274)	26.2113	0.1511	11.2366	0.9398	3.8176	0.9865
15	FNC	0.0770 (0.0162)	0.8947 (0.0212)	11.6451	0.9277	15.6669	0.737	12.8545	0.3797
16	FSA	0.0552 (0.0129)	0.9415 (0.1343)	21.9669	0.3423	22.0685	0.3368	18.3615	0.1305
17	G	0.1130 (0.0207)	0.8788 (0.0205)	21.7589	0.3537	20.5448	0.4243	16.183	0.183
18	GEO	0.0581 (0.0239)	0.9235 (0.0343)	26.9204	0.1093	17.2527	0.6365	7.1647	0.8465
19	IPG	0.0764 (0.0367)	0.8854 (0.0636)	28.241	0.0987	5.4978	0.9994	3.3998	0.992
20	ISP	0.1048 (0.0175)	0.8966 (0.0140)	12.8623	0.8832	21.0717	0.3929	14.3241	0.2804
21	IT	0.0586 (0.0139)	0.9339 (0.0164)	22.1081	0.3346	24.4691	0.2576	19.3877	0.1133
22	LTO	0.1132 (0.0379)	0.8189 (0.0451)	18.2315	0.5721	4.9232	0.9997	2.7138	0.9972
23	LUX	0.0675 (0.0207)	0.9244 (0.0237)	28.8912	0.0956	23.7344	0.2541	13.9576	0.3042
24	MB	0.1067 (0.0313)	0.8697 (0.0376)	16.9691	0.6549	15.7673	0.7309	10.7501	0.5504
25	MED	0.0543 (0.0158)	0.9301 (0.0211)	27.1193	0.1098	14.2428	0.8179	3.8923	0.9853
26	MS	0.0531 (0.0159)	0.9384 (0.0189)	13.5258	0.8537	14.4849	0.805	11.0043	0.5285
27	PC	0.0485 (0.0138)	0.9467 (0.0172)	20.9986	0.3972	9.8382	0.971	7.764	0.8033
28	PLT	0.0461 (0.0153)	0.9416 (0.0172)	17.9797	0.5888	5.1802	0.9996	3.1017	0.9947
29	PMI	0.0778 (0.0382)	0.8977 (0.0597)	12.1308	0.9114	7.946	0.9922	6.9472	0.8611
30	PRY	0.0783 (0.0176)	0.9053 (0.0185)	19.7975	0.4706	27.4171	0.1198	13.7839	0.3147
31	SPM	0.0953 (0.0196)	0.8846 (0.2237)	8.8633	0.9844	13.6736	0.8466	11.2227	0.5099
32	SRG	0.1100 (0.0411)	0.8537 (0.0574)	29.2104	0.0899	28.1461	10.5984	20.2692	0.1019
33	STM	0.0442 (0.0096)	0.9519 (0.0107)	26.1209	0.1618	14.5289	0.8027	8.5747	0.7387
34	STS	0.1221 (0.0398)	0.8452 (0.0362)	27.4943	0.1219	25.5554	0.1809	12.394	0.4145
35	TEN	0.1064 (0.0282)	0.8500 (0.0366)	12.298	0.9054	10.3814	0.9607	5.0749	0.9554
36	TIT	0.0849 (0.0239)	0.9023 (0.0262)	20.4692	0.4289	11.7006	0.9259	5.9236	0.9198
37	TRN	0.0704 (0.0183)	0.9139 (0.0247)	15.7937	0.7293	17.3621	0.6293	12.9452	0.373
38	UBI	0.0824 (0.0196)	0.9083 (0.0215)	20.9975	0.3972	15.5249	0.7456	10.3168	0.5881
39	UCG	0.1085 (0.0227)	0.8916 (0.0203)	19.456	0.4923	17.3197	0.6321	11.1715	0.5142
40	UNI	0.1338 (0.0560)	0.8617 (0.0432)	12.6666	0.8912	6.6912	0.9975	3.9481	0.9844

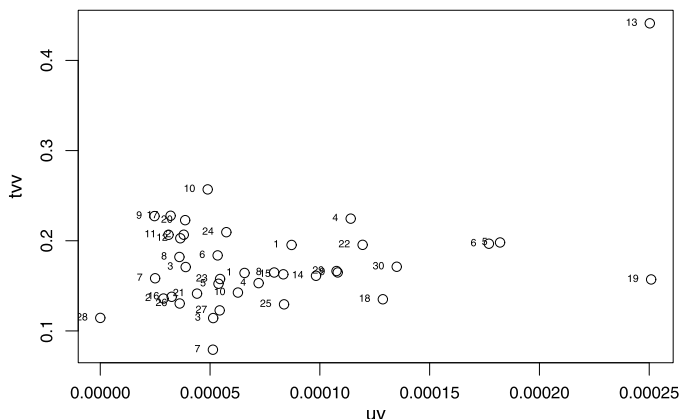


Fig. 5.5. Unconditional volatility and time-varying volatility of the stocks.

