Weighted score-driven fuzzy clustering of time series with a financial application

Roy Cerqueti[a,b], Pierpaolo D'Urso[a], Livia De Giovanni[c], Massimiliano Giacalone[d], Raffaele Mattera[d]

[a] *Department of Social and Economic Sciences, Sapienza University of Rome, Italy*
[b] *School of Business, London South Bank University, United Kingdom*
[c] *Department of Political Sciences, LUISS Guido Carli, Rome, Italy*
[d] *Department of Economics and Statistics, University of Naples "Federico II", Italy*

Abstract

Time series data are commonly clustered based on their distributional characteristics. The moments play a central role among such characteristics because of their relevant informative content. This paper aims to develop a novel approach that faces still open issues in moment-based clustering. First of all, we deal with a very general framework of time-varying moments rather than static quantities. Second, we include in the clustering model high-order moments. Third, we avoid implicit equal weighting of the considered moments by developing a clustering procedure that objectively computes the optimal weight for each moment. As a result, following a fuzzy approach, two weighted clustering models based on both unconditional and conditional moments are proposed. Since the Dynamic Conditional Score model is used to estimate both conditional and unconditional moments, the resulting framework is called weighted score-driven clustering. We apply the proposed method to financial time series as an empirical experiment.

Keywords: Fuzzy clustering, Dynamic Conditional Score, Conditional moments, Unconditional moments, Optimal weighting procedure for clustering

---

Email addresses: roy.cerqueti@uniroma1.it (Roy Cerqueti), pierpaolo.durso@uniroma1.it (Pierpaolo D'Urso), ldegiovanni@luiss.it (Livia De Giovanni), massimiliano.giacalone@unina.it (Massimiliano Giacalone), raffaele.mattera@unina.it (Raffaele Mattera)

Preprint submitted to Expert Systems with Applications

1. Introduction

Clustering is one of the most common ways to discover similar patterns in a given dataset. The time-series databases are often huge and cannot be adequately managed by human inspectors, so clustering techniques are usually necessary for pattern recognition. The primary purpose of clustering procedures is to group similar data according to a similarity measure or a distance. One of the main problems in time series clustering is computing pairwise distances between dynamic objects, accounting for how the time series are represented. Indeed this kind of data have particular features like serial correlation and usually are both noisy and heteroskedastic with the presence of shifts (Aghabozorgi et al., 2015).

Once groups of similar time series are formed, they can be used differently. For example, in the case of financial time series clustering can be used for asset allocation (Tola et al., 2008, Chen and Huang, 2009, Iorio et al., 2018, Khedmati and Azin, 2020), where groups of similar stocks could be seen as portfolios of assets that share similar characteristics. With this respect, once the C clusters of stocks have been identified, we can construct C portfolios in very different ways. For example, each stock can be equally weighted if we build C naive portfolios. Further, we can apply any optimization technique (e.g. mean-variance, minimum-variance, etc.) to the stocks belonging to each cluster.

From a methodological perspective, time series clustering methods can be divided into three main classes: observation-based clustering, feature-based clustering and model-based clustering (Caiado et al., 2015).

The first, namely the observation-based clustering, uses raw data (D'Urso, 2004, Coppi et al., 2010, D'Urso et al., 2018, D'Urso and Massari, 2019). In order to deal with time series of different lengths, the observation-based clustering methods can be built upon the so-called Dynamic Time Warping (Wang et al., 2019, Li et al., 2020) – which is a well-known technique for finding an optimal alignment between two given (time-dependent) sequences under certain restrictions – can be exploited for dealing with time series of different length. In general, the time series clustering methods belonging to this approach are particularly useful with a short time series but are not the most accurate because they miss evaluating important characteristics of the time series.

The second, the feature-based clustering, overcomes the main limitation of the observation-based approaches. Indeed, these methods consider suitable features derived for the time series for clustering. In the case of time-domain features, it is common to account for the autocorrelation function (ACF) (Alonso and Maharaj, 2006, D'Urso and Maharaj, 2009), the partial autocorrelation function (PACF) (Caiado et al., 2006) or the quantile auto-covariance Lafuente-Rego and Vilar (2016). In the frequency domain, the commonly employed features are the periodogram and its transformations (Maharaj and D'Urso, 2011, Caiado et al., 2020), coherence (Maharaj and D'Urso, 2010), the cepstral coefficients (D'Urso et al., 2020) or the quantile cross-spectral density (López-Oriona and Vilar, 2021).

The last class, i.e. model-based clustering, assumes that a certain statistical model generates the time series. Most model-based clustering procedures' spirit is to group objects according to the estimated quantities or parameters. The main advantage of these methods is that the time series does not need to be of equal length. Examples are the ARIMA (Piccolo, 1990, Maharaj, 1996, 2000, D'Urso et al., 2013b), the GARCH (Otranto, 2008, 2010, Caiado and Crato, 2010, D'Urso et al., 2016), the Threshold Autoregressive (TAR) (Aslan et al., 2018), but also the approaches based on copula Disegna et al. (2017), splines (Iorio et al., 2016, D'Urso et al., 2021) or distribution parameters (D'Urso et al., 2017, Mattera et al., 2021) fall within this class of methods.

This paper considers the problem of clustering time series data according to their estimated moments. More in detail, we aim to develop a novel approach that faces some open issues in moment-based clustering.

First of all, we include in the clustering model high-order moments. Following model-based approaches, many authors (e.g. Otranto, 2008, 2010, D'Urso et al., 2016) propose to cluster time series according to conditional variance estimates. Nevertheless, despite being very important for clustering financial time series, we claim that the conditional variance is not the only moment of interest. For example, De Luca and Zuccolotto (2011, 2021) proposed a clustering algorithm for time series with similar tails. Moreover, in the context of feature-based clustering, static higher moments such as skewness and kurtosis are commonly seen as essential features to consider (Fulcher and Jones, 2014, Mori et al., 2015).

Second, we deal with a general framework of time-varying moments rather than static quantities. There are several statistical models explicitly thought to study the dynamic behaviour of the conditional distribution of higher moments (e.g. see Harvey and Siddique, 1999, León et al., 2005). Recently, Creal et al. (2013) and Harvey and Sucarrat (2014) developed the Dynamic

Conditional Score (DCS also called Generalized Autoregressive Score), a very general statistical model that considers the score function of the predictive model density as the driving mechanism for time-varying parameters. The DCS is an essential tool for describing the financial time series's volatility and tails' heaviness. We claim that it is possible to cluster the time series according to the estimated conditional moments obtained using DCS. In other words, similarly to the proposal of clustering time series according to the estimated conditional variances, we propose to cluster time series according to the conditional moments estimated with a DCS[1].

Third, since it is reasonable to assume that each moment has its own relevance in explaining the entire data distribution, we avoid implicit equal weighting by developing a clustering procedure that objectively computes the optimal weight for each moment. Standard fuzzy clustering algorithms consider the common Euclidean distance in defining the dissimilarities among the objects. However, using a simple Euclidean distance would seriously affect the performance of the clustering algorithm because it would assign equal weight to each moment, while it could be that the moments exhibit different degrees of relevance in the definition of the clusters. Therefore, as in D'Urso et al. (2016) and D'Urso and Massari (2019), we develop a data-driven procedure where the optimal weights are computed within the clustering algorithm. This means that, if the assumption of equal importance holds, the algorithm optimally assigns equal weights to all the moments.

More in detail, the proposed clustering model considers two possible approaches in clustering financial time series: *unconditional* and *conditional* moments-based clustering. Indeed, we estimate a DCS with a given underlying distributional assumption for each time series. On the basis of the specified distribution, we obtain a different number of moments (e.g. two moments in the case of Gaussian density or three moments in the case t-student). From the estimated DCS model, we can retrieve both the (static) *unconditional moments*, which are the values of the moments in the long run, and the (time-varying) *conditional moments* which represent how each moment changes over time before reverting to its "unconditional" value. Note that the *unconditional* moment-based clustering has a different goal with respect to the standard static moment-based clustering. In fact, in this case,

[1]Since the quantities estimated by a statistical model become the input of the clustering procedure, the proposed clustering model belongs to the model-based class.

the unconditional moments represent the long-term mean value to which over-time fluctuations will mean-revert. Therefore, with an *unconditional* moment-based clustering, we aim to group time series according to their long-run distribution. Instead, in the *conditional* moment-based clustering, we aim to group time series according to their short-run fluctuations from their long-run value.

Moreover, as an additional point of innovation, differently from previous papers we adopt a fuzzy approach. The fuzzy approach allows each time series to be allocated in two or more clusters. Therefore, we explicitly face the uncertainty related to the assignment of each series to a single cluster. Identifying a clear boundary between clusters is not an easy task in the real world. Consequently, a fuzzy approach seems more attractive than a deterministic one, considering that fuzzy clustering procedures are very efficient at a computational level. Moreover, we use a weighted k-medoids algorithm that is more robust to noise and more insensitive to the outliers than k-means.

The methodological proposal is tested for a large set of financial time series, which have been widely used in clustering through fuzzy methods to classify stocks with a similar rate of return and risk.

The paper is structured as follows. The following section introduces the DCS model, explaining how conditional and unconditional moments could be estimated. Subsequently, in the third section, we explain in detail the proposed *score*-driven clustering approach. Section 4 presents the considered dataset and the employed methodology. Section 5 provides a discussion of the obtained results. The last section highlights the advantages of the new procedure and offers some conclusive remarks.

2. Dynamic Conditional Score model

2.1. Preliminaries

Let be $y_t = (y_t : t = 1, \dots, T)$ a time series generated by the following observation conditional density $p(\cdot)$:

$$y_t \sim p(y_t | f_t, \mathcal{F}_t; \theta), \tag{1}$$

where f_t is a vector of time-varying parameters at time t, $\mathcal{F}_t$ is the available information at time t and θ a vector of static parameters. The length of the vector f_t crucially depends by the assumption we make about the density (1).

As an example, if we specify a Gaussian density, such that $p \sim \mathcal{N}(\mu, \sigma^2)$, we have that $f_t = (\mu_t, \sigma_t^2)$ but with different densities we could obtain more time varying parameters.

The available information $\mathcal{F}_t$, is defined as a collection of the realizations of the time series y_t and of its time varying parameters before time t. Moreover, in the DCS model the dependence of y_t on the parameter θ in (1) is due to the dependence of f_t on θ.

Given two integers $0 \leq n, m \leq T-1$, it is possible to express the DCS of order n and m, $DCS(n,m)$, for the t-th realization f_t of the time-varying parameter vector as follows:

$$f_t = \omega + \sum_{i=1}^{n} \mathbf{A}_i s_{t-i} + \sum_{j=1}^{m} \mathbf{B}_j f_{t-j} \tag{2}$$

where ω is a real vector and the $\mathbf{A}$'s and the $\mathbf{B}$'s are real matrices with an appropriate dimension. All the scalar parameters in $\omega, \mathbf{A}_1, \ldots, \mathbf{A}_n, \mathbf{B}_1, \ldots, \mathbf{B}_m$ are collected in the vector θ introduced before in (1). Moreover, s_t is the *scaled* score of the conditional distribution (1) at time t, and it is a function of the data and the parameters, so that $s_t = s_t(y_t, f_t, \mathcal{F}_t; \theta)$.

Indeed, with the DCS we suppose that the evolution of the time-varying parameter vector f_t is driven by a vector s_t that is proportional to the score of the density (1), namely ∇_t, together with an autoregressive component. Indeed, s_t is defined as:

$$s_t = S_t \cdot \nabla_t \tag{3}$$

where $S_t = S_t(f_t, \mathcal{F}_t; \theta)$ is a positive definite scaling matrix known at time t and $\nabla_t(y_t, f_t, \mathcal{F}_t; \theta)$ is the score of y_t evaluated with respect to f_t, i.e.:

$$\nabla_t = \frac{\partial \log p(y_t | f_t, \mathcal{F}_t; \theta)}{\partial f_t} \tag{4}$$

A common approach of scaling, proposed by Creal et al. (2013), is to consider the score variance. Specifically, the authors proposed to scale using a matrix S_t equal to the inverse of the information matrix of f_t to a power $\gamma \geq 0$:

$$S_t = E_{t-1}\left[\nabla_t \nabla_t'\right]^{-\gamma} \tag{5}$$

where E_{t-1} denotes the expectation at time $t-1$ and the conditional score ∇_t is defined as in (4). The parameter γ usually takes value in the set $\{0, \frac{1}{2}, 1\}$. When $\gamma = 0$, then S_t is the identity matrix and there is no scaling. Differently, if $\gamma = 1$, then the conditional score ∇_t is premultiplied by the inverse to obtain (3) while, if $\gamma = \frac{1}{2}$, ∇_t is scaled to its square-root.

Since the score depends on the complete density and not only on some moments of y_t, the $DCS(n, m)$ model uses the full density structure for updating f_t. We have to highlight that we could get different $DCS(n, m)$ specifications depending on the choice about scaling S_t we make.

A very appealing feature of $DCS(n, m)$ model is that the vector of parameters θ can be estimated by maximum likelihood (see e.g. Creal et al., 2013). As noted by Blasques et al. (2014), stationarity of the underlying time series process guarantees consistency and asymptotic normality for the maximum likelihood estimator.

2.2. Unconditional and conditional moments

Time-varying parameters reflect the moments of the distribution. This is the reason why, henceforth, we refer to the f_t as the time-varying *moments*. According to the model specification (2), it could be highlighted that the time-varying moments f_t are mean-reverting around their long-term mean values, that we define *unconditional moments* κ:

$$\kappa = (I_N - \mathbf{B})^{-1}\omega \tag{6}$$

where I_N is the Identity matrix, $\mathbf{B}$ is the real matrix defined in (2) and each element in the vector κ represents an unconditional moment. By replacing the quantities in (6) with their estimates we obtain the *estimated unconditional moments*. The estimated unconditional moments allow a different representation of the time series y_t. Indeed, let us consider the following collection of time series:

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & \cdots & y_{i,1} & \cdots & y_{N,1} \\ \vdots & \ldots & y_{i,t} & \ldots & \vdots \\ y_{1,T} & \cdots & y_{i,T} & \cdots & y_{N,T} \end{bmatrix} \tag{7}$$

where $\mathbf{Y}$ is a matrix of dimension $T \times N$ with N the number of the original time series. In each column we have an i-th time series of length T, we are able to represent each i-th column as k-th unconditional moments:

$$\mathbf{K} = \begin{bmatrix} \kappa_{1,1} & \dots & \kappa_{1,k} & \dots & \kappa_{1,K} \\ \vdots & \dots & \kappa_{i,k} & \dots & \vdots \\ \kappa_{N,1} & \dots & \kappa_{N,k} & \dots & \kappa_{N,K} \end{bmatrix} \tag{8}$$

where K is the number of unconditional moments. Now, on the i-th row and k-th column we have the k-th unconditional moment of the i-th time series, which has been estimated according to (6).

However, it is also possible to retrieve the time series of the conditional moments from (2). Indeed, once parameters are estimated according to the maximum likelihood approach, we obtain the time series f_t by in-sample predictions. In other words, we assume a $DCS(1,1)$ as follows.

$$\hat{f}_t = \hat{w} + \hat{\mathbf{A}}_1 s_{t-1} + \hat{\mathbf{B}}_1 f_{t-1}, \tag{9}$$

and define the time series $\hat{f}_t$ as *estimated conditional moments* at time t. This fact allows a representation for the time series collection (7) which is different from that of (8):

$$\mathbf{F} = \begin{bmatrix} \hat{f}_{1,1} & \dots & \hat{f}_{k,1} & \dots & \hat{f}_{K,1} \\ \vdots & \dots & \hat{f}_{k,t} & \dots & \vdots \\ \hat{f}_{1,T} & \dots & \hat{f}_{k,T} & \dots & \hat{f}_{K,T} \end{bmatrix} \tag{10}$$

where the element in the t-th row and k-th column is the k-th conditional moments at time t. In other words, we can express each i-th time series in terms of its K conditional moments, that are time series themselves.

Since the K conditional moments are time series, they also have an autoregressive moving average (ARMA) processes representation. Dynamic objects could appear dissimilar (e.g. because of different ARMA orders and/or estimated parameters) even if they share similar properties. A meaningful way to compare time-varying objects is the infinite order autoregressive representation of the process (Piccolo, 1990).

Therefore, once we compute the $AR(\infty)$ representation of each conditional moments, we store the results in the following matrix:

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{1,1} & \dots & \pi_{1,k} & \dots & \pi_{1,K} \\ \vdots & \dots & \pi_{i,k} & \dots & \vdots \\ \pi_{N,1} & \dots & \pi_{N,k} & \dots & \pi_{N,K} \end{bmatrix} \tag{11}$$

In particular, $\pi_{i,k}$ is the $AR(\infty)$ representation of the k-th conditional moment for the i-th time series. Note that, given N time series, we have $N \times K$ conditional moments. Thanks to the representation in (11), we can summarize into one significant number the time-varying behaviour of the estimated moments.

3. Weighted score-driven fuzzy clustering

In what follows, the *Weighted Score-driven Fuzzy C-Medoids Clustering* (WS-FCMd) model is introduced. As briefly mentioned before, a weighted distance measure between all the estimated *unconditional* and *conditional* moments is proposed in this paper, considering that only one moment can be not enough informative about the entire data distribution than all the moments together.

More in detail, our proposal is based on the Fuzzy C-Medoids (FCMd) clustering model (Krishnapuram et al., 1999) specified adequately in terms of weighted conditional and unconditional moments estimated by a given DCS model. As previously stated, the WS-FCMd model computes the weights within the model.

