Article

Distribution-based entropy weighting clustering of skewed and heavy tailed time series

Raffaele Mattera [1] [iD], **Massimiliano Giacalone** [2,*] [iD] **and Karina Gibert Oliveiras** [3] [iD]

1   Department of Economics and Statistics, University of Naples "Federico II", Italy; raffaele.mattera@unina.it
2   Department of Economics and Statistics, University of Naples "Federico II", Italy; massimiliano.giacalone@unina.it
3   Intelligent Data Science and Artificial Intelligence Research Center, Universitat Politecnica de Catalunya, Spain; karina.gibert@upc.edu
*   Correspondence: massimiliano.giacalone@unina.it

Abstract: The goal of clustering is to identify common structures in a data set by forming groups of homogeneous objects. The observed characteristics of many economic time series motivated the development of classes of distributions that can accommodate properties such as heavy tails and skewness. Thanks to its flexibility, the Skewed Exponential Power Distribution (also called Skewed Generalized Error Distribution) ensures a unified and general framework for clustering possibly skewed and heavy tailed time series. This paper develops a clustering procedure of model-based type, assuming that the time series are generated by the same underlying probability distribution but with different parameters. Moreover, we propose to optimally combine the estimated parameters to form the clusters with an entropy weighing *k*-means approach. The usefulness of the proposal is showed by means of application to financial time series, demonstrating also how the obtained clusters can be used to form portfolio of stocks.

1. Introduction

The goal of clustering is to identify common structures in a data set by forming groups of homogeneous data. This objective can be achieved by minimizing the within-group similarity and by maximizing the between-group dissimilarity.

Clustering of time series data is an important tool for data analysis in different areas ranging from engineering to finance and economics. For example, through clustering methods it is possible to build portfolios of similar stocks for financial applications (for example [1–3]). The main clustering approaches for time series can be summarized into three main groups [4]: observation-based, feature-based and model-based.

In the observation-based clustering the raw data are clustered according to a specified distance measure. Several authors proposed fuzzy extensions of common clustering algorithms for raw data (for example [5–9]). The time series involved could have either the same length or not. In the second case, it is common to take advantage of the Dynamic Time Warping (DTW) technique that is used to find an optimal alignment between two series with different lengths. (for example [9,10])

In the feature-based clustering the objects are clustered according to some of the data's features. The main advantage of this class of clustering approaches lies on the fact that the time series length is not an issue because objects with different length can be clustered together. Common time series features considered for clustering are the autocorrelation function (ACF) [11,12], the partial autocorrelation function (PACF) [13], the features of wavelet decomposition of the time series (for example [14,15]) or the cepstral (for example [16,17]).

The model-based clustering approaches assume, instead, that the time series are generated by the same statistical model (for example [18–21]) or that they have the same probability distribution (for example [22,23]). The spirit of most of the model-based clustering procedures is to group objects according to the estimated parameters. Important examples

are the clustering methods based on ARMA process distances (for example [18,19,24]), GARCH-based distances for heteroskedastic time series [19,20,25], estimates of the probability distributions' parameters (for example [22,23]) or, more recently, conditional higher moments (for example see [26]).

This paper develops a clustering procedure of the model-based type, assuming that the time series are generated by the same underlying probability distribution but with different parameters. Clearly, with this aim the specification of a very general distribution is required in order to account for a wide range of possible special cases.

Observed characteristics of many financial and economic time series motivated the development of family of distributions that are enough flexible to accommodate skeweess and heavy-tails, while nesting symmetric and bell-shaped distributions (e.g. the Normal) as special cases.

An important desired property of these classes is that maximum likelihood estimation of the parameters is possible. A class of asymmetric distributions with the desired properties of accommodating heavy tails and skewness is represented by the Skewed Exponential Power Distribution (SEPD) [27–30]. It generalizes the Exponential Power Distribution (also called Generalized Error Distribution, GED) for skewness.

Many financial applications of the GED as well as its skewed extensions have been considered (for example [29–38]). For example, [30] explored moments (also see [29]) as well as measures such as value at risk and expected shortfall that are useful in financial applications. Similarly, [37] proposed a GED-based Value at Risk model, while [38,39] studied the role of the Skewed GED in forecasting volatility.

In general, the Exponential Power Distribution, either symmetric or not, encompasses a very wide variety of special cases. Examples are the Gaussian, the skewed Normal, the Laplace, the skewed Laplace distribution and many others [37,40–42].