Table 5.4
Membership degrees.

	GARCH-FCMdc				GARCH-NC-FCMdc				GARCH-E-FCMdc				GARCH-Tr-FCMdc			
	membership	degrees	crisp		membership	degrees	crisp		membership	degrees	crisp		membership	degrees	crisp	
1	A2A	0.999	0.001	1	0.939	0.050	0.011	1	0.945	0.055	1	0.950	0.050	1		
2	AGL	0.995	0.005	1	0.999	0.000	0.000	1	1.000	0.000	1	1.000	0.000	1		
3	ATL	0.983	0.017	1	0.751	0.223	0.026	1	0.764	0.236	1	0.770	0.230	1		
4	AZM	1.000	0.000	1	0.998	0.001	0.001	1	0.999	0.001	1	0.999	0.001	1		
5	BMPS	1.000	0.000	1	0.999	0.001	0.000	1	0.999	0.001	1	0.999	0.001	1		
6	BP	0.128	0.872	2	0.045	0.928	0.028	2	0.044	0.956	2	0.048	0.952	2		
7	BUL	0.940	0.060	1	0.531	0.030	0.439	1	0.897	0.103	1	0.947	0.053	1		
8	BZU	0.998	0.002	1	0.930	0.058	0.012	1	0.936	0.064	1	0.941	0.059	1		
9	CIR	0.989	0.011	1	0.914	0.070	0.016	2	0.924	0.076	1	0.928	0.072	1		
10	CPR	1.000	0.000	1	1.000	0.000	0.000	1	1.000	0.000	1	1.000	0.000	1		
11	ENEL	0.198	0.802	2	0.000	1.000	0.000	2	0.000	1.000	2	0.000	1.000	2		
12	ENI	0.243	0.757	2	0.000	0.999	0.000	2	0.000	1.000	2	0.000	1.000	2		
13	EXO	0.283	0.717	2	0.002	0.006	0.991	3	0.485	0.515	2	outlier				
14	F	0.998	0.002	1	0.965	0.027	0.008	1	0.971	0.029	1	0.973	0.027	1		
15	FNC	0.999	0.001	1	0.955	0.036	0.009	1	0.960	0.040	1	0.964	0.036	1		
16	FSA	0.997	0.003	1	1.000	0.000	0.000	1	1.000	0.000	1	1.000	0.000	1		
17	G	0.043	0.957	2	0.002	0.991	0.007	2	0.004	0.996	2	0.002	0.998	2		
18	GEO	0.987	0.013	1	0.995	0.002	0.003	1	0.998	0.002	1	0.998	0.002	1		
19	IPG	0.650	0.350	1	0.768	0.119	0.113	3	0.868	0.132	1	0.862	0.138			
20	ISP	0.040	0.960	2	0.001	0.997	0.002	2	0.001	0.999	2	0.001	0.999	2		
21	IT	0.999	0.001	1	1.000	0.000	0.000	1	1.000	0.000	1	1.000	0.000	1		
22	LTO	0.128	0.872	2	0.015	0.979	0.006	2	0.015	0.985	2	0.015	0.985	2		
23	LUX	1.000	0.000	1	0.990	0.007	0.003	1	0.992	0.008	1	0.993	0.007	1		
24	MB	0.060	0.940	2	0.000	1.000	0.000	2	0.000	1.000	2	0.000	1.000	2		
25	MED	0.996	0.004	1	0.997	0.001	0.002	1	0.999	0.001	1	0.999	0.001	1		
26	MS	0.994	0.006	1	0.998	0.001	0.002	1	0.999	0.001	1	0.999	0.001	1		
27	PC	0.992	0.008	1	0.989	0.003	0.008	1	0.996	0.004	1	0.997	0.003	1		
28	PLT	0.973	0.027	1	0.946	0.010	0.044	1	0.983	0.017	1	0.989	0.011	1		
29	PMI	0.986	0.014	1	0.883	0.098	0.019	2	0.895	0.105	1	0.900	0.100	1		
30	PRY	0.886	0.114	1	0.706	0.255	0.039	2	0.730	0.270	1	0.733	0.267	1		
31	SPM	0.224	0.776	2	0.007	0.990	0.003	2	0.008	0.992	2	0.007	0.993	2		
32	SRG	0.165	0.835	2	0.000	1.000	0.000	2	0.000	1.000	2	0.000	1.000	2		
33	STM	0.985	0.015	1	0.960	0.008	0.032	1	0.987	0.013	1	0.992	0.008	1		
34	STS	0.000	1.000	2	0.003	0.991	0.007	2	0.004	0.996	2	0.003	0.997	2		
35	TEN	0.117	0.883	2	0.042	0.929	0.029	2	0.041	0.959	2	0.045	0.955	2		
36	TIT	0.842	0.158	1	0.128	0.852	0.019	2	0.138	0.862	2	0.131	0.869	2		
37	TRN	0.994	0.006	1	0.982	0.013	0.005	1	0.985	0.015	1	0.986	0.014	1		
38	UBI	0.880	0.120	1	0.187	0.789	0.024	2	0.199	0.801	2	0.192	0.808	2		
39	UCG	0.057	0.943	2	0.002	0.991	0.007	2	0.004	0.996	2	0.003	0.997	2		
40	UNI	0.027	0.973	2	0.023	0.775	0.202	2	0.052	0.948	2	0.029	0.971	2		