In the case of an *unconditional moments*-based clustering, the model can be formalized as follows:

$$\begin{aligned} \min : \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^{m} \sum_{k=1}^{K} & \left[w_k (\kappa_{i,k} - \tilde{\kappa}_{c,k}) \right]^2 \\ & \sum_{k=1}^{K} w_k = 1, w_k \geq 0 \\ & \sum_{c=1}^{C} u_{i,c} = 1, u_{i,c} \geq 0 \end{aligned} \tag{12}$$

where $u_{i,c}$ denotes the membership degree of the i-th unit to the c-th cluster,

w_k denotes the weight of the k-th estimated unconditional moments coefficient, the parameter $m > 1$ controls for the fuzziness of the partition, $\kappa_{i,k}$ is the k-th unconditional moments estimated for the i-th time series according to the DCS as in (8) and $\tilde{\kappa}_c$ represents the c-th medoid.

In the case of a *conditional moments*-based clustering, the model is instead specified as follows:

$$\min : \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^{m} \sum_{k=1}^{K} [w_k(\pi_{i,k} - \tilde{\pi}_{c,k})]^2 \tag{13}$$
$$\sum_{k=1}^{K} w_k = 1, w_k \geq 0$$
$$\sum_{c=1}^{C} u_{i,c} = 1, u_{i,c} \geq 0$$

The model (13) is almost the same of the (12). Indeed, the only element of difference is the term $\pi_{i,k}$ and its medoid $\tilde{\pi}_c$ that are the $AR(\infty)$ representation of the k-th conditional moments, while also the constraints in (13) are the same of those in (12). In both the models, the weights w_k are associated with the time series distribution characteristics, captured by each conditional or unconditional moment. The optimal solutions to the models (12) and (13) follow the results of D'Urso et al. (2020).

For the problem (12), we have:

$$u_{i,c} = \frac{1}{\sum_{c'=1}^{C} \left[\frac{\sum_{k=1}^{K} (w_k(\kappa_{i,k} - \tilde{\kappa}_{c,k}))^2}{\sum_{k=1}^{K} (w_k(\kappa_{i,k} - \tilde{\kappa}_{c',k}))^2} \right]^{\frac{1}{m-1}}} \tag{14}$$

for the membership degrees and:

$$w_k = \frac{1}{\sum_{k'=1}^{K} \left[\frac{\sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^{m} (\kappa_{i,k} - \tilde{\kappa}_{c,k})^2}{\sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^{m} (\kappa_{i,k'} - \tilde{\kappa}_{c,k'})^2} \right]} \tag{15}$$

for the unconditional moments' weights.

Similarly, in the case of problem (13), the optimal weights w_k are instead given by:

$$w_k = \frac{1}{\sum_{k'=1}^{K}\left[\frac{\sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^m(\pi_{i,k}-\tilde{\pi}_{c,k})^2}{\sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^m(\pi_{i,k'}-\tilde{\pi}_{c,k'})^2}\right]} \tag{16}$$

The proof of the result (14) can be obtained by maximizing the following Lagrangian function:

$$\mathcal{L}(u_{i,c},\lambda) = \sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^m \sum_{k=1}^{K}\left[w_k(\kappa_{i,k}-\tilde{\kappa}_{c,k})\right]^2 - \lambda\left(\sum_{c=1}^{C} u_{i,c} - 1\right) \tag{17}$$

with respect to $u_{i,c}$ and for fixed values of w_k. The result with its proof are exactly the same if we substitute $\kappa_{i,k}$ with $\pi_{i,k}$.

Starting from (17) we could also derive the optimal weights for the unconditional moments, given fixed values for $u_{i,c}$. To get the solution in (15), let's consider the following Lagrangian function:

$$\mathcal{L}(w_k,\lambda) = \sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^m \sum_{k=1}^{K}\left[w_k(\kappa_{i,k}-\tilde{\kappa}_{c,k})\right]^2 - \lambda\left(\sum_{k=1}^{K} w_k - 1\right)$$

Then it follows that:

$$\frac{\partial\mathcal{L}(w_k,\lambda)}{\partial w_k} = 0 \Longleftrightarrow 2w_k\sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^m(\kappa_{i,k}-\tilde{\kappa}_{c,k})^2 - \lambda = 0$$

$$\frac{\partial\mathcal{L}(w_k,\lambda)}{\partial \lambda} = 0 \Longleftrightarrow \sum_{k=1}^{K} w_k - 1 = 0$$

From which we get:

$$w_k = \frac{\lambda}{2\sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^m(\kappa_{i,k}-\tilde{\kappa}_{c,k})^2}$$

In the end, by substitution, we get the (15). Similarly, we solve problem (13). In this case, we consider the following Lagrangian function:

$$\mathcal{L}(w_k, \lambda) = \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m \sum_{k=1}^{K} [w_k(\pi_{i,k} - \tilde{\pi}_{c,k})]^2 - \lambda \left(\sum_{k=1}^{K} w_k - 1 \right)$$

Then it follows that:

$$\frac{\partial \mathcal{L}(w_k, \lambda)}{\partial w_k} = 0 \iff 2w_k \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m (\pi_{i,k} - \tilde{\pi}_{c,k})^2 - \lambda = 0$$
$$\frac{\partial \mathcal{L}(w_k, \lambda)}{\partial \lambda} = 0 \iff \sum_{k=1}^{K} w_k - 1 = 0$$

From which we get:

$$w_k = \frac{\lambda}{2 \sum_{i=1}^{N} \sum_{c=1}^{C} u_{i,c}^m (\pi_{i,k} - \tilde{\pi}_{c,k})^2}$$

In the end, by substitution, we get the (16).

It is important to mention that in the first step, we should set the starting values within the vector w_k. Without any information, the equal weights $w_k = 1/K$ seems the most reasonable choice even if we could also set other values as long as they satisfy the constraints in (12).

About the computational aspect, we have to mention that the alternating optimization algorithm procedure cannot be adopted for solving the problems (12) and (13) because the necessary conditions cannot be derived by differentiating the objective function with respect to the medoids. Instead, following D'Urso et al. (2020), the solutions to the problems have to be found iteratively by adopting a strategy based on Fu's heuristic algorithm. The algorithm for both the conditional and unconditional moments-based clustering are reported in the Algorithm 1 and Algorithm 2 tables.

Two crucial aspects of the proposed procedure are selecting the fuzziness parameter m and the number of clusters C. To accomplish these tasks, we take advantage of the Fuzzy Silhouette (FS) criterion of Campello and Hruschka (2006). The FS is a well-established cluster validity index that measures the within-cluster cohesion and inter-cluster dispersion. This validation index is commonly used for the selection of the number of clusters (Maharaj et al., 2019).

Algorithm 1 Unconditional moments-based clustering

Estimate unconditional moments with (6);
Fix the number of clusters C, the maximum iterations *max.iter* and the fuzzifier m;
Initialize membership degrees $u_{i,c}$ and weights w_k;
Set iter = 0;
Pick initial medoids $\tilde{\kappa} = (\tilde{\kappa}_1, \ldots, \tilde{\kappa}_c, \ldots, \tilde{\kappa}_C)$
repeat
  Store the current medoids $\tilde{\kappa}_{old} = \tilde{\kappa}$;
  Update the weights w_k with (15);
  Update the membership degrees $u_{i,c}$ with (14);
  Select the new medoids:
  for $c = 1$ to C **do**:

$$q = \text{argmin}_{1 \leqslant i' \leqslant N} \sum_{i''=1}^{N} u_{i''c}^{m} \left[\sum_{k=1}^{K} \left[w_k \left(\tilde{\kappa}_{i'',k} - \tilde{\kappa}_{i',k} \right) \right]^2 \right]$$

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

Algorithm 2 Conditional moments-based clustering

Estimate conditional moments with (10);
Compute the matrix (11);
Fix the number of clusters C, the maximum iterations *max.iter* and the fuzzifier m;
Initialize membership degrees $u_{i,c}$ and weights w_k;
Set iter $= 0$;
Pick initial medoids $\tilde{\pi} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_c, \ldots, \tilde{\pi}_C)$
repeat
  Store the current medoids $\tilde{\pi}_{old} = \tilde{\pi}$;
  Update the weights w_k with (16);
  Update the membership degrees $u_{i,c}$ with (14);
  Select the new medoids:
  for $c = 1$ to C **do**:

$$q = \text{argmin}_{1 \leqslant i' \leqslant N} \sum_{i''=1}^{N} u_{i''c}^{m} \left[\sum_{k=1}^{K} \left[w_k \left(\tilde{\pi}_{i'',k} - \tilde{\pi}_{i',k} \right) \right]^2 \right]$$

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

The FS makes explicit use of the fuzzy partition matrix U with elements $u_{i,c}$ and considers the information on the membership degrees contained in U. In the case of high membership, it stresses the importance of units closely placed with respect to the cluster prototypes. In the case of small membership, it reduces the importance of the units placed in overlapping areas. More precisely, the FS could be defined as follows:

$$FS = \frac{\sum_{i=1}^{N}(u_{i,c} - u_{i,c'})^{\alpha} S_i}{\sum_{i=1}^{N}(u_{i,c} - u_{i,c'})^{\alpha}} \tag{18}$$

with:

$$S_i = \frac{(b_i - a_i)}{\max\{b_i, a_i\}}$$

The value a_i is the average distance between the i-th unit and the units belonging to the cluster $p \in C$ having the highest membership degree with i; b_i is the minimum average distance over the clusters of the i-th unit to all units belonging to the cluster $q \in C$ with $q \neq p$; $u_{i,c}$ and $u_{i,c'}$ are the first and second largest elements of the i-th row of the fuzzy partition matrix, respectively; $\alpha \geq 0$ is a weighting coefficient. The effect of varying the α parameter on the weighting terms in (18) is investigated in Campello and Hruschka (2006).

Accordingly, the best partition is the one associated to the highest FS. Hence, we choose C that maximizes the FS.

Moreover, many heuristic approaches have been proposed in the literature for choosing an appropriate fuzziness parameter $m \in (1, +\infty)$. We have to note that values otf m too close o 1 will result in a partition with all membership values close to 0 or 1. In contrast, excessively large values of m will lead to disproportionate overlap with all memberships close to $1/C$ – where C is the number of clusters. Consequently, a very large value of m – say, $m \to +\infty$ – and too close to $m = 1$ are not a suitable selection. Moreover, when m is high, the mobility of the medoids may be lost. For this reason, a value of m between 1 and 2 is usually recommended (Bezdek, 1981). Simulations carried out by Pal and Bezdek (1995) showed that the most accepted value is $m = 2$. Following D'Urso et al. (2020), in our empirical experiments we choose between $m = 1.5$ and $m = 2$ with the aim of maximizing the FS.

A summary of the proposed method's steps is presented in Fig. 1.

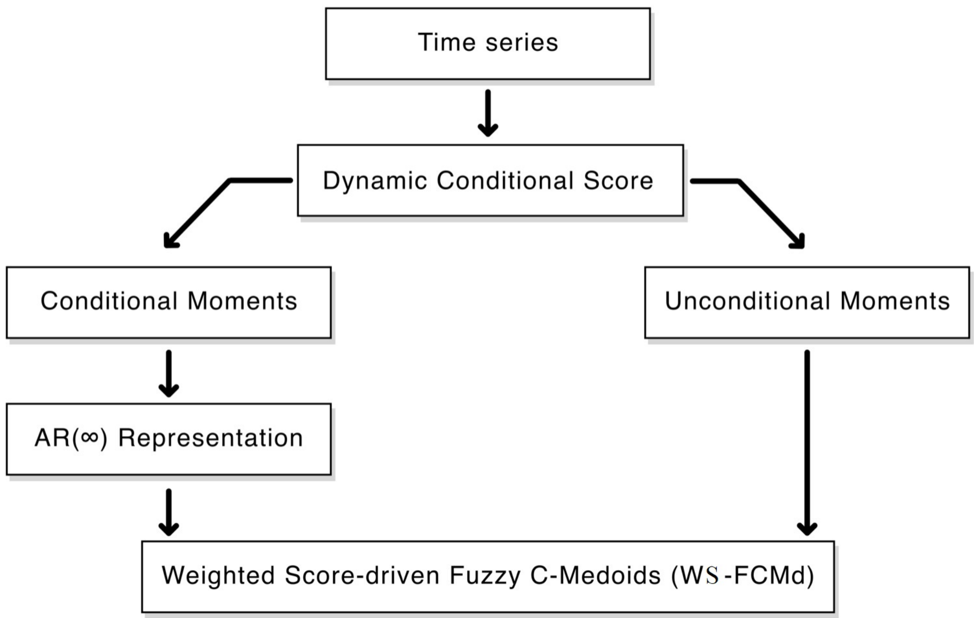

Figure 1: The proposed procedures' flowchart

4. Application to financial time series

4.1. Data

To show the effectiveness of the proposed clustering approach, we evaluate the model with stock market data by considering the Dow Jones Industrial Average Index components[2]. In Fig. 2, we show the daily *returns* computed as:

$$r_{i,t} = \ln\left(\frac{p_{i,t}}{p_{i,t-1}}\right)$$

of the 25 stocks that are components of the Dow Jones Industrial Average Index, being $p_{i,t}$ the daily price of stock i at time t. The considered period

[2]The datasets and codes used for the analysis can be downloaded at the following link `https://www.sites.google.com/view/raffaele-mattera/research`, the name of the file is Weighted score-driven clustering.zip

ranges from January 1st, 2010, to January 1st, 2020. Hence, we have $N = 25$ stocks with $T = 2516$ time-observations.

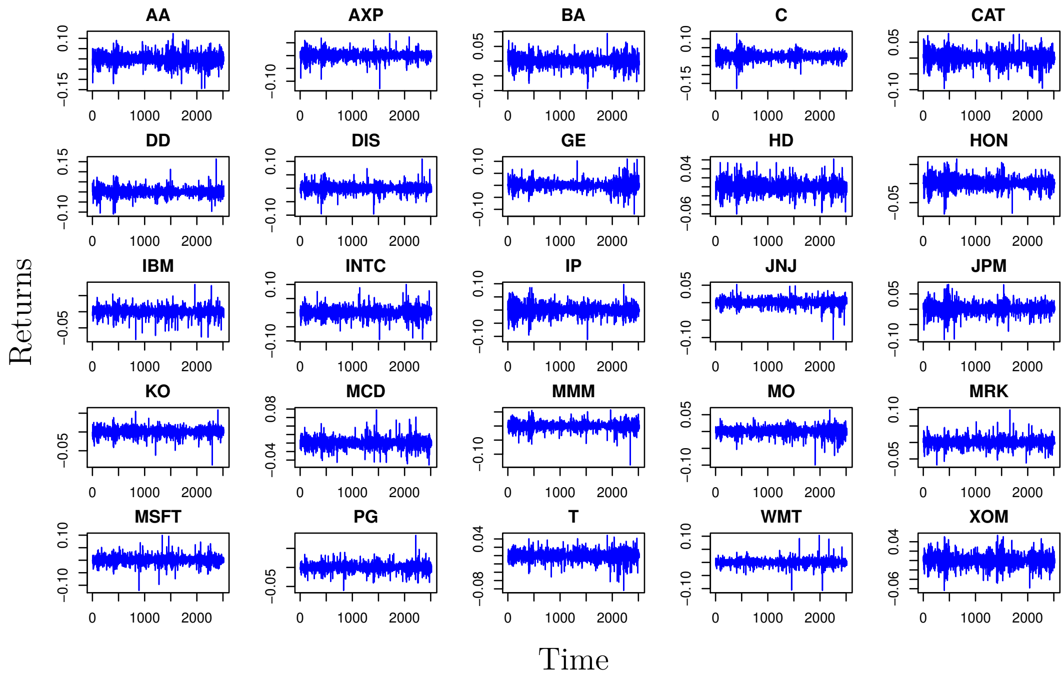

Figure 2: Dow Jones Index components' returns: time series

As Fig. 2 clearly shows, all the considered stock returns are stationary. Moreover, Tab. 1 shows the results of the Augmented Dickey Fuller (ADF) test, finding the absence of a unit root for all the considered stocks in the sample.

Table 1: Results of the unit root ADF test

Stock	ADF statistics
AA	-14.2794***
AXP	-14.1254***
BA	-13.7485***
C	-14.0719***
CAT	-14.1278***
DD	-14.0110***
DIS	-13.6788***
GE	-11.9292***
HD	-14.5836***
HON	-14.1281***
IBM	-13.4206***
INTC	-13.7697***
IP	-14.6593***
JNJ	-14.1532***
JPM	-13.7114***
KO	-14.9122***
MCD	-14.6066***
MMM	-14.1161***
MO	-13.2748***
MRK	-13.9701***
MSFT	-14.1174***
PG	-13.7404***
T	-14.2892***
WMT	-13.6355***
XOM	-14.0907***

Note: Table shows the ADF statistic. *** means p-value< 0.01

With the Fig. 3, instead, we notice that the Dow Jones components are far from being normally distributed, hence confirming also another *stylized fact* of financial securities (e.g. Cerqueti et al., 2019, 2020).