Therefore, in what follows we consider the Skewed Exponential Power Distribution family as the underlying assumption for all the considered time series. Thanks to its flexibility it ensures a unified and general framework for clustering possibly skewed time series.

The paper is structured as follows. In the next section, the entropy weighted clustering algorithm based on the Skewed Exponential Power Distribution is discussed. To show the usefulness of the proposed approach we provide two applications to different financial datasets in the Section 4. Then, in the Section 5 we propose to use the clusters obtained in the Section 4 to build portfolio of stocks. At the end some conclusions are offered.

2. The SEPD-based clustering approach

A very general and flexible family of distribution is represented by the Exponential Power Distribution (also the Generalized Error Distribution or the Exponential Power Function). The EPD random variable Z has the following probability density function [42,43]:

$$f(z) = \frac{\exp(-|\frac{z-\mu}{\sigma}|^p/p)}{2\sigma p^{\frac{1}{p}}\Gamma\left(1+\frac{1}{p}\right)} \tag{1}$$

where $z \in \mathbf{R}$, $\mu \in \mathbf{R}$ is called location parameter, $\sigma > 0$ is called scale parameter, $p > 0$ is a measure of fatness of tails and is called shape parameter (see [40]) and $\Gamma(\cdot)$ is the Gamma function. By construction, this distribution is symmetric and does not allow for skewness (Fig. 1).

Exponential Power Distribution
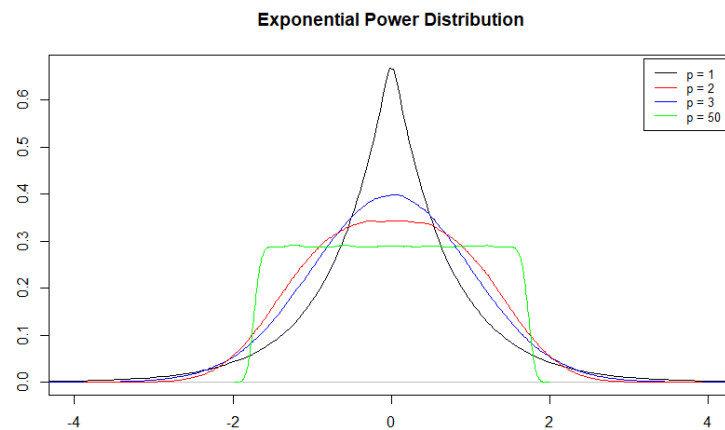
Figure 1. Exponential Power Distribution for different values of shape.

It is possible to write the EPD probability density (1) in more compact form by means of [40]:

$$f(z) = \frac{1}{\sigma} C \exp\left(-\frac{1}{p}\left|\frac{z-\mu}{\sigma}\right|^p\right); \quad C^{-1} = 2p^{1/p}\Gamma(1+1/p) \tag{2}$$

where C is a normalizing constant. The shape parameter p defines the heavy-taildness of the distribution. Hence, with a small value of p we obtain more flat distribution and vice-versa with a large p.

A very important feature of the EPD is that it includes many common distributions as special cases, depending by the value of shape parameter p (Fig. 1).

In particular, the Gaussian distribution is a special case when $p = 2$, and when $p < 2$ the distribution has fatter tails than a Gaussian distribution [37]. Moreover, when $p = 1$ we have a Laplace distribution, and for $p = +\infty$ we have the Uniform distribution [42].

Important contributions that extended the Exponential Power Distribution for skweness are represented by [27,28], where an additional skewness parameter, denoted λ in this paper, is introduced. (see Fig. 2).

Skewed Exponential Power Distribution
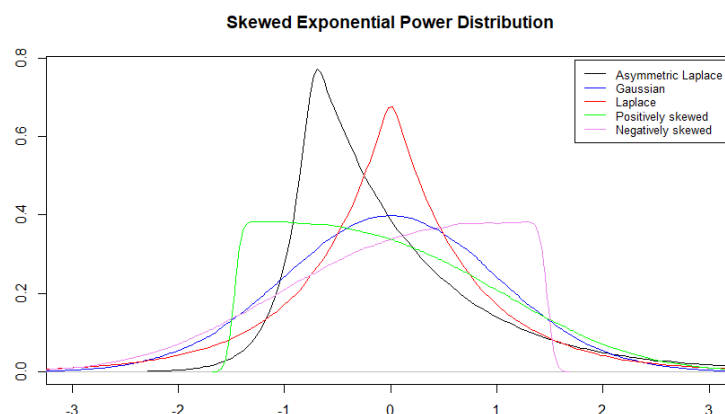
Figure 2. Skewed Exponential Power Distribution for different values of shape and skewness.