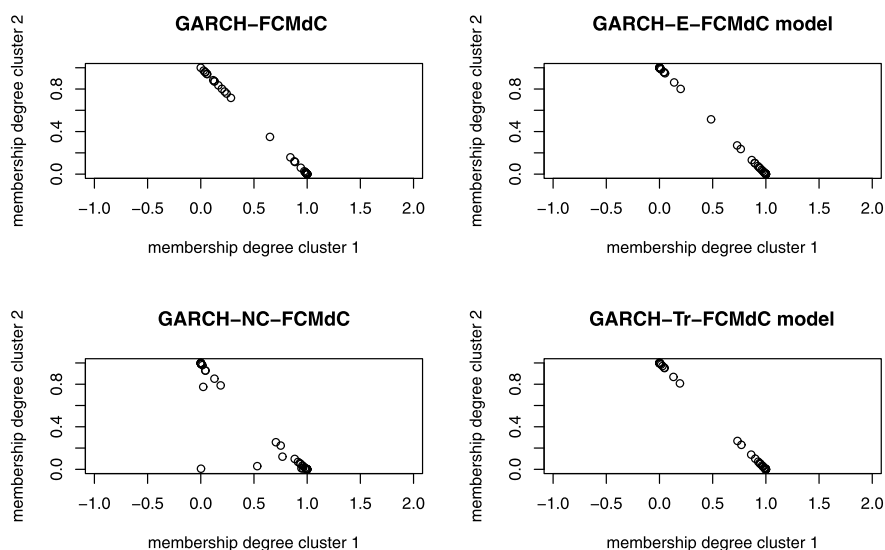


Fig. 5.6. Membership degrees.

As previously, GARCH-NC-FCMdc and GARCH-Tr-FCMdc identify the same outlier, the stock EXO (13). Stock EXO is an outlier as it has a very high value of the variables representing the time-varying volatility and the unconditional volatility (Fig. 5.5). Stock 13 is assigned to cluster 2 in the models GARCH-FCMdc and GARCH-E-FCMdc, but in the model GARCH-E-FCMdc with a low membership, thus not deteriorating the value of the cluster validity index.

The medoids obtained with the robust models GARCH-E-FCMdc, GARCH-NC-FCMdc, GARCH-Tr-FCMdc are (10, 24); the medoids identified with GARCH-FCMdc is (4, 34). For GARCH-FCMdc the medoids are influenced by the presence of stock 13; in fact index 34 has a higher value of tvv with respect to 24. As a consequence stocks 36, 38 move to the other cluster for the reason that they are too far from stock 34. Moreover, beside stock 13 also stock 19 has low membership to the related cluster as the higher value of w_1 leads to consider also the high value of uv as anomalous.

The crisp versions of GARCH-NC-FCMdc and GARCH-Tr-FCMdc assign to cluster 3 or trim also stock 19, beside 13, due to the higher value of w_1 (higher weight to uv).

Separation and compactness of the obtained partitions are illustrated in Fig. 5.6. The analysis of the membership degrees shows that some stocks present uncertain classification in the GARCH-FCMdc and GARCH-E-FCMdc models, specifically stock 13. Stock 13 is in the noise cluster in the model GARCH-NC-FCMdc and is trimmed in the model GARCH-Tr-FCMdc; then the clusters obtained by the highest membership degree are clearly identified.