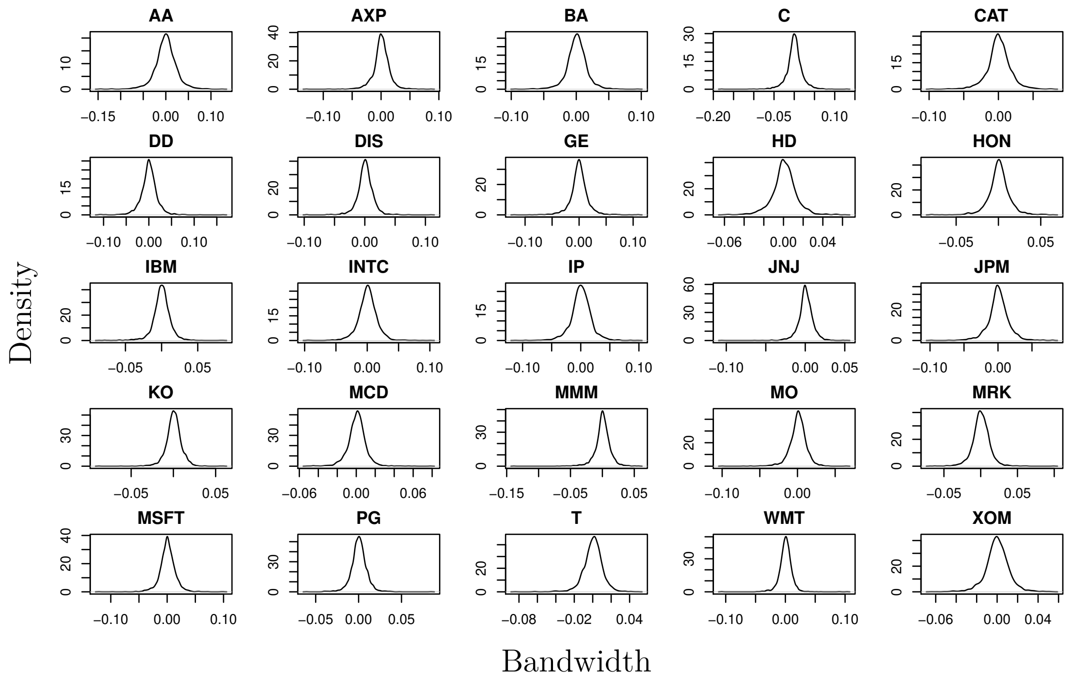

Figure 3: Dow Jones Index components' returns: empirical densities

Specifically, the distributions of the considered stocks show heavy tails and different degrees of skewness. In our context of time-varying high moments of the data distributions, such findings justify the choice of considering different modelling approaches – see the following subsection.

4.2. Employed classes of WS-FCMd clustering

In the empirical applications, we assume three alternative specifications for the density (1): a Gaussian distribution that represents the most simple case, the t-student that is very popular in financial modelling and a Generalized Skew-t distribution that accommodates for both skewness and heavy tails.

In the case of a Gaussian-DCS(1,1) model for all time series we suppose $y_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$ where:

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

assuming, therefore, $f_t = (\mu_t, \sigma_t^2)$. The updating mechanism for time varying parameters μ_t and σ_t^2 could be specified as follows:

$$f_t = \omega + \mathbf{A}s_{t-1} + \mathbf{B}f_{t-1} \tag{19}$$

where s_t is scaled by (5) by setting $\gamma = 1$. In particular, the conditional score vectors are in this case given by:

$$\nabla_t^{(\mu)} = \frac{(y_t - \mu_t)}{\sigma_t^2}$$

$$\nabla_t^{(\sigma)} = \frac{(y_t - \mu_t)^2}{2\sigma_t^4} - \frac{T}{2\sigma_t^2}$$

Summarizing, the model's variables and parameters are:

$$f_t = \begin{pmatrix} \mu_t \\ \sigma_t^2 \end{pmatrix}, \quad \omega = \begin{pmatrix} \omega_\mu \\ \omega_\sigma \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_\mu & 0 \\ 0 & a_\sigma \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} b_\mu & 0 \\ 0 & b_\sigma \end{pmatrix}$$

In order to account for non-normality of the data, we consider a second instance and assume that y_t follows a t-student distribution with location μ_t, scale ϕ_t and degrees of freedom $v_t > 2$ with its density given by:

$$p(y_t|f_t, \mathcal{F}_t; \theta) = \frac{\Gamma\left(\frac{v_t+1}{2}\right)}{\Gamma\left(\frac{v_t}{2}\right)\phi_t\sqrt{\pi v_t}} \left(1 + \frac{(y_t - \mu_t)^2}{v_t \phi_t}\right)^{\frac{v_t+1}{2}}$$

Now we have that $f_t = (\mu_t, \phi_t, v_t)$. Assuming a t-DCS(1,1), the time varying mechanism of the parameters is the same as in (19), with the conditional score vectors are equal to:

$$\nabla_t^{(\mu)} = \frac{(v_t + 1)(y_t - \mu_t)}{v_t\phi_t + (y_t - \mu_t)^2}$$

$$\nabla_t^{(\phi)} = \frac{1}{2\phi_t}\left[\frac{(v_t + 1)(y_t - \mu_t)^2}{v_t\phi_t + (y_t - \mu_t)^2} - 1\right]$$

$$\nabla_t^{(v)} = \frac{1}{2}\left\{\psi\left(\frac{v_t + 1}{2}\right) - \psi\left(\frac{v_t}{2}\right) - \frac{1}{v_t} - \log\left(1 + \frac{(y_t - \mu_t)^2}{v_t\phi_t}\right) + \frac{(v_t + 1)(y_t - \mu_t)^2}{v_t\left[v_t\phi_t + (y_t - \mu_t)^2\right]}\right\}$$

where $\psi(\cdot)$ is the Digamma function. By scaling the conditional score with $\gamma = 1$, the model's variables and parameters are:

$$f_t = \begin{pmatrix} \mu_t \\ \phi_t \\ v_t \end{pmatrix}, \quad \omega = \begin{pmatrix} \omega_\mu \\ \omega_\phi \\ \omega_v \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_\mu & 0 & 0 \\ 0 & a_\phi & 0 \\ 0 & 0 & a_v \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} b_\mu & 0 & 0 \\ 0 & b_\phi & 0 \\ 0 & 0 & b_v \end{pmatrix}$$

This model has been also denoted by Beta-t-EGARCH model by Harvey and Sucarrat (2014).

In the end, we specify a third method that aims to include skewness together with heavy tails in the analysis. To this aim, we consider the Skew-t distribution developed by Fernández and Steel (1998), which is characterized by the following density:

$$\begin{aligned} p(y_t|f_t, \mathcal{F}_t; \theta) &= \frac{2}{\gamma_t + \frac{1}{\gamma_t}} \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)(\pi\nu_t)^{1/2}} \phi_t^{-1} \\ \times &\left[1 + \frac{(y_t - \mu_t)^2}{\nu_t \phi_t^2} \left\{ \frac{1}{\gamma_t^2} I_{[0,\infty)}(y_t - \mu_t) + \gamma_t^2 I_{(-\infty,0)}(y_t - \mu_t) \right\} \right]^{-(\nu_t+1)/2} \end{aligned} \tag{20}$$

where μ_t is the location, ϕ_t the scale, v_t the shape, γ_t the skewness, $I_{(-\infty,0)}(y_t - \mu_t)$ the indicator function for $(y_t - \mu_t) < 0$ and $I_{[0,\infty)}(y_t - \mu_t)$ for $(y_t - \mu_t) \geq 0$. The thickness of the distribution's tails is determined by the parameter v_t, while γ_t determines the amount of mass on both sides of the location μ_t. Symmetric distributions are obtained for $\gamma = 1$. The score of the density (20) can be obtained following Zhu and Galbraith (2010) and related papers. In this last case, the model's variables and parameters are given by:

$$f_t = \begin{pmatrix} \mu_t \\ \phi_t \\ v_t \\ \gamma_t \end{pmatrix}, \quad \omega = \begin{pmatrix} \omega_\mu \\ \omega_\phi \\ \omega_v \\ \omega_\gamma \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_\mu & 0 & 0 & 0 \\ 0 & a_\phi & 0 & 0 \\ 0 & 0 & a_v & 0 \\ 0 & 0 & 0 & a_\gamma \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} b_\mu & 0 & 0 & 0 \\ 0 & b_\phi & 0 & 0 \\ 0 & 0 & b_v & 0 \\ 0 & 0 & 0 & b_\gamma \end{pmatrix}$$

The parameters contained in ω, $\mathbf{A}$ and $\mathbf{B}$, are estimated by MLE and then replaced within the conditional moments' equation (19) to obtain in-sample predictions.

4.3. Alternative clustering models and validation

As stated in the introduction, there are many other approaches for clustering time series. A first approach is to consider the Fuzzy C-medoids (FCMd, see Bezdek, 1981) with simple Euclidean distance between an i-th stocks with its centroid:

$$\min : \sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^{m} d_{i,c}^{2} = \sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^{m} \left[\sum_{t=1}^{T}(r_{i,t} - r_{c,t})^2\right] \tag{21}$$

Since the model (21) is based on raw time series, we define it as *Raw data-based FCMd*.

Within the class of feature-based clustering approaches, an established clustering model is based on the use of the time series' auto-correlation function (e.g. see D'Urso and Maharaj, 2009). By defining $\rho_{i,l}$ the auto-correlation at l-th lag of the i-th time series, the clustering model can be defined as follows:

$$\min : \sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^{m} d_{i,c}^{2} = \sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^{m} \left[\sum_{l=1}^{L}\left(\rho_{i,l} - \rho_{c,l}\right)^2\right] \tag{22}$$

The model (22) is called *ACF-based FCMd*.

In the end, we consider a FCMd algorithm that employs the GARCH-based distance of Caiado and Crato (2010). Let $\mathbf{T}_i = (\hat{\alpha}_i, \hat{\beta}_i)$ a matrix containing the estimated parameters of a GARCH(1,1) process[3] for the i-th time series, the *GARCH-based FCMd* algorithm (see D'Urso et al., 2013a) can be written as:

$$\min : \sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^{m} d_{i,c}^{2} = \sum_{i=1}^{N}\sum_{c=1}^{C} u_{i,c}^{m} \left[\left(\mathbf{T}_i - \mathbf{T}_c\right)' \Omega^{-1} \left(\mathbf{T}_i - \mathbf{T}_c\right)\right] \tag{23}$$

where $\left(\mathbf{T}_i - \mathbf{T}_c\right)' \Omega^{-1} \left(\mathbf{T}_i - \mathbf{T}_c\right)$ is the squared Mahalanobis-like distance between the estimated GARCH parameters and Ω^{-1} is a scaling matrix, equal to the covariance of the estimated parameters.

[3]The GARCH(1,1) is a parsimonious representation of an ARCH(∞) (see Bollerslev, 1986).

In the experiment with real data, we compare the validity of the proposed clustering models with the aforementioned alternatives (21), (22) and (23).

In a controlled environment, where the *ground truth* is available, the quality of the partition can be computed employing the external validation indices, such as the Adjusted Rand Index (ARI, see Hubert and Arabie, 1985). However, studying the validity of a clustering algorithm is a difficult task because this ground truth is unknown. The main idea underlying most clustering validity indices is to measure the within-cluster cohesion and inter-cluster dispersion. The FS of Campello and Hruschka (2006), which has been already described in Section 3, is a well-established and widely used cluster validity index for fuzzy partitions. Accordingly, the best clustering algorithm is the one with the maximum Silhouette.

Therefore, we first compare the proposed approaches based on the FS index. Moreover, we also compare the groups by means of non-parametric tests (see Demšar, 2006). More in detail, we generated 10 datasets with 20 randomly stocks selected from the whole sample. The clustering models are applied to obtain partitions for each of the 10 new sub-samples. Then, the Friedman (1937) test is used to understand if the clustering models perform differently over the 10 random datasets, while the Nemenyi (1963) post-hoc analysis is performed in order to analyze the differences quantitatively.

5. Results

In this section, the clustering results are reported. First, we analyze the case of the Score driven Fuzzy C-Medoids model under Gaussian density for both the securities. Then, we show the results related to the heavy-tailed specification with t-student density and under a Generalized Skew-t distribution, that includes skewness. In the end, also a comparison of different clustering algorithms is presented. More in detail, the performances of the proposed clustering are compared, in terms of Fuzzy Silhouette, with the standard FCMd algorithm, based on a simple Euclidean distance, the ACF-based FCMd approach and the GARCH-based FCMd, both commonly implemented for clustering financial time series.

5.1. Clustering with Gaussian density

The first step of the procedure is to estimate a Gaussian-DCS(1,1) for all the Dow Jones components. Parameters estimates are reported in Appendix A of the paper and will not be shown here to save space.

According to Fig. 1, once the DCS parameters have been estimated, we could either extrapolate the unconditional moments or the conditional moments depending on the type of clusters we want to get.

First, let us analyze the case of *unconditional moments*-based WS-FCMd clustering. The empirical counterpart of the matrix (8), the matrix containing the estimated unconditional moments, is reported in the Table 2. Since we have specified a Gaussian density $K = 2$, so we have a $N \times 2$ matrix.

Stock	$\kappa_{i,1}$	$\kappa_{i,2}$
AA	-0.000216	0.000566
AXP	0.000502	0.000206
BA	0.000795	0.000245
C	0.000371	0.000383
CAT	0.000477	0.000293
DD	0.000290	0.000340
DIS	0.000654	0.000176
GE	0.000004	0.000277
HD	0.000901	0.000155
HON	0.000696	0.000164
IBM	0.000122	0.000153
INTC	0.000541	0.000239
IP	0.000349	0.000315
JNJ	0.000443	0.000090
JPM	0.000564	0.000249
KO	0.000386	0.000087
MCD	0.000574	0.000096
MMM	0.000402	0.000150
MO	0.000568	0.000126
MRK	0.000496	0.000146
MSFT	0.000742	0.000205
PG	0.000409	0.000087
T	0.000342	0.000112
WMT	0.000410	0.000118
XOM	0.000132	0.000136

Table 2: Unconditional moments estimated by a Gaussian-DCS(1,1) for the Dow Jones Index components' daily returns

The first column, $\kappa_{i,1}$ represents the estimated unconditional *first* moment for the i-th asset (i-th row), while the second column, $\kappa_{i,2}$, contains the estimated unconditional *second* moment for the i-th asset.

We define the number of clusters according to the FS approach of Campello and Hruschka (2006). The results are showed in the Fig. 4 for both values $m = 1.5$ and $m = 2$ of the fuzziness parameter.

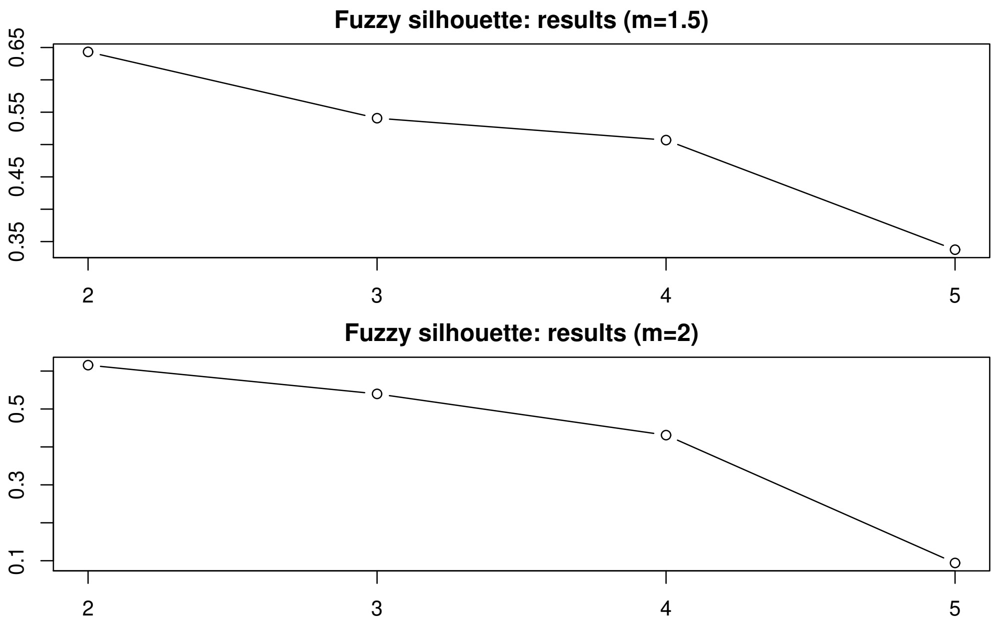

Figure 4: Fuzzy silhouette of unconditional moment-based clustering in the case of a Gaussian-DCS(1,1) with both $m = 1.5$ and $m = 2$

Since, for Gaussian density, the FS takes a higher value for $m = 1.5$ and $C = 2$, we select these values in our WS-FCMd algorithm. Another essential feature of the proposed algorithm is the weights assigned to each unconditional or conditional moment in forming the groups. Table 3 shows the weights assigned to the first and the second estimated unconditional moments.

	$\kappa_{i,1}$	$\kappa_{i,2}$
Weights	30.74%	69.25%

Table 3: Weights of the two unconditional moments (fuzziness $m = 1.5$)

As it appears clearly, the weights associated with the estimated unconditional variance (second moment $\kappa_{i,2}$) is much higher than the unconditional mean (first moment $\kappa_{i,1}$), meaning that this information is more critical in clusters formation.

The cluster analysis results are summarized in the Table 4, where in the last two columns, one can find the estimated membership degrees to the cluster assignment.