Some papers (for example [29,30,34,40]) constructed seemingly different classes of SEPD distributions. However, as suggested by [40], all of them are actually reparametrizations of the SEPD proposed by [27,28].

In this paper, following [34], we say that a random variable Z has a Skewed Exponential Power Distribution if its probability density function is the following:

$$f(z) = \frac{p}{\sigma \Gamma\left(1 + \frac{1}{p}\right)} \frac{\lambda}{1 + \lambda^2} \exp\left(-\frac{\lambda^p}{\sigma^p}[(z - \mu)^+]^p - \frac{1}{\sigma^p \lambda^p}[(z - \mu)^-]^p\right) \tag{3}$$

where:

$$(z - \mu)^+ = \max(z - \mu; 0) \quad \text{and} \quad (z - \mu)^- = \max(\mu - z; 0)$$

The parameters μ and σ correspond to location and scale, respectively, while λ controls skewness, and p is the shape parameter. For $\lambda = 1$, the distribution is symmetric about μ so we obtain the symmetric exponential power distribution. In the case $\lambda \neq 1$, by letting $p = 1$ we obtain the skewed Laplace distribution with density [34]:

$$f(z) = \frac{1}{\sigma} \frac{\lambda}{1 + \lambda^2} \begin{cases} \exp(-\frac{\lambda}{\sigma}|z - \mu|) & \text{for} \quad z \geq \mu, \\ \exp(-\frac{1}{\sigma\lambda}|z - \mu|) & \text{for} \quad z < \mu \end{cases} \tag{4}$$

For $p = 2$ and $\lambda \neq 1$, instead, we obtain the skewed normal distribution as defined in [44]. More details about the SEPD and the skewed Laplace distribution can be found in [34].

The great flexibility of the SEPD can be successfully exploited in the clustering process if the aim is to form distribution-based clusters. Distribution-based clustering could be of interest for a variety of applications (for example [22,23]).

In what follows, following in the spirit the contribution of [23], we propose a clustering algorithm that uses the estimated moments from the Skewed Exponential Power Distribution here introduced to form clusters. In other words, time series with similar estimated parameters are be placed in the same cluster. Moreover, since the underlying distribution has more than one parameter, following [7,45], we propose to optimally weight each parameter that represents a different feature of the data distribution.

The clustering model can be presented as follows. Let's assume to have $N(n = 1, \ldots, N)$ time series that are generated by a Skewed Exponential Power Distribution of parameters μ_n, σ_n, p_n and λ_n. We can store the estimated parameters in the following matrix:

$$\mathbf{X} = \begin{bmatrix} \mu_1 & \sigma_1 & p_1 & \lambda_1 \\ \mu_2 & \sigma_2 & p_2 & \lambda_2 \\ \vdots & \vdots & \vdots & \vdots \\ \mu_n & \sigma_n & p_n & \lambda_n \\ \vdots & \vdots & \vdots & \vdots \\ \mu_N & \sigma_N & p_N & \lambda_N \end{bmatrix} \tag{5}$$

that we can be used to compute the time series' dissimilarities.

As briefly stated before, since the SEPD has more than one parameter, a natural question is how would we use this information. Indeed, it is surely possible to cluster the time series only according to the location estimates or with respect to the scale parameter. Similarly, we can be interested in clustering time series with similar skewness or shape.

In this paper, we don't cluster the time series according to a single parameter but, instead, we aim to optimally combine them.

An useful approach for optimally weighting different features is represented by the weighted k-means (WKM) algorithm of [46]. The WKM algorithm proposes to incorporate a weighted distance function within the usual k-means algorithm. The main idea is that the weights are a measure of the relative importance of each feature with respect to the membership of the observations to a given cluster.

Formally, the Weighted k-Means algorithm (WKM) can be formalized as follows:

$$\min : \sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{m=1}^{M} u_{n,c} w_{m,c}^{\beta} D_{m,c}^{2} \tag{6}$$

under the constraints:

$$\sum_{c=1}^{C} u_{n,c} = 1, \quad u_{n,c} \geq 0, \tag{7}$$

$$\sum_{m=1}^{M} w_{m,c} = 1, \quad 0 \leq w_{m,c} \leq 1 \tag{8}$$

where $u_{n,c} \in \{0,1\}$ is binary and takes value of 1 if the n-th object belongs to the c-th cluster, $w_{m,c}$ represents the weight of the m-th feature in determining the c-th cluster and $D_{m,c} = d(x_{n,m}, x_{c,m})$, represents the (Euclidean) distance between the m-th feature of the n-th time series and the one of the c-th centroid.