6. Final remarks and future perspectives

In the literature the task of time series classification has been addressed following different methodological approaches. In this paper, considering a fuzzy approach and adopting a partitioning around medoids procedure we have considered the so-called model-based approach to time series clustering. In particular, different robust clustering models have been proposed for classifying heteroskedastic time series. They are based on a GARCH parametric modeling of the time series, i.e. the unconditional volatility and time-varying volatility GARCH representation of the time series. We firstly have suggested a timid robustification of the fuzzy clustering, i.e. the GARCH-based Fuzzy C-Medoids Clustering (GARCH-FCMdc model). It represents the basic model for the robust clustering models suggested successively. In fact, based on the GARCH-FCMdc model, we have proposed three robust fuzzy clustering models belonging, respectively, to the so-called metric, noise and trimmed approaches, i.e. the GARCH-based Exponential Fuzzy C-Medoids Clustering model (GARCH-E-FCMdc model), the GARCH-based Fuzzy C-Medoids Clustering with Noise Cluster model (GARCH-NC-FCMdc model) and the GARCH-based Trimmed Fuzzy C-Medoids Clustering model (GARCH-Tr-FCMdc model). Each model neutralizes in a different manner the negative effects of the

outliers in the clustering process. In particular, the GARCH-E-FCMdC model achieves its robustness with respect to outliers by taking into account a “robust” distance measure; the GARCH-NC-FCMdC model achieves its robustness with respect to outliers by introducing a noise cluster represented by a noise prototype; the GARCH-Tr-FCMdC model achieves its robustness with respect to outliers by trimming away a certain fraction of outlying time series data.

The usefulness and effectiveness of the proposed clustering models have been illustrated by means of a simulation study and two applications in finance and economics. The simulation and empirical studies have shown the good performances of the robust clustering models in defying properly the optimal partitions of the heteroskedastic time series and in capturing and neutralizing the negative effects of possible outliers in the clustering process.

In the future, we will investigate, in a fuzzy framework, new clustering approaches for classifying time series with non-stationary and/or non-linear structure.

Acknowledgements

The authors thank the editors and the referees for their useful comments and suggestions which helped to improve the quality and presentation of this manuscript.