	Membership degrees	
Stock	Cluster 1	Cluster 2
AA	0.9114	0.0886
AXP	0.0314	0.9686
BA	0.3168	0.6832
C	0.9986	0.0014
CAT	0.9728	0.0272
DD	1.0000	0.0000
DIS	0.0033	0.9967
GE	0.9296	0.0704
HD	0.0184	0.9816
HON	0.0027	0.9973
IBM	0.0244	0.9756
INTC	0.3094	0.6906
IP	0.9993	0.0007
JNJ	0.0026	0.9974
JPM	0.4753	0.5247
KO	0.0038	0.9962
MCD	0.0019	0.9981
MMM	0.0001	0.9999
MO	0.0002	0.9998
MRK	0.0000	1.0000
MSFT	0.0460	0.9540
PG	0.0035	0.9965
T	0.0017	0.9983
WMT	0.0005	0.9995
XOM	0.0162	0.9838

Table 4: Partitions by the unconditional moments-based clustering procedures with $m = 1.5$ for the Dow Jones Index components' daily returns

Overall, the second group is the most numerous, with 80% of the asset. Moreover, for most of the stocks, the cluster assignment is not very uncertain. However, for both JPM and BA stocks, the group's assignment shows fuzziness since the membership degree of being in both groups 1, and 2 is very close.

Let us consider now the case of clustering with conditional moments. As previously explained, in the case of the *conditional moments*-based WS-FCMd clustering algorithm, a further step is necessary after parameter estimation. Indeed, while in the case of the unconditional moments, we could immediately represent the time series matrix $\mathbf{Y}$ in (7) with the matrix $\mathbf{K}$ in (8), this does not happen with conditional moments.

More specifically, each column of the matrix $\mathbf{Y}$ can be represented as a matrix of dimension $T \times K$ of conditional moments. Since conditional moments are time series themselves, the proposed procedure compares each k-th moment for the i-th time series by an $AR(\infty)$ representation. The empirical $AR(\infty)$ representations, hence the matrix $\mathbf{\Pi}$ in (11), is shown in the Table 7.

Stock	$\pi_{i,1}$	$\pi_{i,2}$
AA	-0.98100	-0.57400
AXP	-0.49700	-0.31700
BA	-0.50000	-0.48900
C	-0.51000	-0.87600
CAT	-0.97900	-0.82400
DD	-0.47500	-0.55900
DIS	-0.48200	-0.24400
GE	-0.50000	-0.76300
HD	-0.50200	-0.81100
HON	-0.46200	-0.62300
IBM	-0.47900	-0.99000
INTC	-0.53900	-0.99300
IP	-0.51400	-0.77600
JNJ	-0.47100	-0.09300
JPM	-0.50400	-0.77700
KO	-0.42900	-0.99000
MCD	-0.41700	-0.99100
MMM	-0.48300	-0.01500
MO	-0.92600	-0.57100
MRK	-0.45300	-0.07800
MSFT	-0.46700	-0.12500
PG	-0.45600	-0.30400
T	-0.46900	-0.99300
WMT	-0.44600	-0.98700
XOM	-0.49700	-0.91900

Table 5: $AR(\infty)$ representation of the conditional moments estimated by a Gaussian-DCS(1,1) for the Dow Jones Index components' daily returns

Where $\pi_{i,1}$ is the AR(∞) representation of the *first* conditional moment for the i-th asset and, similarly, $\pi_{i,2}$ is the AR(∞) representation of the *second* conditional moment.

The number of clusters has to be selected by the FS approach of Campello and Hruschka (2006). According to the FS we selected $C = 2$ and fuzziness $m = 2$.

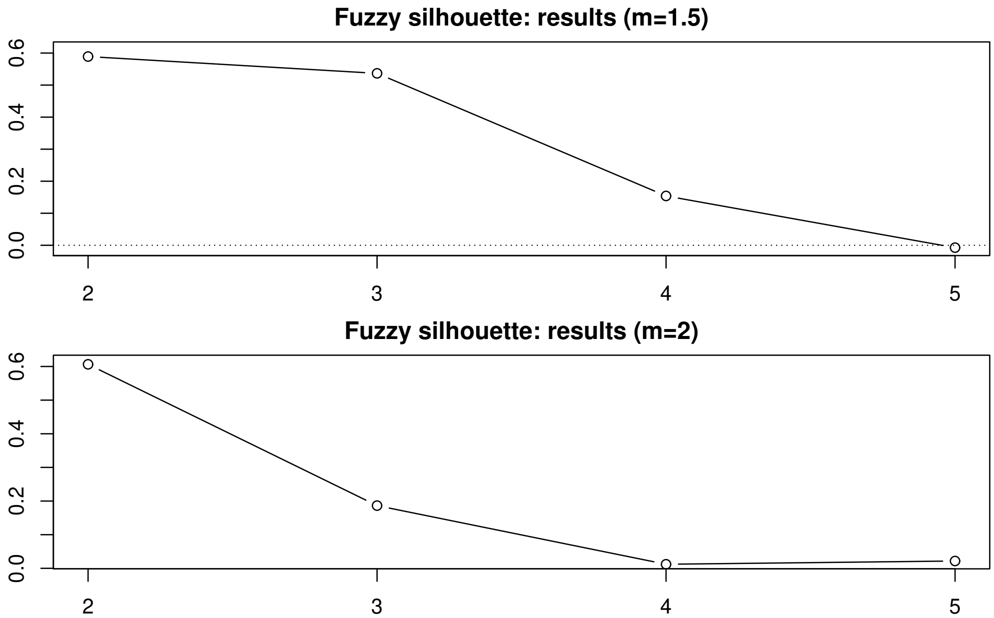

Figure 5: Fuzzy silhouette of conditional moment-based clustering in the case of a Gaussian-DCS(1,1) with both $m = 1.5$ and $m = 2$

Table 6 shows the weights assigned to the first and the second estimated unconditional moments.

	$\pi_{i,1}$	$\pi_{i,2}$
Weights	22.36%	77.63%

Table 6: Weights of the two conditional moments (fuzziness $m = 2$)

As previously, we get that the second conditional moment contains much more information than the first one in clustering formation. Moreover, with respect to the previous experiment with unconditional moments, the relevance of the second moment increases from 69% of Table 3 to 77% of Table 6.

The final assignments are summarized in Table 7, where in the last two columns, we report the estimated membership degrees to the cluster assignment.

Overall, the second group is again the most numerous, with almost 70% of the assets. On the other hand, more assets are placed in the second group according to the conditional moment-based clustering.

Clustering results are different with respect to the previous case. First of all, the medoids are different. While in the first case, DD and MRK are the medoids, in the second one, we have DIS and C stocks. Second, the classification of some stocks is more uncertain in the case of conditional moments based clustering. For example, the stocks AA has a membership of 0.46 to the cluster $c = 1$ and 0.53 for $c = 2$, while in the unconditional moment-based clustering, it belongs to the cluster $c = 1$ with a membership of 0.9. The same applies to other stocks like DD and MO.

These differences are crucial in terms of the applicability of these clusters, for example, in the case of asset allocation. Indeed, as we have already mentioned in the introduction, once the C clusters of stocks have been identified, it is possible to construct C portfolios by applying a naive diversification strategy or a specific optimization technique (e.g. mean-variance, minimum variance, etc.). Clearly, different cluster assignments result in different portfolio construction and, therefore, different out-of-sample performances in terms of risk-return trade-off. Moreover, an investor can exclude stocks whose classification is too uncertain.

	Membership degrees	
Stock	Cluster 1	Cluster 2
AA	0.4627	0.5373
AXP	0.9830	0.0170
BA	0.7135	0.2865
C	0.0000	1.0000
CAT	0.1583	0.8417
DD	0.5042	0.4958
DIS	1.0000	0.0000
GE	0.0454	0.9546
HD	0.0130	0.9870
HON	0.3107	0.6893
IBM	0.0234	0.9766
INTC	0.0243	0.9757
IP	0.0341	0.9659
JNJ	0.9641	0.0359
JPM	0.0334	0.9666
KO	0.0267	0.9733
MCD	0.0282	0.9718
MMM	0.9340	0.0660
MO	0.4661	0.5339
MRK	0.9582	0.0418
MSFT	0.9754	0.0246
PG	0.9884	0.0116
T	0.0248	0.9752
WMT	0.0243	0.9757
XOM	0.0042	0.9958

Table 7: Partitions by the conditional moments-based clustering procedures with $m = 2$ for the Dow Jones Index components' daily returns

5.2. Clustering with heavy-tailed density

Stock returns are far to be normally distributed as they appear clearly from the empirical densities shown in Fig. 3. Therefore, to account for the empirical distribution's heavy tails now, we suppose a t-DCS(1,1) model to estimate parameters and, hence, obtain both conditional and unconditional moments.

In this case the matrix $\mathbf{K}$ is of dimension 25×3, because the specified density $p(\cdot) \sim t$, so we have the location μ, the scale ϕ and the shape v.

Estimates are showed in the Table 8.

Stock	$\kappa_{i,1}$	$\kappa_{i,2}$	$\kappa_{i,4}$
AA	-0.00012	0.00042	8.00000
AXP	0.00090	0.00015	7.99999
BA	0.00099	0.00018	8.00000
C	0.00054	0.00029	7.99999
CAT	0.00068	0.00022	8.00000
DD	0.00043	0.00026	8.00000
DIS	0.00085	0.00013	7.99999
GE	0.00004	0.00021	7.99999
HD	0.00107	0.00012	8.00000
HON	0.00080	0.00012	7.99999
IBM	0.00065	0.00011	7.99999
INTC	0.00068	0.00018	8.00000
IP	0.00058	0.00024	8.00000
JNJ	0.00055	0.00007	8.00000
JPM	0.00070	0.00019	7.99999
KO	0.00052	0.00007	8.00000
MCD	0.00062	0.00007	8.00000
MMM	0.00109	0.00011	7.99999
MO	0.00074	0.00009	8.00000
MRK	0.00060	0.00011	8.00000
MSFT	0.00078	0.00015	8.00000
PG	0.00044	0.00007	8.00000
T	0.00048	0.00008	8.00000
WMT	0.00048	0.00009	8.00000
XOM	0.00017	0.00010	8.00000

Table 8: Unconditional moments estimated by a t-DCS(1,1) for the Dow Jones Index components' daily returns

Where we define $\kappa_{i,4}$ as the unconditional *fourth* moment (the shape v) for the i-th asset. This last column suggests that despite all the assets sharing a different unconditional mean and variance, they have very close unconditional shape (fourth moment) values.

Moreover, the first two estimated unconditional moments $\kappa_{i,1}, \kappa_{i,2}$ in the Table 8 are very different from those of the Table 2 since the underlying statistical model is different and, hence, the estimated parameters too. Moreover,

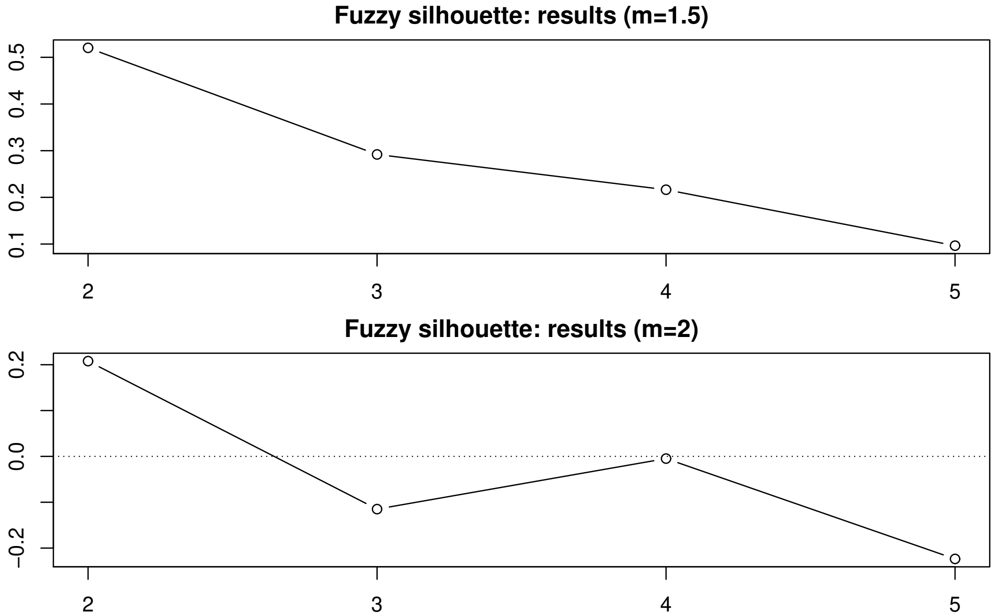

Figure 6: Fuzzy silhouette of unconditional moment-based clustering in the case of a t-DCS(1,1) with both $m = 1.5$ and $m = 2$

the unconditional shape looks the same for all the considered stocks. This fact makes the unconditional shape $\bar{v}$ less useful in clustering different time series.

However, despite the unconditional levels of the shape being similar, the time-varying behaviour among time series are very different. Appendix B of the paper shows the estimated conditional moments, and it is clear that the conditional shape $\hat{v}_t$ is not flat over time and has different patterns for the different stocks.

The Fig. 6 shows the FS values for different clusters. The highest value is reached with fuzziness parameter $m = 1.5$ and $C = 2$, so we use them for the algorithm.

The Table 9, instead, shows the weights assigned to the four estimated unconditional moments.

	$\kappa_{i,1}$	$\kappa_{i,2}$	$\kappa_{i,4}$
Weights	23.13%	56.73%	20.13%

Table 9: Weights of the three unconditional moments (fuzziness $m = 1.5$)

In this case, the unconditional second moment has a principal role in clustering decision making, while the unconditional shape is the lowest since the values are very close among the assets. In other words, different group assignments are mainly due to the differences in the unconditional location and scale parameters.

The group assignment is presented in Table 10. In the case of t-student density, we have that most of the stocks are placed in the second group. Moreover, the uncertainty in the group assignment is a bit higher in the case of heavy-tailed distributional assumption.

	Membership degrees	
Stock	Cluster 1	Cluster 2
AA	0.9137	0.0863
AXP	0.0674	0.9326
BA	0.3726	0.6274
C	0.9977	0.0023
CAT	0.9681	0.0319
DD	1.0000	0.0000
DIS	0.0074	0.9926
GE	0.8952	0.1048
HD	0.0165	0.9835
HON	0.0043	0.9957
IBM	0.0017	0.9983
INTC	0.3274	0.6726
IP	0.9971	0.0029
JNJ	0.0099	0.9901
JPM	0.5412	0.4588
KO	0.0119	0.9881
MCD	0.0076	0.9924
MMM	0.0216	0.9784
MO	0.0072	0.9928
MRK	0.0000	1.0000
MSFT	0.0378	0.9622
PG	0.0117	0.9883
T	0.0086	0.9914
WMT	0.0074	0.9926
XOM	0.0282	0.9718

Table 10: Partitions by the unconditional moments-based clustering procedures with $m = 1.5$ for the Dow Jones Index components' daily returns

Now we analyse the case of conditional moments-based clustering. The $AR(\infty)$ representation, the $\pi_{i,k}$ elements, are showed in the Table 11.

Stock	$\pi_{i,1}$	$\pi_{i,2}$	$\pi_{i,4}$
AA	-0.28600	-0.53000	-0.47200
AXP	-0.66400	-0.81700	-0.47400
BA	-0.64700	-0.85700	-0.61500
C	-0.62300	-0.74500	-0.65900
CAT	-0.35100	-0.80700	-0.58800
DD	-0.60900	-0.75800	-0.64200
DIS	-0.64300	-0.83300	-0.61600
GE	-0.60700	-0.75200	-0.45400
HD	-0.63600	-0.85800	-0.63300
HON	-0.59400	-0.71200	-0.44100
IBM	-0.63000	-0.88300	-0.59500
INTC	-0.70300	-0.51100	-0.60700
IP	-0.66900	-0.77700	-0.60800
JNJ	-0.63600	-0.86600	-0.39600
JPM	-0.66000	-0.73400	-0.67600
KO	-0.60600	-0.66300	-0.62300
MCD	-0.58000	-0.65800	-0.48600
MMM	-0.65800	-0.80100	-0.59700
MO	-0.88200	-0.88600	-0.38500
MRK	-0.61200	-0.88700	-0.61900
MSFT	-0.64100	-0.88100	-0.61700
PG	-0.60200	-0.88300	-0.39600
T	-0.65000	-0.78900	-0.33600
WMT	-0.62500	-0.87900	-0.43000
XOM	-0.66800	-0.80200	-0.63800

Table 11: $AR(\infty)$ representation of the conditional moments estimated by a t-DCS(1,1) for the Dow Jones Index components' daily returns

The first column $\pi_{i,1}$ represents the infinite order AR representation for the conditional first moment time series, the second column $\pi_{i,2}$ reports the same AR(∞) representation for the second conditional moment while the last column $\pi_{i,4}$ contains the infinite AR for the conditional shape of the i-th asset.

As a result of the conditional shape's time variation, we observe different values in the last column of Table 11. Then the numbers of clusters are selected with the Campello and Hruschka (2006) procedure.