Applied to the context of the distribution-based clustering, the weights $w_{m,c}$ are suitable values associated to each parameter m in the matrix **X** shown in (5) of the specified distribution within the c-th cluster.

Note that the weight $w_{m,c}$ is intrinsically associated with the squared distance $D_{n,c}^{2}$ for the specified distribution parameters. This makes possible to optimally weighting each distribution's feature in calculating the dissimilarities. Moreover, another appealing feature is that each c-th group has its own optimal weight vector.

Then, the exponent β has to be analyzed. With $\beta = 0$ we obtain the usual k-means clustering algorithm, while with a value of $\beta = 1$, we have that the weights associated to the feature with the smallest value of the weighted dissimilarity is equal to 1 and all the others $w_{m,c}$ are equal to zero.

When $\beta > 1$, the larger the D_m, the smaller the weight w_m. With a $\beta < 0$, we have that the larger D_m the larger the weight w_m. Then, if $0 < \beta < 1$ the larger the features' dissimilarity, the larger is the weight w_m and this is against the variable weighting principal [46].

Therefore we cannot choose $0 < \beta < 1$, $\beta = 0$ or $\beta = 1$ but in the WKM algorithm suitable values are $\beta < 0$ or $\beta > 1$.

However the exponent β is an artificial device, lacking a strong theoretical justification [7]. Note that the value of β in the formula (6) is similar to the fuzzines parameter in the fuzzy c-means algorithm. To overcome this problem, the usage of a regularization term has been proposed [7,45]. In this case, the burden represented by β is shifted to the regularization term obtaining, in such a way, a factor that multiplies the regularization contribution to the clusters formation.

With this respect, [45] proposed a clustering algorithm where the weight of a given feature in a cluster represents the relevance of each feature in determining the clusters.

Therefore, [45] modified the objective function (6) by adding the weight entropy term such that, at the same time, we minimize the within cluster dispersion and maximize the negative weight entropy. Hence, we force more features to contribute in the formation of the groups [47].

The new objective function can be written as follows:

$$\min : \sum_{c=1}^{C} \left[\sum_{n=1}^{N} \sum_{m=1}^{M} u_{n,c} w_{n,m} D_{m,c}^{2} + \gamma \sum_{m=1}^{M} w_{n,m} \log(w_{n,m}) \right] \tag{9}$$

subject to the constraints:

$$\sum_{c=1}^{C} u_{n,c} = 1, \quad u_{n,c} \geq 0, \tag{10}$$

$$\sum_{m=1}^{M} w_{m,c} = 1, \quad 0 \le w_{m,c} \le 1 \tag{11}$$

where $u_{n,c} \in \{0,1\}$ is binary, if a hard clustering procedure is developed, and takes value of 1 if the n-th object belongs to the c-th cluster, $w_{m,c}$ represents the weight of the m-th feature in determining the c-th cluster and $D_{m,c} = d(x_{n,m}, x_{c,m})$, represents the (Euclidean) distance between the m-th feature in the matrix $\mathbf{X}$ shown in (5) of the n-th time series and the one of the c-th centroid.

The first term in (9) is the sum of the within cluster dispersion, while the other one is the negative weight entropy. The positive parameter γ controls for the size of the weights, meaning that with γ we decide the degree of discrimination between the features [45].

The algorithm works as follows. An initial set of k means are identified as the starting centroids. An initial cluster is defined considering that the observations are clustered to the nearest centroid according to the Euclidean distance measure among distribution parameter estimates (5). The centroids are identified based on these clusters, while the weights are computed for each time series in any given cluster. Then, we compute the new centroids and, by using an updated weighted distance, each time series is clustered to its nearest new centroid. This steps are repeated until the algorithm converges.