References

- [1] P. D'Urso, Dissimilarity measures for time trajectories, *Stat. Methods Appl.* 9 (1–3) (2000) 53–83.
- [2] P. D'Urso, Fuzzy C-means clustering models for multivariate time-varying data: different approaches, *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 12 (03) (2004) 287–326.
- [3] T.W. Liao, Clustering of time series data: a survey, *Pattern Recognit.* 38 (11) (2005) 1857–1874.
- [4] R. Coppi, P. D'Urso, Fuzzy unsupervised classification of multivariate time trajectories with the Shannon entropy regularization, *Comput. Stat. Data Anal.* 50 (6) (2006) 1452–1477.
- [5] J. Caiado, N. Crato, A GARCH-based method for clustering of financial time series: international stock markets evidence, in: C.H. Skiadas (Ed.), *Recent Advances in Stochastic Modeling and Data Analysis*, World Scientific Publishing, New Jersey, 2007, pp. 542–551.
- [6] P. D'Urso, L. De Giovanni, Temporal self-organizing maps for telecommunications market segmentation, *Neurocomputing* 71 (13) (2008) 2880–2892.
- [7] E. Otranto, Clustering heteroskedastic time series by model-based procedures, *Comput. Stat. Data Anal.* 52 (10) (2008) 4685–4698.
- [8] E. Otranto, A. Trudda, Classifying Italian pension funds via GARCH distance, in: C. Perna, M. Sibillo (Eds.), *Mathematical and Statistical Methods in Insurance and Finance*, Springer, 2008, pp. 189–197.
- [9] J. Caiado, N. Crato, Identifying common dynamic features in stock returns, *Quant. Finance* 10 (7) (2010) 797–807.
- [10] R. Coppi, P. D'Urso, P. Giordani, A fuzzy clustering model for multivariate spatial time series, *J. Classif.* 27 (1) (2010) 54–88.
- [11] E.A. Maharaj, P. D'Urso, D. Galagedera, Wavelets-based fuzzy clustering of time series, *J. Classif.* 27 (2010) 231–275.
- [12] E. Otranto, Identifying financial time series with similar dynamic conditional correlation, *Comput. Stat. Data Anal.* 54 (1) (2010) 1–15.
- [13] E.A. Maharaj, P. D'Urso, Fuzzy clustering of time series in the frequency domain, *Inf. Sci.* 181 (7) (2011) 1187–1211.
- [14] P. D'Urso, E.A. Maharaj, Wavelets-based clustering of multivariate time series, *Fuzzy Sets Syst.* 193 (2012) 33–61.
- [15] P. D'Urso, C. Cappelli, D. Di Lallo, R. Massari, Clustering of financial time series, *Physica A: Stat. Mech. Appl.* 392 (9) (2013) 2114–2129.
- [16] P. D'Urso, L. De Giovanni, R. Massari, D. Di Lallo, Noise fuzzy clustering of time series by autoregressive metric, *Metron* 71 (3) (2013) 217–243.
- [17] P. D'Urso, D. Di Lallo, E.A. Maharaj, Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks, *Soft Comput.* 17 (1) (2013) 83–131.
- [18] P. D'Urso, L. De Giovanni, E.A. Maharaj, R. Massari, Wavelet-based self-organizing maps for classifying multivariate time series, *J. Chemom.* 28 (1) (2014) 28–51.
- [19] J. Caiado, E.A. Maharaj, P. D'Urso, Time series clustering, in: C. Hennig, M. Meila, F. Murtagh, R. Rocci (Eds.), *Handbook of Cluster Analysis*, Chapman and Hall/CRC, 2015, pp. 241–264.
- [20] R. Coppi, P. D'Urso, Fuzzy K-means clustering models for triangular fuzzy time trajectories, *Stat. Methods Appl.* 11 (1) (2002) 21–40.
- [21] R. Coppi, P. D'Urso, Three-way fuzzy clustering models for LR fuzzy time trajectories, *Comput. Stat. Data Anal.* 43 (2) (2003) 149–177.
- [22] P. D'Urso, Fuzzy clustering for data time arrays with inlier and outlier time trajectories, *IEEE Trans. Fuzzy Syst.* 13 (5) (2005) 583–604.
- [23] Y.-S. Jeong, M.K. Jeong, O.A. Omiaomu, Weighted dynamic time warping for time series classification, *Pattern Recognit.* 44 (9) (2011) 2231–2240.
- [24] A.M. Alonso, E.A. Maharaj, Comparison of time series using subsampling, *Comput. Stat. Data Anal.* 50 (10) (2006) 2589–2599.
- [25] J. Caiado, N. Crato, D. Peña, A periodogram-based metric for time series classification, *Comput. Stat. Data Anal.* 50 (10) (2006) 2668–2684.
- [26] J. Caiado, N. Crato, D. Peña, Comparison of times series with unequal length in the frequency domain, *Commun. Stat., Simul. Comput.* 38 (3) (2009) 527–540.
- [27] E.A. Maharaj, P. D'Urso, A coherence-based approach for the pattern recognition of time series, *Physica A: Stat. Mech. Appl.* 389 (17) (2010) 3516–3537.
- [28] D. Piccolo, A distance measure for classifying ARIMA models, *J. Time Ser. Anal.* 11 (2) (1990) 153–164.
- [29] E.A. Maharaj, A significance test for classifying ARMA models, *J. Stat. Comput. Simul.* 54 (4) (1996) 305–331.