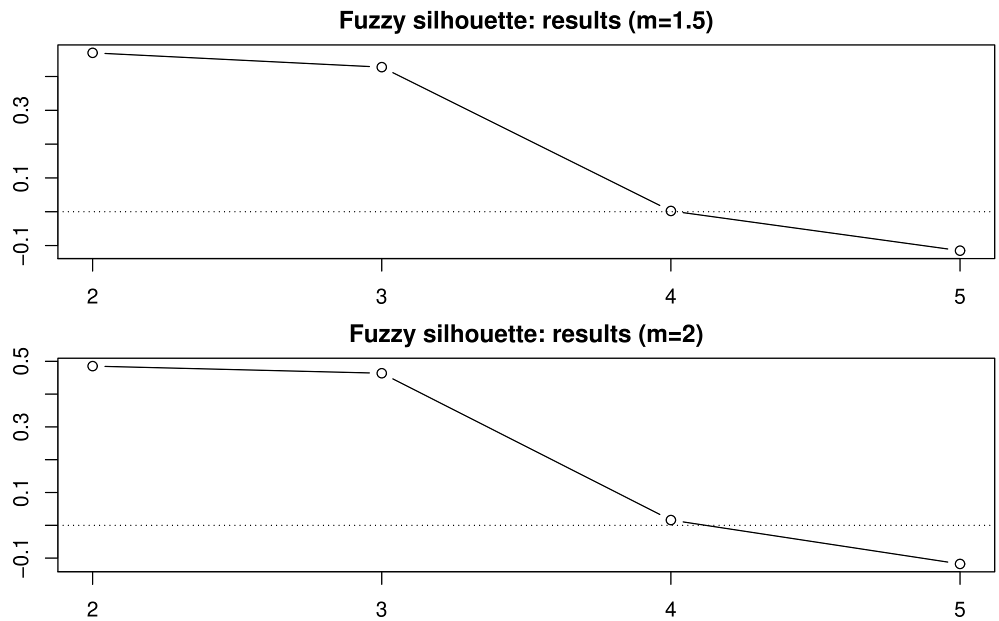

Figure 7: Fuzzy silhouette of conditional moment-based clustering in the case of a t-DCS(1,1) with both $m = 1.5$ and $m = 2$

According to Fig. 7, in both cases $m = 1.5$ and $m = 2$ we get the highest FS values with $C = 2$. However the FS is higher for $m = 2$, hence we select this value of fuzziness for the WS-FCMd algorithm.

The relative importance of the conditional moments in clustering formation is shown in Table 12.

	$\pi_{i,1}$	$\pi_{i,2}$	$\pi_{i,4}$
Weights	10.46%	6.80%	82.72%

Table 12: Weights of the three conditional moments (fuzziness $m = 2$)

Clusters formation, shown in the Table 13, results in a lower level of uncertainty than the unconditional moment-based approach. Moreover, groups have a different composition than the previous case, with the second group that, differently from the Gaussian density specification, is again the most numerous.

	Membership degrees	
Stock	Cluster 1	Cluster 2
AA	0.1558	0.8442
AXP	0.0896	0.9104
BA	0.9999	0.0001
C	0.9650	0.0350
CAT	0.9268	0.0732
DD	0.9840	0.0160
DIS	1.0000	0.0000
GE	0.0255	0.9745
HD	0.9929	0.0071
HON	0.0103	0.9897
IBM	0.9834	0.0166
INTC	0.9751	0.0249
IP	0.9970	0.0030
JNJ	0.0234	0.9766
JPM	0.9430	0.0570
KO	0.9931	0.0069
MCD	0.1690	0.8310
MMM	0.9869	0.0131
MO	0.0526	0.9474
MRK	0.9988	0.0012
MSFT	0.9995	0.0005
PG	0.0235	0.9765
T	0.1019	0.8981
WMT	0.0000	1.0000
XOM	0.9886	0.0114

Table 13: Partitions by the conditional moments-based clustering procedures with $m = 2$ for the Dow Jones Index components' daily returns

5.3. Skewed and heavy-tailed distribution

Although the popularity of the t-student distribution for stock returns modelling, Fig. 3 shows that almost all the stocks in the sample are characterized by a given degree of skewness. Therefore, in what follows, we consider the clustering under skewed and heavy-tailed distribution by assuming a Skew-t distribution.

As usual, let us analyse first the case of unconditional-moments based clustering. Now we have that the matrix $\mathbf{K}$ is of dimension 25×4 because the Fernández and Steel (1998) is characterized by 4 parameters: location μ, scale ϕ, skewness γ and shape v. We define the unconditional skewness for the i-th stock as $\kappa_{i,3}$. Tab. 14 shows the estimated unconditional moments.

	$\kappa_{i,1}$	$\kappa_{i,2}$	$\kappa_{i,3}$	$\kappa_{i,4}$
AA	-0.000122	0.023796	1.000000	7.999991
AXP	0.000897	0.014353	0.999999	7.999986
BA	0.000988	0.015638	1.000000	7.999989
C	0.000542	0.019573	1.000000	7.999986
CAT	0.000670	0.017128	1.000000	7.999990
DD	0.000405	0.018448	1.000000	7.999987
DIS	0.000850	0.013254	1.000000	7.999986
GE	0.000136	0.016629	0.999999	7.999910
HD	0.001074	0.012467	1.000000	7.999990
HON	0.000798	0.012816	1.000000	7.999986
IBM	0.000646	0.012367	0.999999	7.999984
INTC	0.000677	0.015449	1.000000	7.999990
IP	0.000582	0.017741	0.999999	7.999990
JNJ	0.000550	0.009490	1.000000	7.999998
JPM	0.000698	0.015788	1.000000	7.999988
KO	0.000492	0.009333	1.000000	7.999998
MCD	0.000619	0.009821	1.000000	7.999997
MMM	0.001089	0.012248	0.999999	7.999984
MO	0.000741	0.011228	1.000000	7.999998
MRK	0.000596	0.012092	1.000000	7.999989
MSFT	0.000784	0.014302	1.000000	7.999987
PG	0.000442	0.009319	1.000000	7.999998
T	0.000478	0.010578	1.000000	7.999998
WMT	0.000480	0.010862	1.000000	7.999997
XOM	0.000168	0.011682	1.000000	7.999998

Table 14: Unconditional moments estimated by a Skew-t-DCS(1,1) for the Dow Jones Index components' daily returns

Tab. 14 suggests that almost all the stocks are unconditionally symmetric, i.e. in the long run, they have symmetric distributions. Indeed, all the stocks have a long-run skewness parameter equal (or very close) to 1, which

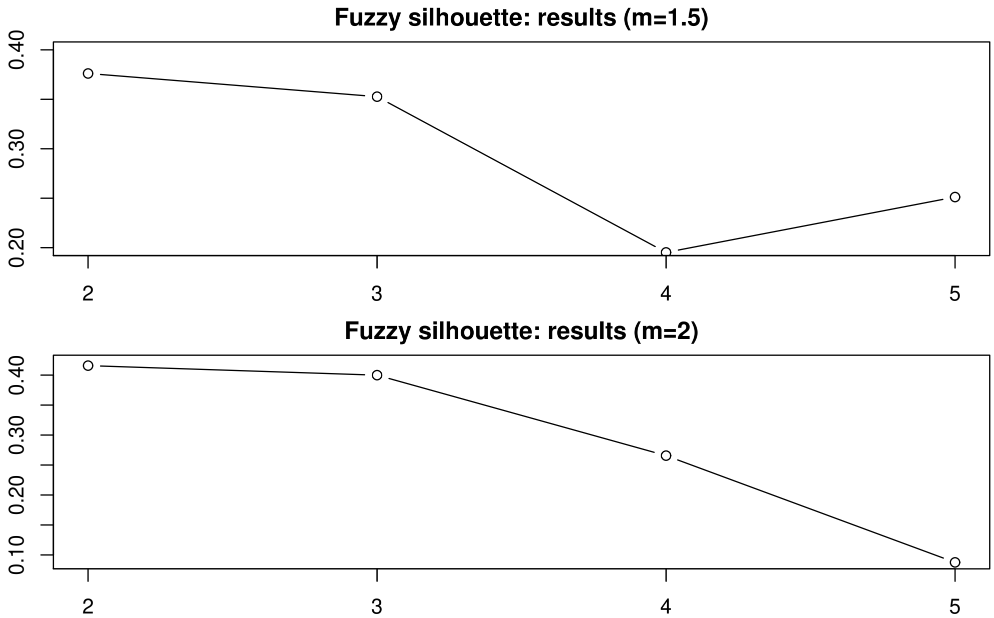

Figure 8: Fuzzy silhouette of unconditional moment-based clustering in the case of a Skew-t-DCS(1,1) with both $m = 2$ and $m = 2$

corresponds to a symmetric version of the distribution shown in (20). Moreover, the stocks also have a similar degree of shape (see column 4 of Tab. 14) in the long run. This means that both the conditional skewness and shape of the time series differentiate for the short-run fluctuations around the long-run values, as it will be clear soon. Nevertheless, they show different unconditional (long-run) values for both mean and variance. Fig. 8 shows the values of the FS. The maximum Silhouette is reached with fuzziness $m = 2$ and $C = 2$ clusters.

Table 15 shows the weights assigned to the four estimated unconditional moment in the case of WS-FCMd clustering.

	$\kappa_{i,1}$	$\kappa_{i,2}$	$\kappa_{i,3}$	$\kappa_{i,4}$
Weights	22.17%	30.27%	25.01%	22.54%

Table 15: Weights of the four unconditional moments (fuzziness $m = 2$)

The optimal weighting is close to an equally weighting scheme. Nevertheless, the process scale seems to be the most important determinant in cluster definition. The membership degrees of the final clustering are reported in Tab. 16.

	Membership degrees	
Stock	Cluster 1	Cluster 2
AA	0.5882	0.4118
AXP	0.6399	0.3601
BA	0.7723	0.2277
C	0.7578	0.2422
CAT	0.9103	0.0897
DD	0.7497	0.2503
DIS	0.5259	0.4741
GE	0.5252	0.4748
HD	0.5061	0.4939
HON	0.2605	0.7395
IBM	0.5528	0.4472
INTC	1.0000	0.0000
IP	0.7805	0.2195
JNJ	0.1793	0.8207
JPM	0.9252	0.0748
KO	0.1914	0.8086
MCD	0.1649	0.8351
MMM	0.5594	0.4406
MO	0.1916	0.8084
MRK	0.0000	1.0000
MSFT	0.4326	0.5674
PG	0.2134	0.7866
T	0.1466	0.8534
WMT	0.1327	0.8673
XOM	0.3003	0.6997

Table 16: Partitions by the unconditional moments-based clustering procedures with $m = 2$ for the Dow Jones Index components' daily returns

The unconditional moment-based clustering under skew-t distribution highlights more fuzzy units regarding previous cases. Indeed, a stock like DIS, GE and HD belong with low membership to cluster 1 (i.e. 0.52 and 0.51). Other stocks like MSFT and MMM have a slightly higher membership degree than 0.5 (0.57 and 0.55, respectively).

Let us now analyze the case of conditional moments-based clustering under the Skew-t assumption. We start by considering the AR(∞) representation of the conditional moments that are shown in Tab. 17.

Stock	$\pi_{i,1}$	$\pi_{i,2}$	$\pi_{i,3}$	$\pi_{i,4}$
AA	-0.35900	-0.06900	-0.62800	-0.79700
AXP	-0.66400	-0.84800	-0.45600	-0.87500
BA	-0.64400	-0.77200	-0.88100	-0.85200
C	-0.62700	-0.87200	-0.65500	-0.59700
CAT	-0.37200	-0.80100	-0.69800	-0.58200
DD	-0.62600	-0.88400	-0.51100	-0.86300
DIS	-0.63700	-0.84900	-0.79300	-0.86100
GE	-0.61300	-0.84300	-0.65800	-0.86400
HD	-0.63900	-0.85600	-0.58600	-0.64300
HON	-0.58800	-0.87700	-0.65500	-0.78700
IBM	-0.62100	-0.68500	-0.85500	-0.80700
INTC	-0.73000	-0.08000	-0.68100	-0.88100
IP	-0.67200	-0.85300	-0.68600	-0.60400
JNJ	-0.63000	-0.67100	-0.56900	-0.88400
JPM	-0.65600	-0.87500	-0.82800	-0.83800
KO	-0.58700	-0.11900	-0.32900	-0.84600
MCD	-0.59600	-0.11200	-0.46500	-0.86800
MMM	-0.68000	-0.84900	-0.60000	-0.85300
MO	-0.85600	-0.62600	-0.67800	-0.73100
MRK	-0.61400	-0.78900	-0.28500	-0.81300
MSFT	-0.63800	-0.69500	-0.27400	-0.87700
PG	-0.57800	-0.13000	-0.55500	-0.87000
T	-0.62600	-0.12800	-0.47300	-0.88200
WMT	-0.63200	-0.70000	-0.74500	-0.86200
XOM	-0.64400	-0.88200	-0.63600	-0.41700

Table 17: $AR(\infty)$ representation of the conditional moments estimated by a Skew-t-DCS(1,1) for the Dow Jones Index components' daily returns

The third column shows the conditional skewness's $AR(\infty)$ representation. It is interesting to note that, despite all stocks having unconditionally symmetric distribution, the degree of skewness is not constant but changes

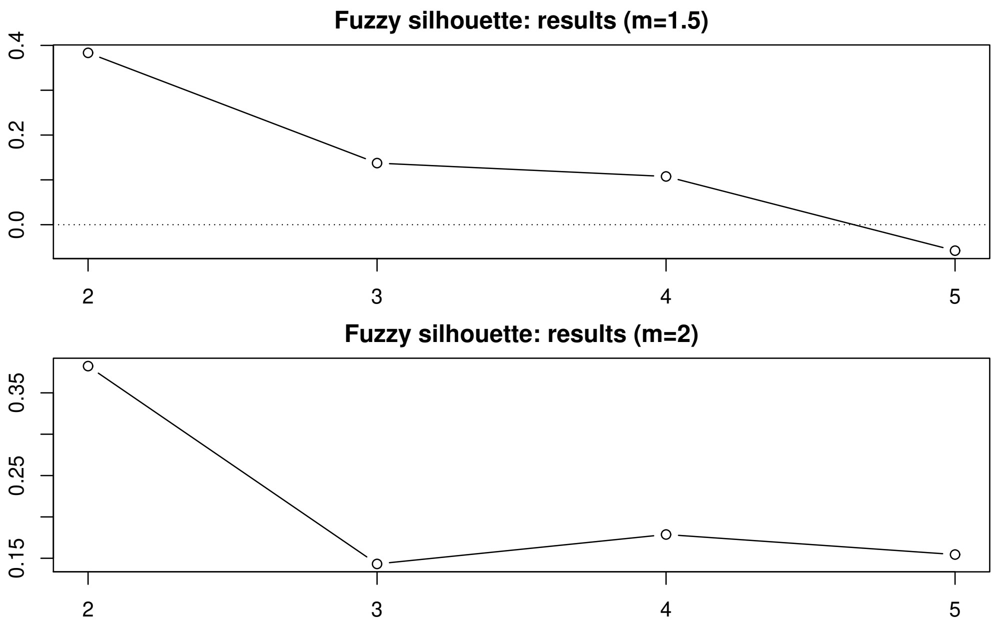

Figure 9: Fuzzy silhouette of conditional moment-based clustering in the case of a Skew-t-DCS(1,1) with both $m = 1.5$ and $m = 2$

over time. Moreover, the stocks differ by the time pattern of the skewness. As a result, we find different conditional values in the third column of Table 17. Then, these values become the input of the WS-FCMd clustering algorithm.

The optimal number of clusters C is selected with the Campello and Hruschka (2006) procedure (see Fig. 9). The maximum Silhouette is achieved with fuzziness $m = 1.5$ and $C = 2$ clusters.

Table 18 shows the weights assigned to the four estimated unconditional moment in the case of WS-FCMd clustering.

	$\pi_{i,1}$	$\pi_{i,2}$	$\pi_{i,3}$	$\pi_{i,4}$
Weights	0.24%	0.22%	0.30%	0.24%

Table 18: Weights of the four unconditional moments (fuzziness $m = 2$)

Also, in this case, the optimal weighting is close to an equally weighting scheme. Nevertheless, by accounting from conditional moments, we now get that most of the clustering result is driven by the time variation in the skewness. At the same time, the primary source was the unconditional scale of the processes.

The resulting partition is shown in Tab. 19.

	Membership degrees	
Stock	Cluster 1	Cluster 2
AA	0.4770	0.5230
AXP	0.9018	0.0982
BA	0.2886	0.7114
C	0.1452	0.8548
CAT	0.2392	0.7608
DD	0.8883	0.1117
DIS	0.2296	0.7704
GE	0.2803	0.7197
HD	0.1784	0.8216
HON	0.0000	1.0000
IBM	0.2273	0.7727
INTC	0.7490	0.2510
IP	0.1854	0.8146
JNJ	1.0000	0.0000
JPM	0.2305	0.7695
KO	0.7658	0.2342
MCD	0.8264	0.1736
MMM	0.8180	0.1820
MO	0.5523	0.4477
MRK	0.7119	0.2881
MSFT	0.7807	0.2193
PG	0.8076	0.1924
T	0.8562	0.1438
WMT	0.3517	0.6483
XOM	0.2811	0.7189

Table 19: Partitions by the conditional moments-based clustering procedures with $m = 1.5$ for the Dow Jones Index components' daily returns

With conditional moments-based clustering, the fuzziness of the partition reduces. For example, MSFT now belongs to its cluster with a high membership, but it applies to MMM, DIS, GE and HD stocks. The only unit with a fuzzy classification is the stock MO with membership equal to 0.55.

5.4. Comparisons

Let us discuss the differences in the final classification of the proposed models with respect to the established alternatives briefly discussed in section 4.3. Clearly, for each of the alternative models discussed in section 4.3, we select the number of clusters C with maximum FS criterion (see Fig. 10).