In the case of Skewed Exponential Power Distribution, the optimal weights of the SEPD-DWEKM model, obtained by the solution of the optimization problem (9), are equal to:

$$w_{m',c} = \frac{\exp(\frac{-D_{m',c}}{\gamma})}{\sum_{m=1}^{4} \exp(\frac{-D_{m,c}}{\gamma})} \tag{12}$$

The proof of (12) can easily be derived by following [46]. Similarly to the standard k-means algorithm $u_{n,c}$ is updated as follows:

$$\begin{cases} u_{n,c} = 1 & \text{if} \quad \sum_{m=1}^{4} w_{m,c} D_{m,c}^2 \le \sum_{m=1}^{4} w_{m,c'} D_{m,c'}^2 \\ u_{n,c} = 0 & \text{otherwise} \end{cases}$$

where $u_{n,c} = 1$ means that the n-th object is assigned to the c-th cluster, so we have an hard, not fuzzy, final assignment. If a time series is equidistant from two clusters, we assign it to the one with the smallest index.

From (12) we understand the role played by the parameter γ, that is used to control for the size of the weights. Indeed, if $\gamma > 0$, the weights $w_{m,c}$ are inversely proportional to squared distance $D_{m,c}^2$. Therefore, the smaller $D_{m,c}^2$, the larger the weights $w_{m,c}$ and, hence, the more important the corresponding dimension m. Instead, if $\gamma < 0$, the weights $w_{m,c}$ is proportional to the distance $D_{m,c}^2$. Therefore, the larger is the distance the larger is the associated weight. This is a contradictory result and, hence, γ cannot be smaller than zero. In the end, γ can be set equal to zero. In this case, the dimension m' with the smallest distance has a weight equal to 1, $w_{m',c} = 1$, while all the others are zero $w_{m,c} = 0$. Therefore, each cluster contains only one important dimension.

A final crucial aspect of the any clustering procedure is the selection of the number of clusters (C). With this respect we compute the Silhouette Width Criterion (SWC) of [48]. Clearly, the best partition is expected to be pointed out when the SWC is maximized, which implies the minimization of the intra-group distance the maximization of the inter-group distance.

3. Application to financial time series

To show the effectiveness of the proposed clustering approach, in what follows we provide an application to stock market data. The role of skewness and kurtosis in modeling financial data is well documented (for a review see [49]).

Therefore, financial market data represent a clear example of the possible application

since the empirical densities of the financial time series are proven to be non-Gaussian, asymmetric and heavy tailed [50][1].

In what follows we provide empirical applications of the proposed clustering approach to two different financial datasets. In the first expertiment, we consider the FTSE100's stocks, while in the second we consider the industrial sector's stocks belonging to the S&P500 index.

3.1. FTSE100 stocks

The first application with real data aims to cluster the stocks belonging to the FTSE100 index. At this aim we consider the daily stock returns over the last 10 years, from the 1-st January 2011 to the 1-st January 2021 (Fig. 3).
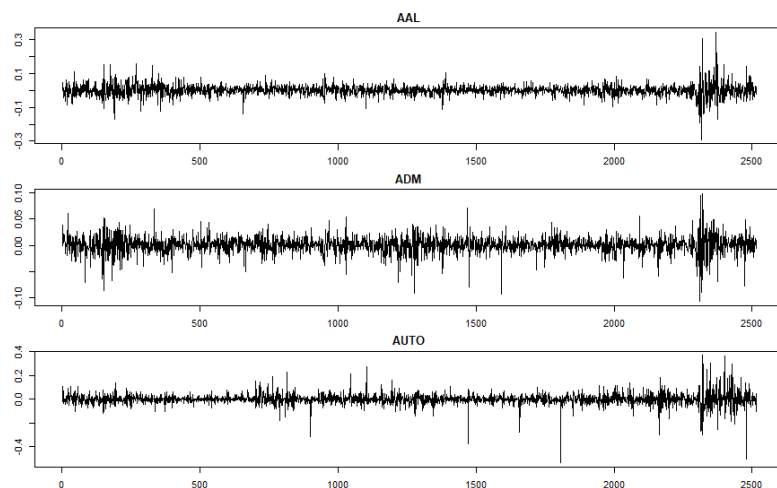
Figure 3. Sample of stock returns time series included in the dastaset under consideration (FTSE100 data).

In particular, over the 100 stocks we selected those without missing values within the considered sampling period, hence getting as result $N = 25$ stocks. The list of the stocks included in the sample is shown in Tab. 12 in the Appendix A.

To empirically motivate the peculiar distributional characteristics of the stock returns included in the sample, we show some estimated empirical densities (Fig. 4).