- [30] E.A. Maharaj, Comparison and classification of stationary multivariate time series, *Pattern Recognit.* 32 (7) (1999) 1129–1138.
- [31] E.A. Maharaj, Cluster of time series, *J. Classif.* 17 (2) (2000) 297–314.
- [32] K. Kalpakis, D. Gada, V. Puttagunta, Distance measures for effective clustering of ARIMA time-series, in: *Proceedings IEEE International Conference on Data Mining*, 2001, ICDM 2001, IEEE, 2001, pp. 273–280.
- [33] Y. Xiong, D.-Y. Yeung, Time series clustering with ARMA mixtures, *Pattern Recognit.* 37 (8) (2004) 1675–1689.
- [34] P. D'Urso, L. De Giovanni, R. Massari, Time series clustering by a robust autoregressive metric with application to air pollution, *Chemom. Intell. Lab. Syst.* 141 (2015) 107–124.
- [35] G.M. James, C.A. Sugar, Clustering for sparsely sampled functional data, *J. Am. Stat. Assoc.* 98 (462) (2003) 397–408.
- [36] L.Á. García-Escudero, A. Gordaliza, A proposal for robust curve clustering, *J. Classif.* 22 (2) (2005) 185–201.
- [37] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *J. Econom.* 31 (3) (1986) 307–327.
- [38] R. Coppi, M.A. Gil, H.A. Kiers, The fuzzy approach to statistical analysis, *Comput. Stat. Data Anal.* 51 (1) (2006) 1–14.
- [39] P. D'Urso, E.A. Maharaj, Autocorrelation-based fuzzy clustering of time series, *Fuzzy Sets Syst.* 160 (24) (2009) 3565–3589.
- [40] H. Hwang, W.S. Desarbo, Y. Takane, Fuzzy clusterwise generalized structured component analysis, *Psychometrika* 72 (2) (2007) 181–198.
- [41] A. McBratney, A. Moore, Application of fuzzy sets to climatic classification, *Agric. For. Meteorol.* 35 (1) (1985) 165–185.
- [42] M. Wedel, W.A. Kamakura, *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Press, Boston, 1998.
- [43] W.J. Heiser, P.J. Groenen, Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima, *Psychometrika* 62 (1) (1997) 63–83.
- [44] B. Everitt, S. Landau, M. Leese, *Cluster Analysis*, 4th edition, Arnold Press, London, 2001.
- [45] R. Coppi, P. D'Urso, P. Giordani, Fuzzy C-medoids clustering models for time-varying data, in: B. Bouchon-Meunier, G. Coletti, R. Yager (Eds.), *Modern Information Processing: from Theory Applications*, Elsevier, 2006, pp. 195–206.
- [46] L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, vol. 344, John Wiley & Sons, 2005.
- [47] R. Krishnapuram, J. Kim, A note on the Gustafson–Kessel and adaptive fuzzy clustering algorithms, *IEEE Trans. Fuzzy Syst.* 7 (4) (1999) 453–461.
- [48] R. Krishnapuram, A. Joshi, O. Nasraoui, L. Yi, Low-complexity fuzzy relational clustering algorithms for web mining, *IEEE Trans. Fuzzy Syst.* 9 (4) (2001) 595–607.
- [49] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, Norwell, MA, USA, 1981.
- [50] R.N. Davé, Characterization and detection of noise in clustering, *Pattern Recognit. Lett.* 12 (11) (1991) 657–664.
- [51] J. Kim, R. Krishnapuram, R. Davé, Application of the least trimmed squares technique to prototype-based clustering, *Pattern Recognit. Lett.* 17 (6) (1996) 633–641.
- [52] K.-L. Wu, M.-S. Yang, Alternative c-means clustering algorithms, *Pattern Recognit.* 35 (10) (2002) 2267–2278.
- [53] F. Klawonn, F. Höppner, Fuzzy cluster analysis from the viewpoint of robust statistics, in: R. Seising (Ed.), *Views on Fuzzy Sets and Systems from Different Perspectives*, Springer, Berlin, 2009, pp. 439–455.
- [54] R. Winkler, F. Klawonn, R. Kruse, Fuzzy clustering with polynomial fuzzifier function in connection with M-estimators, *Appl. Comput. Math.* 10 (1) (2011) 146–163.
- [55] R. Coppi, P. D'Urso, P. Giordani, Fuzzy and possibilistic clustering for fuzzy data, *Comput. Stat. Data Anal.* 56 (4) (2012) 915–927.
- [56] J.M. Leski, Fuzzy c-ordered-means clustering, *Fuzzy Sets Syst.* 286 (2016) 114–133.
- [57] L.Á. García-Escudero, A. Gordaliza, Robustness properties of k means and trimmed k means, *J. Am. Stat. Assoc.* 94 (447) (1999) 956–969.
- [58] D.-Q. Zhang, S.-C. Chen, A comment on “Alternative c-means clustering algorithms”, *Pattern Recognit.* 37 (2) (2004) 173–174.
- [59] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (8) (1991) 841–847.
- [60] T. Kamdar, A. Joshi, On creating adaptive web servers using weblog mining, Tech. rep. TR-CS-00-05, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, 2000.
- [61] M. Meilă, Comparing clusterings – an information based distance, *J. Multivar. Anal.* 98 (5) (2007) 873–895.
- [62] R.J. Campello, A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment, *Pattern Recognit. Lett.* 28 (7) (2007) 833–841.
- [63] D.T. Anderson, J.C. Bezdek, M. Popescu, J.M. Keller, Comparing fuzzy, probabilistic, and possibilistic partitions, *IEEE Trans. Fuzzy Syst.* 18 (5) (2010) 906–918.