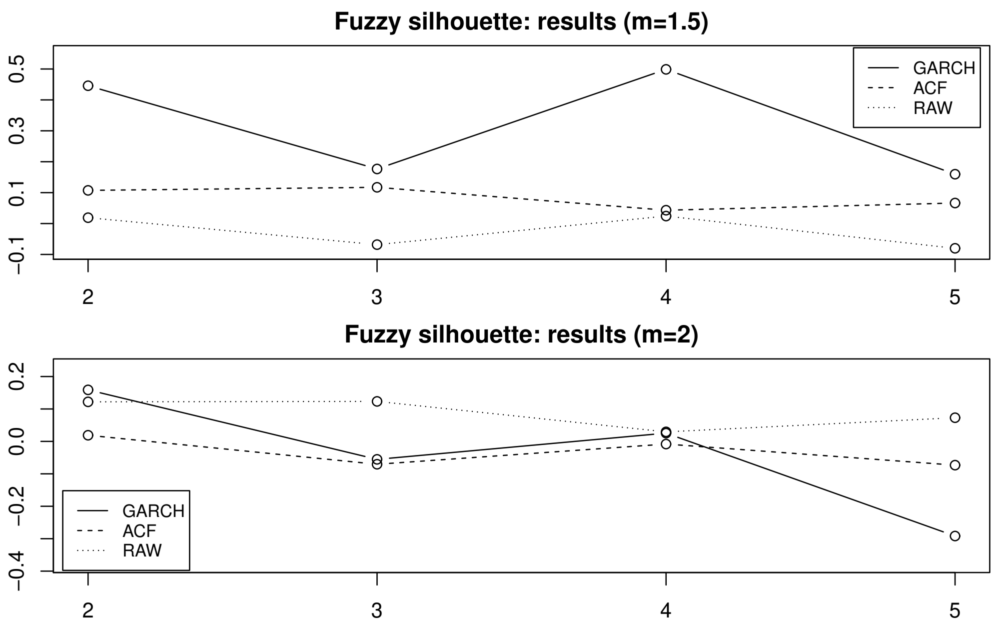

Figure 10: Fuzzy silhouette of the alternative clustering models with both $m = 1.5$ and $m = 2$

Tab. 20 shows the ranks of clustering methods in terms of the values of the selected cluster validity measure, i.e. the FS of Campello and Hruschka (2006) index.

Models		Fuzzy Silhouette
WS-FCMd:		
	Unconditional Gaussian	**0.6432**
	Unconditional t	0.5203
	Unconditional Skew-t	0.4158
	Conditional Gaussian	0.6064
	Conditional t	0.4853
	Conditional Skew-t	0.3834
Raw data-based FCMd		0.0242
ACF-based FCMd		0.1234
GARCH-based FCMd		0.4980

Table 20: Campello and Hruschka (2006) Fuzzy Silhouette and Xie and Beni (1991) indices for the proposed clustering models. Best methods are highlighted in bold.

Tab. 20 shows that the score-driven models are the best approaches with respect to the alternatives according to the Fuzzy Silhouette criterion. However, it is not clear which alternative WS-FCMd specifications outperform the others. Indeed, according to the FS, the WS-FCMd based on the unconditional Gaussian moments is the best model. As expected, among the considered alternatives, the GARCH-based FCMd performs much better than the others, while the raw-data based approach is the worst model. Moreover, the GARCH-based FCMd is ranked third in terms of FS. The Skew-t assumption does not generate improvements in the partitions' quality.

We also evaluate the clustering accuracy with non-parametric tests to better understand the differences. As explained in Subsection 4.3, we generated ten datasets with twenty randomly stocks selected from the whole sample. Then, the clustering models are applied to obtain partitions for each of the ten new sub-samples. The FS values of the clustering models for each dataset is reported in Tab. 21.

Panel A:		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
WS-FCMd:						
	Unconditional Gaussian	0.81	0.73	0.71	0.74	0.77
	Unconditional t	**0.93**	0.90	**0.91**	0.91	**0.91**
	Unconditional Skew-t	0.91	0.79	0.74	0.90	0.90
	Conditional Gaussian	0.90	0.90	0.93	0.91	0.86
	Conditional t	0.83	0.59	0.39	0.67	0.75
	Conditional Skew-t	0.88	**0.93**	0.39	**0.94**	0.90
Raw data-based FCMd		0.06	0.19	0.01	0.21	0.17
ACF-based FCMd		0.24	0.12	0.22	0.14	0.22
GARCH-based FCMd		0.59	0.38	0.65	0.51	0.61
Panel B:		Dataset 6	Dataset 7	Dataset 8	Dataset 9	Dataset 10
WS-FCMd:						
	Unconditional Gaussian	0.69	0.70	0.75	0.71	0.78
	Unconditional t	0.87	0.87	0.90	0.91	0.89
	Unconditional Skew-t	0.65	0.84	0.77	0.94	0.92
	Conditional Gaussian	**0.92**	**0.94**	**0.90**	0.91	0.92
	Conditional t	0.49	0.60	0.56	0.92	0.65
	Conditional Skew-t	0.44	0.38	0.55	**0.97**	**0.93**
Raw data-based FCMd		0.05	0.19	0.02	0.18	0.04
ACF-based FCMd		0.23	0.21	0.17	0.21	0.16
GARCH-based FCMd		0.61	0.56	0.59	0.62	0.42

Table 21: Campello and Hruschka (2006) Fuzzy Silhouette for the proposed clustering models. Best methods are highlighted in bold.

Tab. 21 highlights that for all the considered experiments, the WS-FCMd clustering approaches always provide the most accurate results. More in detail, the WS-FCMd consistently outperform the alternative regardless of the specified probability distribution for most of the experiments.

Then, we rank for each experiment the clustering algorithms in terms of FS. The ranks are reported in Tab. 22.

Panel A:		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
WS-FCMd:						
	Unconditional Gaussian	6	5	4	5	5
	Unconditional t	1	2	2	2	1
	Unconditional Skew-t	2	4	3	4	2
	Conditional Gaussian	3	3	1	3	4
	Conditional t	5	6	6	6	6
	Conditional Skew-t	4	1	7	1	3
Raw data-based FCMd		9	8	9	8	9
ACF-based FCMd		8	9	8	9	8
GARCH-based FCMd		7	7	5	7	7
Panel B:		Dataset 6	Dataset 7	Dataset 8	Dataset 9	Dataset 10
WS-FCMd:						
	Unconditional Gaussian	3	4	4	6	5
	Unconditional t	2	2	2	4	4
	Unconditional Skew-t	4	3	3	2	3
	Conditional Gaussian	1	1	1	5	2
	Conditional t	6	5	6	3	6
	Conditional Skew-t	7	7	7	1	1
Raw data-based FCMd		9	9	9	9	9
ACF-based FCMd		8	8	8	8	8
GARCH-based FCMd		5	6	5	7	7

Table 22: Ranks in terms of Campello and Hruschka (2006) Fuzzy Silhouette for the proposed clustering models.

Then we perform the Friedman (1937) test where the number of data sets is ten, and the number of methods is 9. The Friedman test compares the average ranks of the nine algorithms for the ten datasets. Under the null hypothesis, the nine algorithms have equal ranks over the ten datasets. The test statistics[4] is equal to $\tau_F = 62.027$ with a p-value of $1.863e^{-10}$. Hence, we reject the null hypothesis of equal performances across the different experiments. Then the Nemenyi (1963) post-hoc test has been used to analyze the differences among the nine alternative methods. The Critical Differences (CDs) obtained from the Nemenyi post-hoc test (see Demšar, 2006) are reported in Fig. 11. The clustering algorithms that show no significant differences are grouped using bold horizontal lines.

[4]Given N datasets and k models, the Friedman statistics follows a Chi-square distribution with k-1 and (k-1)(N-1) degrees of freedom.

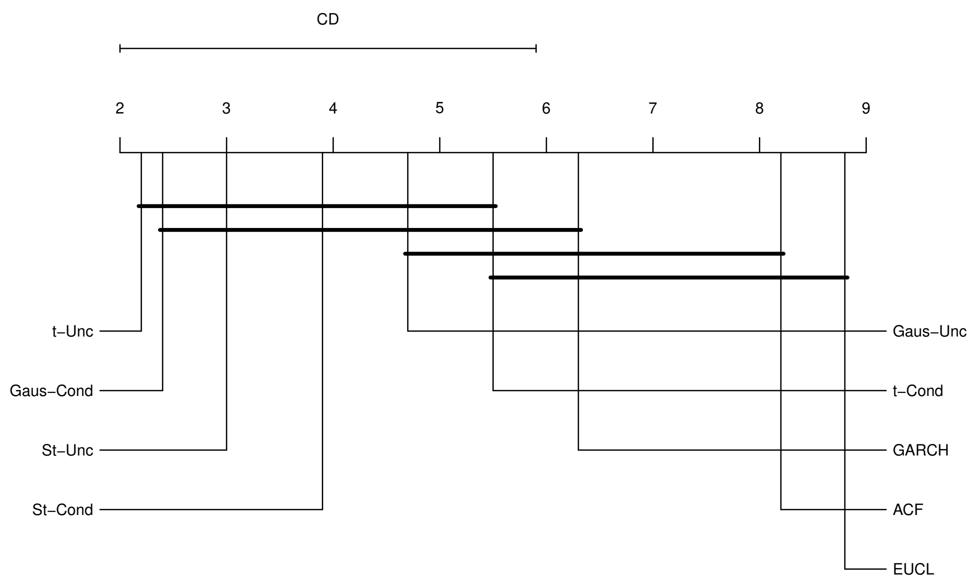

Figure 11: Critical Differences: results

From Fig. 11 we observe that the two best clustering models are the unconditional moments-based WS-FCMd with t-student distribution and the conditional moments-based WS-FCMd algorithm with Gaussian distribution. The average ranks are equal to 2.2 and 2.4, respectively. It is interesting to note that both approaches provide not statistically different results with respect to most of the other WS-FCMd algorithms, which can be considered equally good. With this respect, the most important result is that the alternative clustering models have significantly different and lower ranks than the WS-FCMd models. For example, the Raw-data based and the ACF-based approaches are consistently the worst ones (average ranks equal to 8.8 and 8.2). The GARCH-based FCMd has an average rank equal to 6.3 and provides a statistically different and worst classification than the majority of the WS-FCMd algorithms. The only models with statistically similar performances, i.e. the worst WS-FCMd algorithms, are based on conditional moments under Gaussian distribution and the unconditional moments under t distribution.

Overall, the results highlighted in Fig. 11 suggest that the proposed clustering models provide better and statistically different classification than the proposed alternatives. Nevertheless, there is insufficient evidence to assess which of the best two models outperforms the other. Therefore, based on these results, we would suggest a t distribution for clustering time series based on the long-run distribution characteristics, while a Gaussian distribution for clustering according to short-run deviations.

6. Final remarks

This paper proposes a new approach for clustering financial time series based on the DCS parametric modelling. This general statistical model considers the predictive model density's score function as the driving mechanism for time-varying parameters. For each time series, we estimate the DCS with a specific distribution. Based on the specified distribution, we obtain different moments (e.g. two in the Gaussian distribution or three in the case of t-student one). From the estimated DCS model, we get both the *unconditional moments*, that is, the value of the moment in the long run, and the *conditional moments* which represent how each moment changes over time before reverting to its "unconditional" value.

In this framework, we adopt a fuzzy clustering perspective. In doing so, we admit that each time series can be in more than one cluster with a certain probability level. Indeed, the fuzzy approach implicitly indicates the presence of a second-best cluster. This is a missing property in the traditional clustering methods. Moreover, identifying a clear boundary between clusters is not an easy task in the real world, so a fuzzy approach appears more attractive than a deterministic one.

We present an application to a real financial dataset based on the Dow Jones Industrial Average's stock returns. We find that the two best clustering models are the unconditional moments-based WS-FCMd with t-student distribution and the conditional moments-based WS-FCMd algorithm with Gaussian distribution. Nevertheless, both approaches provide not statistically different results with respect to most of the other WS-FCMd algorithms, which can be considered equally good. An impressive result is that the alternative clustering models have significantly different and lower ranks than the proposed WS-FCMd algorithms.

The proposed approach presents some noticeable practical implications in the context of finance. Specifically, the proposed clustering algorithm can be used to build financial portfolios. Indeed, clustering is recently being used for asset allocation (e.g. Raffinot, 2017), especially in an high-dimensional context. The weighted score-driven approach can be used to construct financial portfolios based on asset returns' distribution by exploiting the evidence documenting time variation in returns' higher moments (e.g. see Jondeau and Rockinger, 2003, Ergün and Jun, 2010, Jondeau and Rockinger, 2012). Indeed, an important stylized fact of financial returns is related to their empirical distribution that is non-Gaussian and heavy-tailed (Cont, 2001). More details about the relevance of time variability of higher moments for asset allocation can be found in Jondeau and Rockinger (2012).

In the future, we will investigate possible robust versions of the proposed DCS-based models to neutralize the harmful effects of possible outliers time series in the clustering process. Moreover, another interesting research direction is based on the development of a clustering model that, accounting for the intrinsic differences between conditional and unconditional moments, allows for a reasonable weighting between two quantities. Finally, future research could also be devoted to the proposed clustering procedure's possible applications in other real-world frameworks. Indeed, even if the proposed clustering approach is effective in clustering the time series of financial nature, it is general. The procedure presented here could also be applied in other disciplines such as engineering, like in the relevant case of signal processes clustering.

References

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

Alonso, A. M. and Maharaj, E. A. (2006). Comparison of time series using subsampling. *Computational statistics & data analysis*, 50(10):2589–2599.

Aslan, S., Yozgatligil, C., and Iyigun, C. (2018). Temporal clustering of time series via threshold autoregressive models: application to commodity prices. *Annals of Operations Research*, 260(1-2):51–77.

Bezdek, J. C. (1981). Objective function clustering. In *Pattern recognition with fuzzy objective function algorithms*, pages 43–93. Springer.

Blasques, F., Koopman, S. J., and Lucas, A. (2014). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Electronic Journal of Statistics*, 8(1):1088–1112.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

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

Caiado, J., Crato, N., and Peña, D. (2006). A periodogram-based metric for time series classification. *Computational Statistics & Data Analysis*, 50(10):2668–2684.

Caiado, J., Crato, N., and Poncela, P. (2020). A fragmented-periodogram approach for clustering big data time series. *Advances in Data Analysis and Classification*, 14(1):117–146.

Caiado, J., Maharaj, E. A., and D'Urso, P. (2015). Time-series clustering. In *Handbook of cluster analysis*, pages 262–285. Chapman and Hall/CRC.

Campello, R. J. and Hruschka, E. R. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, 157(21):2858–2875.

Cerqueti, R., Giacalone, M., and Mattera, R. (2020). Skewed non-gaussian garch models for cryptocurrencies volatility modelling. *Information Sciences*.

Cerqueti, R., Giacalone, M., and Panarello, D. (2019). A generalized error distribution copula-based method for portfolios risk assessment. *Physica A: Statistical Mechanics and its Applications*, 524:687–695.

Chen, L.-H. and Huang, L. (2009). Portfolio optimization of equity mutual funds with fuzzy return rates and risks. *Expert Systems with Applications*, 36(2):3720–3727.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Coppi, R., D'Urso, P., and Giordani, P. (2010). A fuzzy clustering model for multivariate spatial time series. *Journal of Classification*, 27(1):54–88.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

De Luca, G. and Zuccolotto, P. (2011). A tail dependence-based dissimilarity measure for financial time series clustering. *Advances in Data Analysis and Classification*, 5(4):323–340.

De Luca, G. and Zuccolotto, P. (2021). Hierarchical time series clustering on tail dependence with linkage based on a multivariate copula approach. *International Journal of Approximate Reasoning*, 139:88–103.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7:1–30.

Disegna, M., D'Urso, P., and Durante, F. (2017). Copula-based fuzzy clustering of spatial time series. *Spatial Statistics*, 21:209–225.

D'Urso, P. (2004). Fuzzy c-means clustering models for multivariate time-varying data: different approaches. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(03):287–326.

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

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

D'Urso, P., García-Escudero, L. A., De Giovanni, L., Vitale, V., and Mayo-Iscar, A. (2021). Robust fuzzy clustering of time series based on b-splines. *International Journal of Approximate Reasoning*.

D'Urso, P., Maharaj, E. A., and Alonso, A. M. (2017). Fuzzy clustering of time series using extremes. *Fuzzy Sets and Systems*, 318:56–79.

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

D'Urso, P., De Giovanni, L., Massari, R., D'Ecclesia, R. L., and Maharaj, E. A. (2020). Cepstral-based clustering of financial time series. *Expert Systems with Applications*, 161:113705.

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

D'Urso, P. and Maharaj, E. A. (2009). Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems*, 160(24):3565–3589.

D'Urso, P. and Massari, R. (2019). Fuzzy clustering of mixed data. *Information Sciences*, 505:513–534.

Ergün, A. T. and Jun, J. (2010). Time-varying higher-order conditional moments and forecasting intraday var and expected shortfall. *The Quarterly Review of Economics and Finance*, 50(3):264–272.

Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the american statistical association*, 93(441):359–371.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200):675–701.

Fulcher, B. D. and Jones, N. S. (2014). Highly comparative feature-based time-series classification. *IEEE Transactions on Knowledge and Data Engineering*, 26(12):3026–3037.

Harvey, A. and Sucarrat, G. (2014). Egarch models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis*, 76:320–338.

Harvey, C. R. and Siddique, A. (1999). Autoregressive conditional skewness. *Journal of financial and quantitative analysis*, pages 465–487.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1):193–218.

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

Iorio, C., Frasso, G., D'Ambrosio, A., and Siciliano, R. (2018). A p-spline based clustering approach for portfolio selection. *Expert Systems with Applications*, 95:88–103.

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic dynamics and Control*, 27(10):1699–1737.