---

[1]   We have to highlight that this statement is not always true. For example, it is known that most of monthly stock indices show a Gaussian distribution. However, it is similarly accepted that daily stock returns are not normally, heavy tailed and asymmetric distributed. Therefore, in this paper we deal with daily returns data.
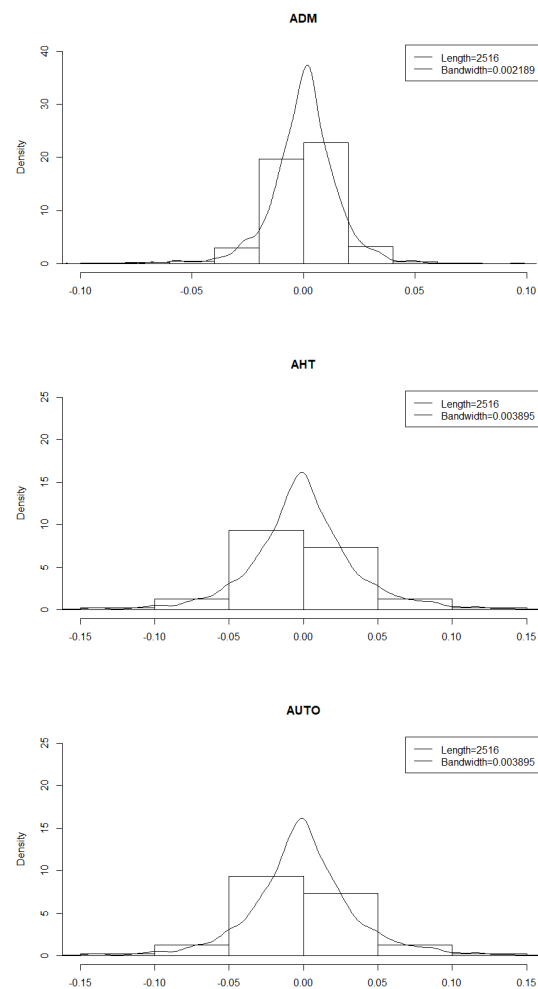
Figure 4. Empirical densities of the stocks shown in Fig. 3.

Moreover, in the Tab. 1 we report the sample estimators for mean, standard deviation, skewness and kurtosis as well as the Jarque-Bera [51] normality test. The results of the conducted normality tests suggest to reject the null hypothesis of normal distribution for all the stocks (see JB test column of Tab. 1). Accordingly, it can be highlighted that any stock shows a symmetric distribution and the majority of them are negatively skewed. Furthermore, the stocks show very high leptokurtic distributions with fatter tails than the Gaussian. Indeed, within the sample only one stock shows a kurtosis lower than 3 (i.e. IAG) while all the others have much highier values.

|	Mean	St. Dev.	Skewness	Kurtosis	JB test	Length
AAL	0.0002	0.0326	0.4403	13.4706	19141.2841***	2516
ADM	0.0003	0.0161	-0.5162	5.4498	3233.0866***	2516
AHT	-0.0010	0.0508	6.8974	177.1057	3313541.0732***	2516
AUTO	-0.0002	0.0488	-0.7128	20.1631	42911.4151***	2516
AZN	0.0005	0.0151	-0.4973	13.6616	19707.8758***	2516
BHP	0.0000	0.0208	-0.3470	5.9551	3777.1783***	2516
BME	0.0006	0.0143	0.1503	10.5756	11758.3256***	2516
CPG	-0.0011	0.0344	-1.7316	41.2409	179865.2429***	2516
CRH	0.0004	0.0204	-0.7531	9.0937	8925.6488***	2516
DGE	0.0003	0.0117	-0.7707	8.7590	8309.3523***	2516
EVR	0.0005	0.0229	-0.1605	10.2988	11152.9051***	2516
FERG	0.0006	0.0189	0.0795	64.6506	438904.5890***	2516
GSK	0.0002	0.0122	-0.6519	8.6100	7966.7005***	2516
IAG	-0.0006	0.0383	-0.0550	2.7651	805.4615***	2516
IHG	0.0005	0.0192	-0.6035	15.9591	26903.8564***	2516
III	0.0002	0.0314	-0.9284	20.7329	45506.6850***	2516
MNG	-0.0003	0.0195	-0.3442	9.4997	9530.2226***	2516
MRO	-0.0004	0.0319	-2.7604	64.0144	433505.1309***	2516
NWG	-0.0003	0.0272	-0.9661	11.1644	13485.1878***	2516
PRU	0.0002	0.0213	-0.8713	16.1174	27602.6225***	2516
RIO	0.0002	0.0218	0.0577	3.8344	1547.1291***	2516
SVT	0.0001	0.0261	0.7359	15.7558	26301.1559***	2516
TSCO	0.0007	0.0185	-0.1470	11.6275	14210.8902***	2516
VOD	0.0000	0.0160	-0.4344	11.4206	13780.2067***	2516
WPP	0.0001	0.0194	-1.7115	16.9938	31561.2513***	2516

Note: *** means significance at 1% confidence level.

Table 1: Descriptive statistics and normality test of [51] for the FTSE100 stocks.

Therefore, for clustering time series with similar distributions we use the approach based on the Skewed Exponential Power Distribution presented in the previous Section. The first step of the clustering procedure requires the estimation of the SEPD's parameters. Then, the number of clusters has to be choosen.
At this aim, we consider the average silhouette width criterion (SWC). In the Fig. 5 is reported the final result.
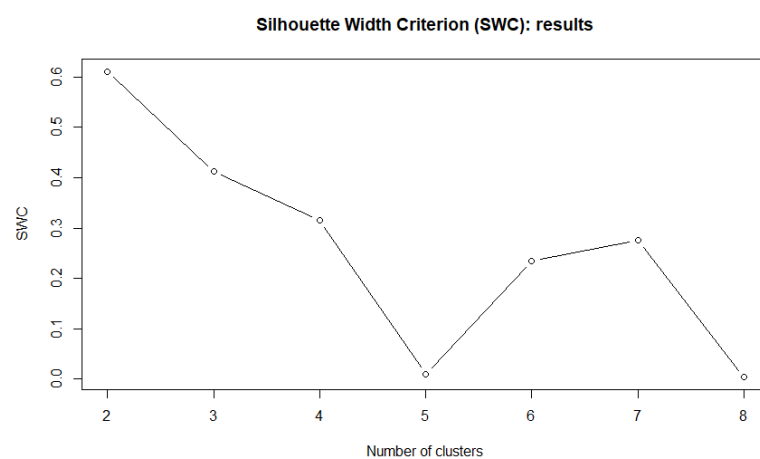
Figure 5. Silhouette Width Criterion for different number of clusters *C* (distribution-based clustering) – experiment with FTSE100 stocks.

Accordingly, the parameters estimated by Maximum Likelihood (MLE)[2] as well as the final clustering results are reported in the Tab. 2.

From the Tab. 2 is evident that the second cluster contains the majority of the stocks. Moreover, the two groups mainly differentiate each other in terms of their shapes. Indeed, in the second cluster we have the stocks characterized by the lowest shape parameters p and by a skewness parameter λ always greater than 1. In general, sorting by shape is in this case more informative than sorting by the degree of skewness that, however, still reveals important information about the distribution of the stocks placed within each group.

Moreover, some additional comments about data heterogeneity within each cluster can be provided by looking at the Tab. 2. Indeed, the second cluster seems to be the one with the highest degree of heterogeneity. To see why, we can look at the column of the estimated skewness in Tab. 2. While in the first cluster we have all values of λ close to 1, in the case of cluster 2 the values range from $\lambda = 0.88$ to $\lambda = 1.03$. A similar discussion can be provided for the shape values p, since in the cluster 1 all the stocks have low shape's parameters p.

	Location	Scale	Shape	Skewness	Cluster
AHT	0.000938	0.038008	0.634380	1.020410	1
FERG	0.000396	0.003676	0.737628	1.087450	1
III	0.000215	0.030700	0.802500	1.005290	1
SVT	0.005273	0.024976	0.571017	1.206776	1
AAL	0.000173	0.031175	0.970264	0.978047	2
ADM	0.000326	0.015780	0.980135	0.923800	2
AUTO	-0.007066	0.043761	0.840452	0.885689	2
AZN	0.000486	0.014322	1.005312	0.994350	2
BHP	0.000069	0.020380	1.131097	0.928945	2
BME	0.000587	0.013721	0.980685	1.002382	2
CPG	-0.001029	0.032234	0.879535	0.976905	2
CRH	0.000370	0.019841	1.089496	0.981539	2
DGE	0.000287	0.011343	1.076945	0.925662	2
EVR	0.000544	0.022103	1.022613	0.964463	2
GSK	0.000174	0.011886	1.067266	0.966571	2
IAG	-0.000660	0.038030	1.171752	1.034922	2
IHG	0.000484	0.018176	0.883195	0.953825	2
MNG	-0.000304	0.018965	1.066270	0.969299	2
MRO	-0.000229	0.029521	0.937861	0.975776	2
NWG	-0.000333	0.026422	1.009514	0.946825	2
PRU	0.000284	0.020122	0.912030	0.975315	2
RIO	0.000231	0.021566	1.139757	0.987474	2
TSCO	0.000734	0.017782	1.042405	0.981452	2
VOD	0.000242	0.015358	0.999362	1.011196	2
WPP	0.000128	0.018252	0.952139	0.951217	2

Table 2: MLE parameters estimation from a SEPD and assigned clusters according to the EWKM – FTSE100 data.