Jondeau, E. and Rockinger, M. (2012). On the importance of time variability in higher moments for asset allocation. *Journal of Financial Econometrics*, 10(1):84–123.

Khedmati, M. and Azin, P. (2020). An online portfolio selection algorithm using clustering approaches and considering transaction costs. *Expert Systems with Applications*, 159:113546.

Krishnapuram, R., Joshi, A., and Yi, L. (1999). A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering. In *FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No. 99CH36315)*, volume 3, pages 1281–1286. IEEE.

Lafuente-Rego, B. and Vilar, J. A. (2016). Clustering of time series using quantile autocovariances. *Advances in Data Analysis and classification*, 10(3):391–415.

León, Á., Rubio, G., and Serna, G. (2005). Autoregresive conditional volatility, skewness and kurtosis. *The Quarterly Review of Economics and Finance*, 45(4-5):599–618.

Li, H., Liu, J., Yang, Z., Liu, R. W., Wu, K., and Wan, Y. (2020). Adaptively constrained dynamic time warping for time series classification and clustering. *Information Sciences*, 534:97–116.

López-Oriona, Á. and Vilar, J. A. (2021). Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series. *Expert Systems with Applications*, 185:115677.

Maharaj, E. A. (1996). A significance test for classifying arma models. *Journal of Statistical Computation and Simulation*, 54(4):305–331.

Maharaj, E. A. (2000). Cluster of time series. *Journal of Classification*, 17(2):297–314.

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

Maharaj, E. A. and D'Urso, P. (2010). A coherence-based approach for the pattern recognition of time series. *Physica A: Statistical mechanics and its Applications*, 389(17):3516–3537.

Maharaj, E. A. and D'Urso, P. (2011). Fuzzy clustering of time series in the frequency domain. *Information Sciences*, 181(7):1187–1211.

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

Mori, U., Mendiburu, A., and Lozano, J. A. (2015). Similarity measure selection for clustering time series databases. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):181–195.

Nemenyi, P. B. (1963). *Distribution-free multiple comparisons*. Princeton University.

Otranto, E. (2008). Clustering heteroskedastic time series by model-based procedures. *Computational Statistics & Data Analysis*, 52(10):4685–4698.

Otranto, E. (2010). Identifying financial time series with similar dynamic conditional correlation. *Computational Statistics & Data Analysis*, 54(1):1–15.

Pal, N. R. and Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy systems*, 3(3):370–379.

Piccolo, D. (1990). A distance measure for classifying arima models. *Journal of Time Series Analysis*, 11(2):153–164.

Raffinot, T. (2017). Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management*, 44(2):89–99.

Tola, V., Lillo, F., Gallegati, M., and Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, 32(1):235–258.

Wang, X., Yu, F., Pedrycz, W., and Yu, L. (2019). Clustering of interval-valued time series of unequal length based on improved dynamic time warping. *Expert Systems with Applications*, 125:293–304.

Xie, X. L. and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence*, 13(8):841–847.

Zhu, D. and Galbraith, J. W. (2010). A generalized asymmetric student-t distribution with application to financial econometrics. *Journal of Econometrics*, 157(2):297–305.

	ω_μ	ω_σ	a_μ	a_σ	b_μ	b_σ
AA	-9.428E-06	-7.683E-01	1.000E-06	1.898E-01	9.800E-01	8.972E-01
AXP	5.093E-04	-1.013E+00	1.000E-06	1.898E-01	5.000E-01	8.807E-01
BA	5.118E-04	-1.130E+00	1.000E-06	1.898E-01	5.000E-01	8.641E-01
C	3.177E-04	-2.876E-01	1.000E-06	1.898E-01	5.000E-01	9.634E-01
CAT	1.215E-05	-7.012E-01	1.000E-06	1.898E-01	9.800E-01	9.138E-01
DD	4.117E-04	-5.562E-01	1.000E-06	1.898E-01	5.000E-01	9.303E-01
DIS	2.907E-04	-1.604E+00	1.000E-06	1.898E-01	5.000E-01	8.145E-01
GE	8.954E-05	-4.351E-01	1.000E-06	1.898E-01	5.000E-01	9.469E-01
HD	5.614E-04	-7.560E-01	1.000E-06	1.898E-01	5.000E-01	9.138E-01
HON	5.948E-04	-6.070E-01	1.000E-06	1.898E-01	5.000E-01	9.303E-01
IBM	6.086E-05	-1.757E-01	1.000E-06	1.000E-04	5.000E-01	9.800E-01
INTC	2.442E-04	-1.668E-01	1.000E-06	1.000E-04	5.497E-01	9.800E-01
IP	2.587E-04	-4.282E-01	1.000E-06	1.898E-01	5.166E-01	9.469E-01
JNJ	2.262E-04	-1.420E+00	1.000E-06	1.898E-01	5.000E-01	8.476E-01
JPM	5.563E-04	-5.779E-01	1.000E-06	1.898E-01	5.000E-01	9.303E-01
KO	1.911E-04	-1.869E-01	1.000E-06	1.000E-04	5.000E-01	9.800E-01
MCD	2.863E-04	-1.848E-01	1.000E-06	1.000E-04	5.000E-01	9.800E-01
MMM	2.886E-04	-2.216E+00	1.000E-06	1.898E-01	5.000E-01	7.483E-01
MO	4.761E-05	-1.071E+00	1.000E-06	1.898E-01	9.303E-01	8.807E-01
MRK	4.220E-04	-3.100E+00	1.000E-06	1.898E-01	5.000E-01	6.490E-01
MSFT	5.111E-04	-2.420E+00	1.000E-06	1.898E-01	5.000E-01	7.152E-01
PG	2.255E-04	-1.890E+00	1.000E-06	1.898E-01	5.000E-01	7.979E-01
T	1.712E-04	-1.820E-01	1.000E-06	1.000E-04	5.000E-01	9.800E-01
WMT	2.061E-04	-1.809E-01	1.000E-06	1.000E-04	5.000E-01	9.800E-01
XOM	2.785E-04	-4.726E-01	1.000E-06	1.898E-01	5.000E-01	9.469E-01

Table A.23: Parameter estimates from a Gaussian-DCS(1,1) model - stock market data

	ω_μ	ω_ϕ	ω_v	a_μ	ϕ	a_v	b_μ	b_ϕ	b_v
AA	-2.438E-06	-1.553E-01	-4.703E-02	1.000E-06	1.898E-01	5.691E-01	9.800E-01	9.800E-01	9.800E-01
AXP	5.344E-04	-4.660E-01	-4.705E-02	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.469E-01	9.800E-01
BA	6.282E-04	-5.993E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.303E-01	5.000E-01
C	4.144E-04	-2.981E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.634E-01	5.000E-01
CAT	1.341E-05	-4.472E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	9.800E-01	9.469E-01	5.000E-01
DD	4.020E-04	-4.393E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.469E-01	5.000E-01
DIS	3.334E-04	-6.223E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.303E-01	5.000E-01
GE	1.306E-04	-3.100E-01	-4.706E-02	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.634E-01	9.800E-01
HD	5.424E-04	-7.808E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.138E-01	5.000E-01
HON	4.283E-04	-4.780E-01	-4.703E-02	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.469E-01	9.800E-01
IBM	2.760E-04	-7.822E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.138E-01	5.000E-01
INTC	4.091E-04	-1.726E-01	-1.176E+00	1.000E-06	1.898E-01	1.898E-01	5.331E-01	9.800E-01	5.000E-01
IP	3.385E-04	-4.435E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.469E-01	5.000E-01
JNJ	3.162E-04	-8.278E-01	-4.703E-02	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.138E-01	9.800E-01
JPM	5.766E-04	-4.559E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.469E-01	5.000E-01
KO	2.857E-04	-3.522E-01	-1.176E+00	1.000E-06	1.898E-01	1.898E-01	5.000E-01	9.634E-01	5.000E-01
MCD	3.507E-04	-3.485E-01	-4.703E-02	1.000E-06	1.898E-01	1.898E-01	5.000E-01	9.634E-01	9.800E-01
MMM	5.009E-04	-4.828E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.469E-01	5.000E-01
MO	1.282E-04	-9.522E-01	-4.703E-02	1.000E-06	3.794E-01	3.794E-01	8.641E-01	8.972E-01	9.800E-01
MRK	3.110E-04	-9.370E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	8.972E-01	5.000E-01
MSFT	6.156E-04	-7.571E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.138E-01	5.000E-01
PG	2.440E-04	-1.150E+00	-4.703E-02	1.000E-06	3.794E-01	1.898E-01	5.000E-01	8.807E-01	9.800E-01
T	2.713E-04	-6.538E-01	-4.703E-02	1.000E-06	1.898E-01	1.898E-01	5.000E-01	9.303E-01	9.800E-01
WMT	2.742E-04	-1.113E+00	-4.703E-02	1.000E-06	3.794E-01	1.898E-01	5.000E-01	8.807E-01	9.800E-01
XOM	1.340E-04	-6.399E-01	-1.176E+00	1.000E-06	3.794E-01	1.898E-01	5.000E-01	9.303E-01	5.000E-01

Table A.24: Parameter estimates from a t-DCS(1,1) model - stock market data

	ω_μ	ω_ϕ	ω_γ	ω_v	a_μ	ϕ	a_γ	a_v	b_μ	b_ϕ	b_γ	b_v
AA	1.376E-06	-7.476E-02	-5.404E-08	-4.703E-02	1.000E-06	1.000E-04	1.898E-01	5.500E+00	9.800E-01	9.800E-01	9.634E-01	9.800E-01
AXP	5.550E-04	-9.980E-01	-1.372E-05	-4.704E-02	1.000E-06	1.898E-01	1.000E-04	5.500E+00	5.000E-01	7.648E-01	9.800E-01	9.800E-01
BA	6.490E-04	-9.779E-01	-1.536E-07	-4.704E-02	1.000E-06	1.898E-01	1.898E-01	4.172E+00	5.000E-01	7.648E-01	8.145E-01	9.800E-01
C	3.742E-04	-5.344E-01	-1.446E-07	-4.704E-02	1.000E-06	1.898E-01	1.898E-01	5.500E+00	5.000E-01	8.641E-01	5.000E-01	9.800E-01
CAT	1.341E-05	-8.891E-01	-6.574E-07	-4.703E-02	1.000E-06	1.898E-01	1.898E-01	4.172E+00	9.800E-01	7.814E-01	5.497E-01	9.800E-01
DD	4.653E-04	-6.086E-01	4.018E-06	-4.704E-02	1.000E-06	1.898E-01	1.898E-01	4.362E+00	5.000E-01	8.476E-01	9.634E-01	9.800E-01
DIS	3.012E-04	-8.736E-01	2.749E-07	-4.704E-02	1.000E-06	1.898E-01	1.898E-01	4.362E+00	5.000E-01	7.979E-01	9.138E-01	9.800E-01
GE	1.647E-04	-4.210E-01	-4.000E-07	-4.704E-02	1.000E-06	1.898E-01	1.898E-01	5.500E+00	5.000E-01	8.972E-01	5.000E-01	9.800E-01
HD	5.663E-04	-1.249E+00	-3.048E-06	-4.704E-02	1.000E-06	1.898E-01	1.898E-01	4.741E+00	5.000E-01	7.152E-01	9.634E-01	9.800E-01
HON	4.293E-04	-5.920E-01	-5.196E-08	-4.703E-02	1.000E-06	1.898E-01	1.898E-01	4.172E+00	5.000E-01	8.641E-01	5.000E-01	9.800E-01
IBM	2.790E-04	-1.324E+00	-2.990E-07	-4.703E-02	1.000E-06	1.898E-01	1.898E-01	2.655E+00	5.000E-01	6.986E-01	8.972E-01	9.800E-01
INTC	3.506E-04	-8.342E-02	-1.189E-06	-4.704E-02	1.000E-06	1.000E-04	1.000E-04	5.500E+00	5.662E-01	9.800E-01	5.000E-01	9.800E-01
IP	3.225E-04	-8.815E-01	-1.803E-06	-4.704E-02	1.000E-06	1.898E-01	1.898E-01	3.603E+00	5.000E-01	7.814E-01	5.000E-01	9.800E-01
JNJ	3.198E-04	-1.018E+00	1.704E-06	-4.703E-02	1.000E-06	1.898E-01	1.898E-01	5.500E+00	5.000E-01	7.814E-01	9.634E-01	9.800E-01
JPM	5.619E-04	-7.010E-01	1.345E-06	-4.704E-02	1.000E-06	1.898E-01	1.898E-01	5.500E+00	5.000E-01	8.310E-01	6.986E-01	9.800E-01
KO	2.732E-04	-9.349E-02	-4.257E-06	-4.703E-02	1.000E-06	1.000E-04	1.000E-04	5.500E+00	5.000E-01	9.800E-01	9.800E-01	9.800E-01
MCD	3.933E-04	-9.269E-02	-3.923E-05	-4.707E-02	1.000E-06	1.000E-04	1.898E-01	5.500E+00	5.000E-01	9.800E-01	9.800E-01	9.800E-01
MMM	4.836E-04	-7.439E-01	-2.131E-06	-4.703E-02	1.000E-06	1.898E-01	1.898E-01	5.500E+00	5.166E-01	8.310E-01	5.000E-01	9.800E-01
MO	9.553E-05	-1.799E+00	-4.542E-07	-4.703E-02	1.000E-06	1.898E-01	3.794E-01	5.500E+00	8.972E-01	5.993E-01	9.634E-01	9.800E-01
MRK	3.158E-04	-1.404E+00	5.944E-07	-4.703E-02	1.000E-06	1.898E-01	1.000E-04	5.500E+00	5.000E-01	6.821E-01	9.800E-01	9.800E-01
MSFT	6.642E-04	-1.069E+00	2.380E-05	-4.704E-02	1.000E-06	1.898E-01	1.000E-04	4.741E+00	5.000E-01	7.483E-01	9.800E-01	9.800E-01
PG	2.470E-04	-9.355E-02	-2.077E-07	-4.704E-02	1.000E-06	1.000E-04	1.000E-04	5.500E+00	5.000E-01	9.800E-01	5.000E-01	9.800E-01
T	2.588E-04	-9.098E-02	-4.154E-06	-4.703E-02	1.000E-06	1.000E-04	1.898E-01	5.500E+00	5.000E-01	9.800E-01	9.800E-01	9.800E-01
WMT	2.869E-04	-1.662E+00	-3.325E-07	-4.703E-02	1.000E-06	1.898E-01	1.898E-01	4.172E+00	5.000E-01	6.324E-01	5.662E-01	9.800E-01
XOM	1.491E-04	-8.991E-01	8.602E-07	-4.703E-02	1.000E-06	1.898E-01	1.000E-04	1.897E+00	5.000E-01	7.979E-01	9.469E-01	9.800E-01

Table A.25: Parameter estimates from a Skew-t-DCS(1,1) model - stock market data

Appendix B. Estimated conditional moments

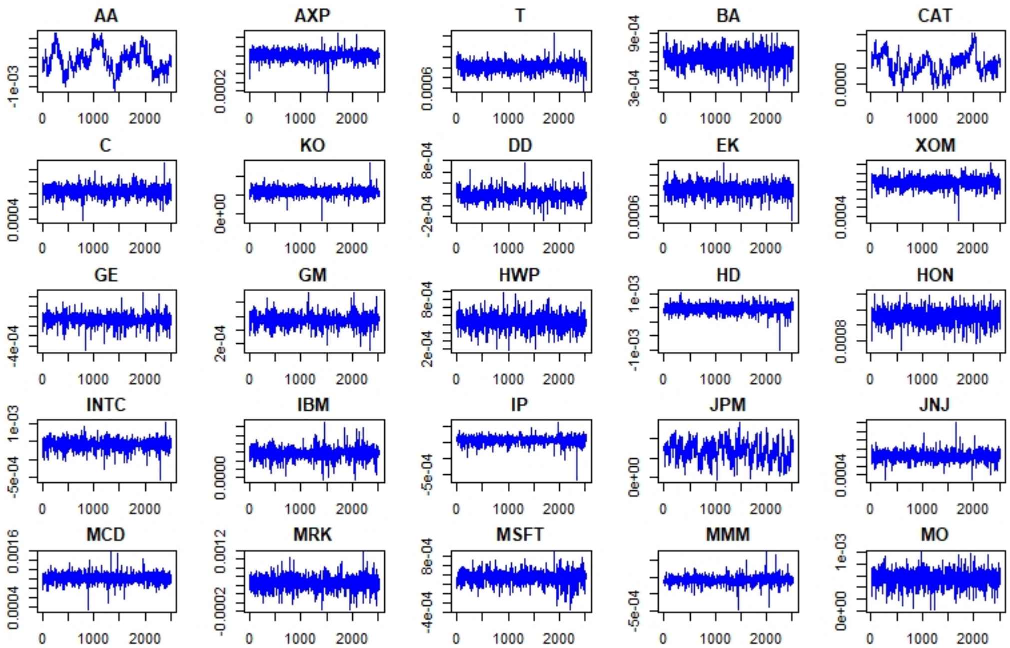

Figure B.12: Conditional mean estimated by the Gaussian-DCS(1,1) for the Dow Jones Index components' daily returns

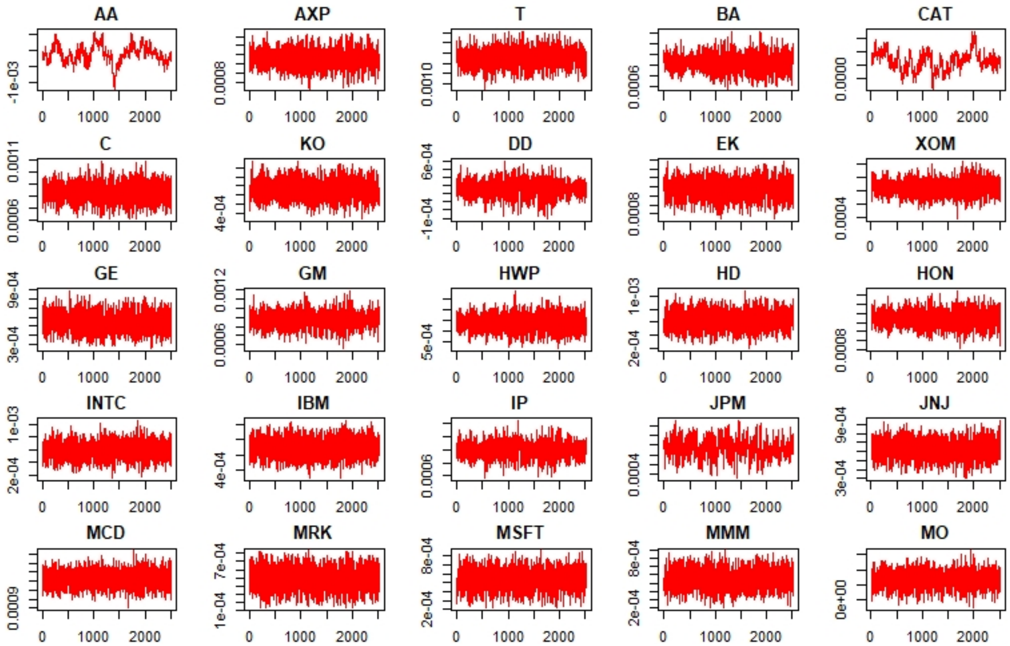

Figure B.13: Conditional mean estimated by the t-DCS(1,1) for the Dow Jones Index components' daily returns

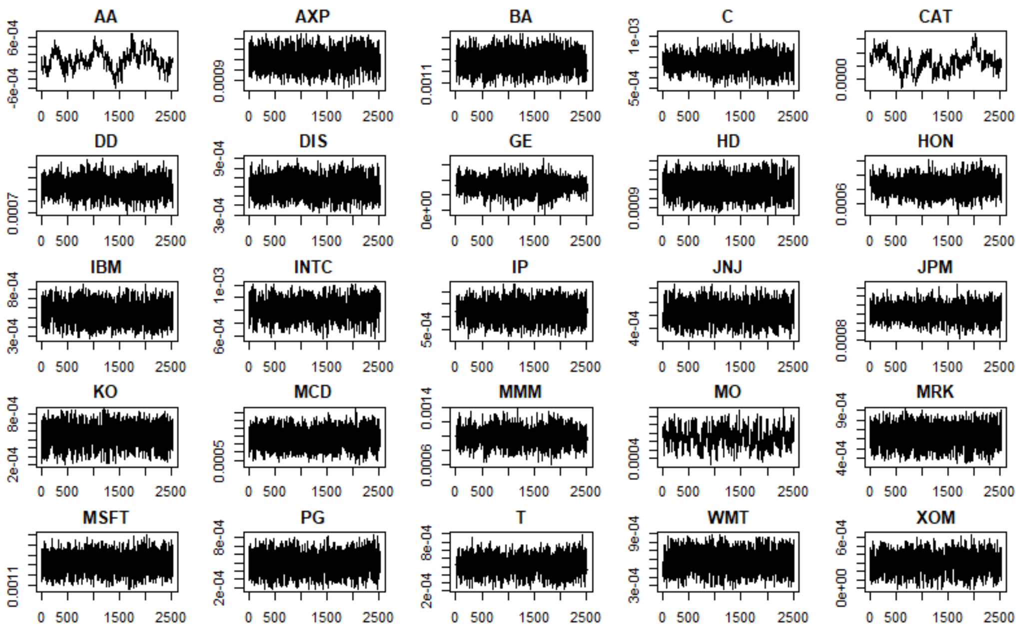

Figure B.14: Conditional mean estimated by the Skew-t-DCS(1,1) for the Dow Jones Index components' daily returns

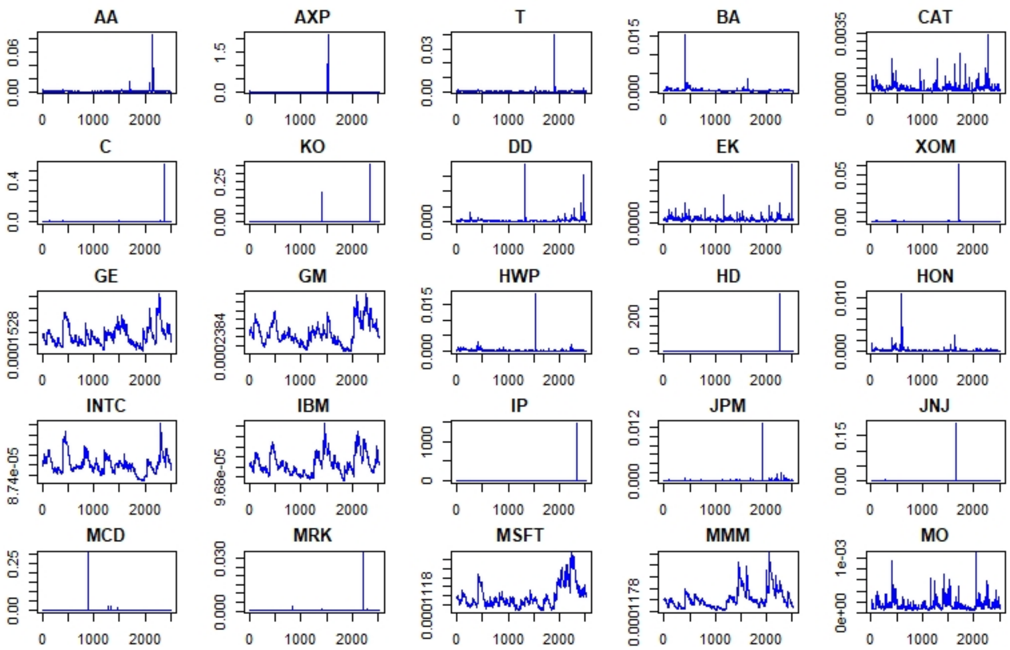

Figure B.15: Conditional variance estimated by the Gaussian-DCS(1,1) for the Dow Jones Index components' daily returns

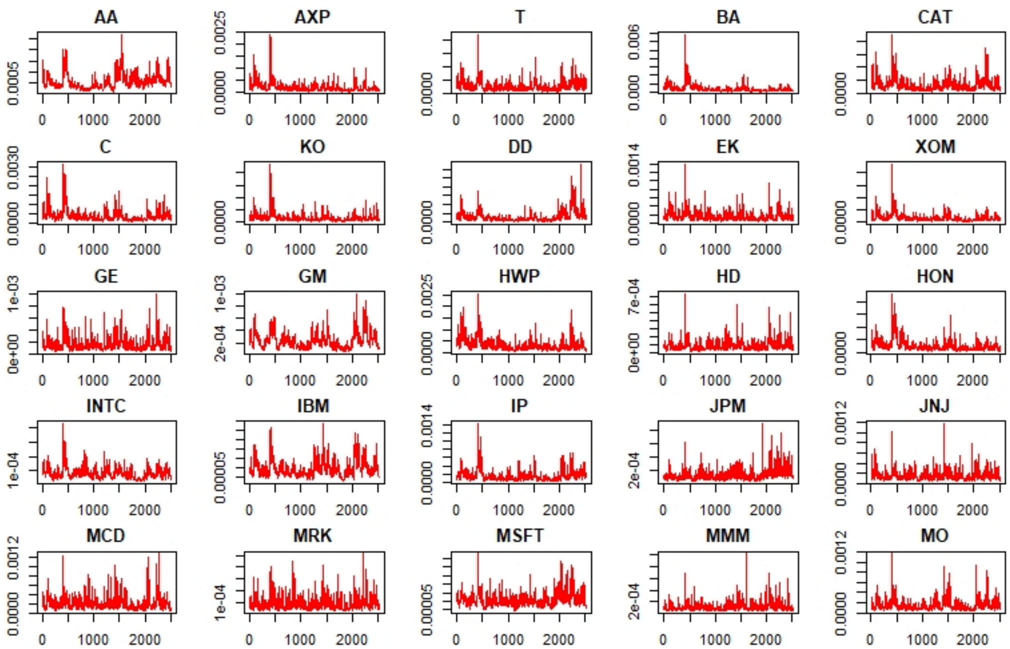

Figure B.16: Conditional variance estimated by the t-DCS(1,1) for the Dow Jones Index components' daily returns

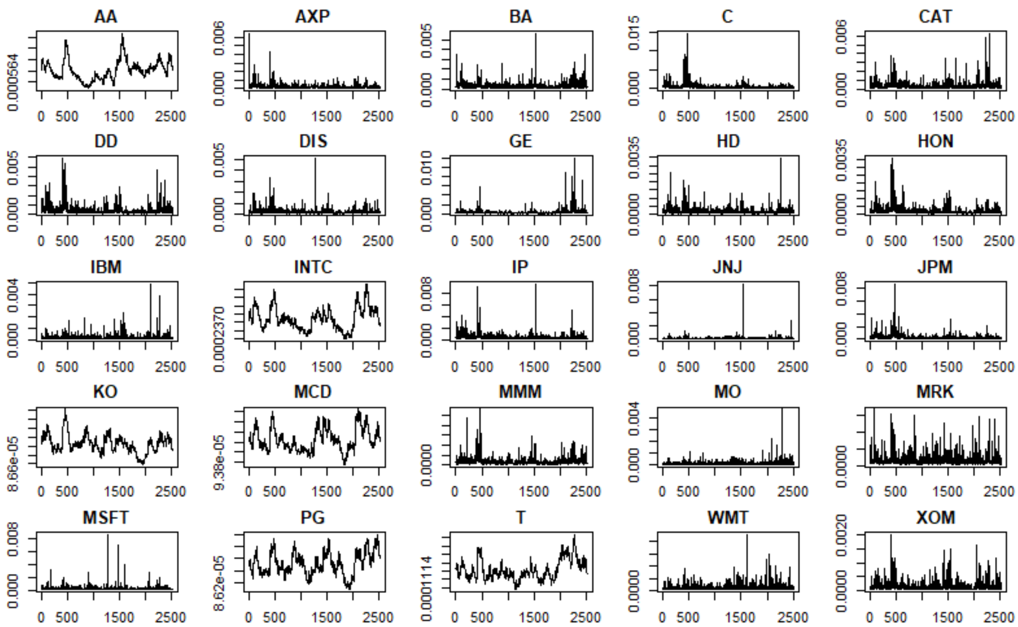

Figure B.17: Conditional variance estimated by the Skew-t-DCS(1,1) for the Dow Jones Index components' daily returns

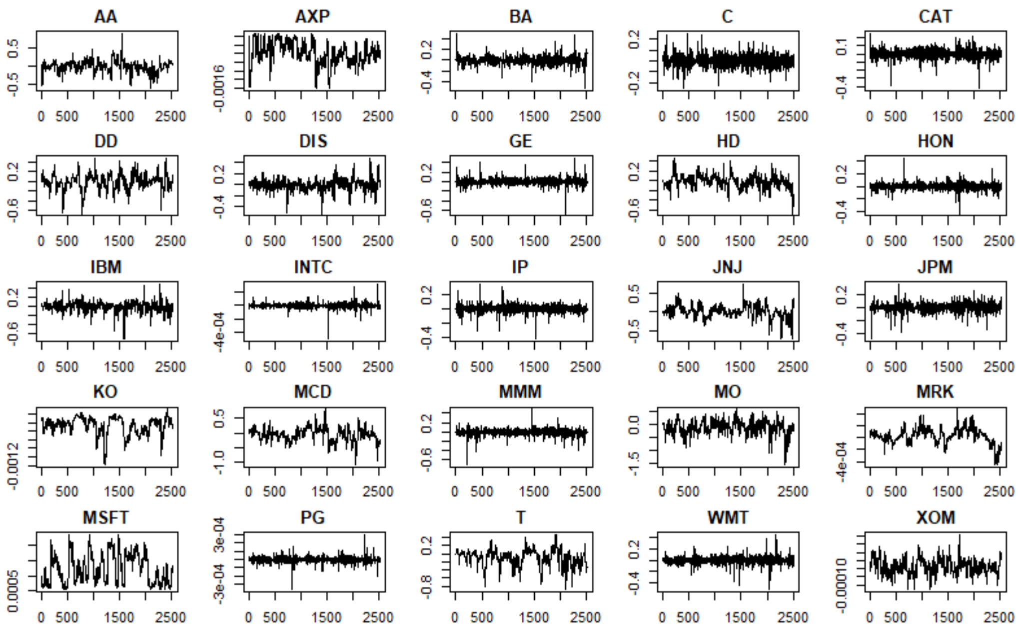

Figure B.18: Conditional skewness estimated by the Skew-t-DCS(1,1) for the Dow Jones Index components' daily returns

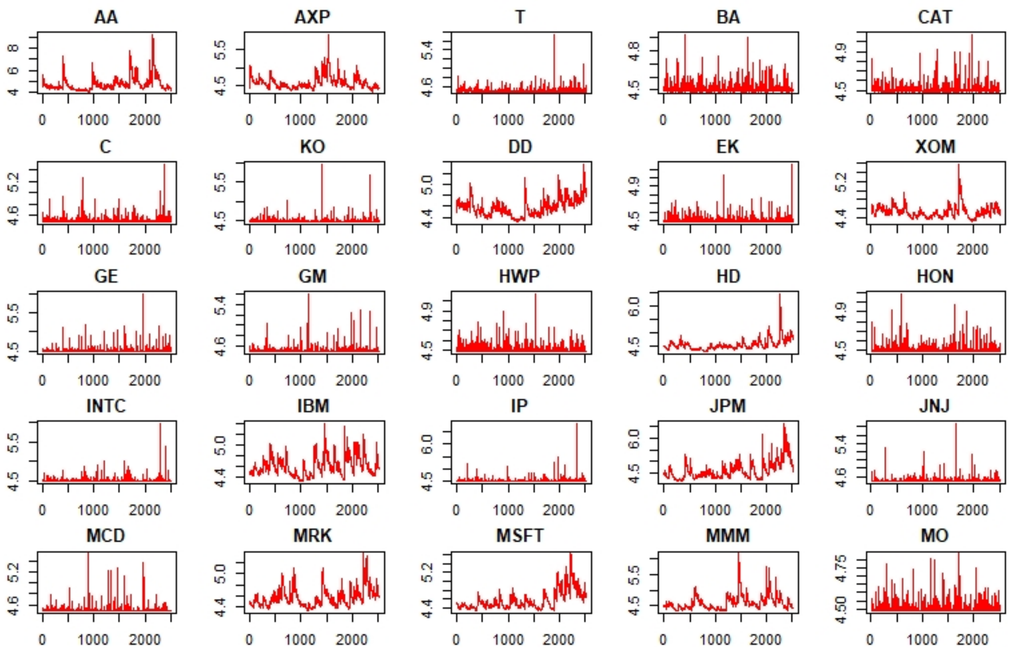

Figure B.19: Conditional shape estimated by the t-DCS(1,1) for the Dow Jones Index components' daily returns

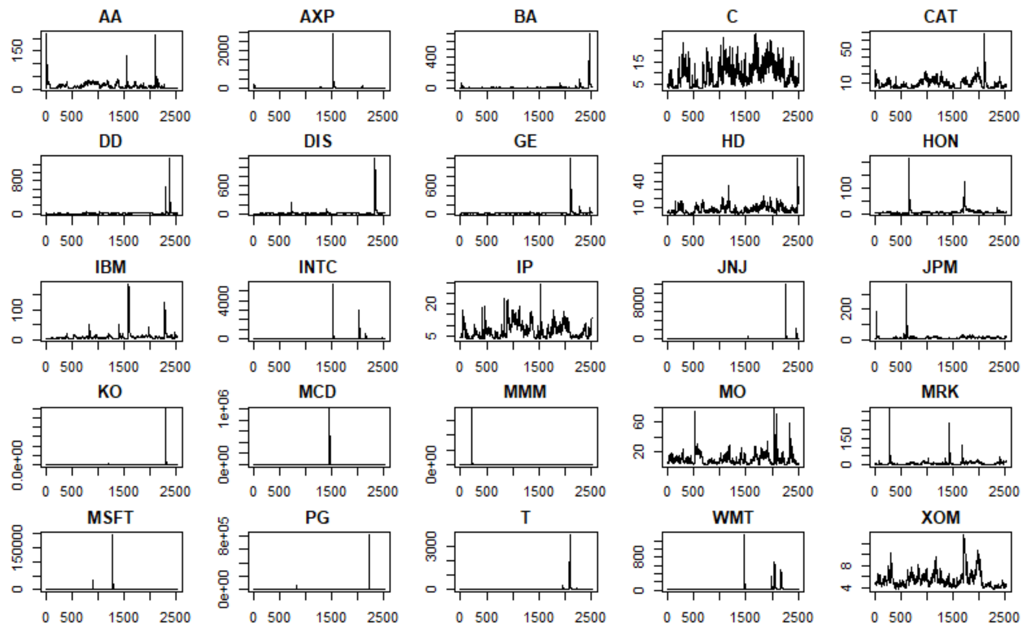

Figure B.20: Conditional shape estimated by the Skew-t-DCS(1,1) for the Dow Jones Index components' daily returns