In general, the weights obtained by means of the entropy weighted k-means algorithm (EKWM) reflect, as discussed in the previous Section, this heterogeneity. Indeed, the weights are inversely proportional to squared distances such that to small distances are associated larger weights.

The Tab. 3 shows the optimal weights computed with respect the selected $C = 2$ clusters. According to the arguments presented so far, the weights effectively reflect the degree of heterogeneity of the features. Indeed, in the cluster 2 the shape's weight w_p is the lowest

---

[2] We use the software R to obtain the parameter estimates. More in details, the function *nlminb* is used in order to maximize the log-likelihood function of the SEPD. As starting values for the function we use the sample estimates for location and scale parameters, while we set $p = 2$ and $\lambda = 1$ (symmetric distribution) for shape and skweness, respectively, such that the starting values correspond to the normal distribution.

one since the distances in terms of shape parameters in the second cluster are higher than the same shape-based distances in the first cluster.

In the case of other parameters (i.e. location, scale and skewness) the weights assigned in the two groups are very similar. In other words, the Tab. 3 highlights that the two clusters mainly differentiate each other because of the distribution's shape.

|	Location	Scale	Shape	Skewness
Cluster 1	0.253624	0.253463	0.245611	0.247302
Cluster 2	0.261146	0.260772	0.222667	0.255415

Table 3: Distribution-based EWKM for FTSE100 stocks: resulting weights

However, one can ask whether a distribution-based clustering approach for time series is more convenient than other common approaches available. Clearly there is not an easy answer to this question since the usefulness of a clustering approach depends by its aim and by the researcher's goal.

However, in what follows we provide an in-sample comparison of a well-established clustering approach for financial time series based on the stock returns correlations (e.g. see [1]). In particular, assuming a k-medoids approach, we cluster the time series according to the following correlation-based distance:

$$d_{n,j} = \sqrt{2(1 - \rho_{n,j})} \tag{13}$$

that depends by the correlation $\rho_{n,j}$ between the n-th stock returns $r_{n,t}$ and the j-th returns $r_{j,t}$. In the Fig. 6 is reported the SWC criterion for different clusters C. The number of clusters with highest validity are $C = 7$.
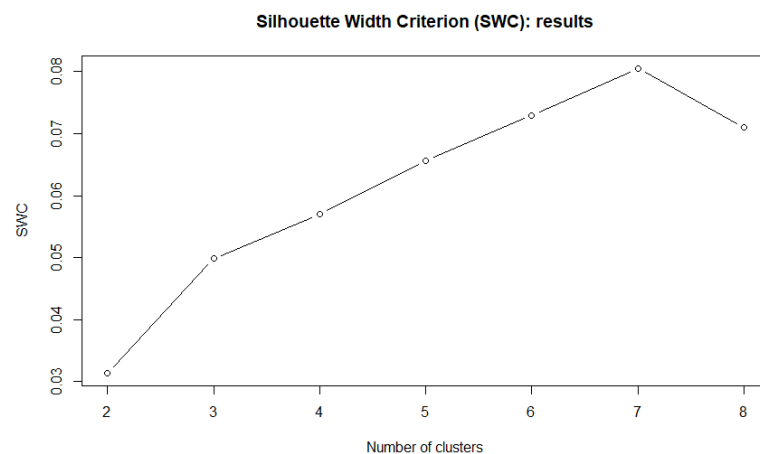
Figure 6. Silhouette Width Criterion for different number of clusters (correlation based clustering) – experiment with the FTSE100 stocks.

However the highest SWC is equal to 0,08 and is dramatically lower than the SWC value in Fig. 5 that is equal to 0,6. The differences between the two classifications are showed in the Tab. 4.

In general, according to the SWC criterion, we can argue that the clusters obtained by means of the distribution-based approach are much more accurate than those obtained with a correlation-based approach, that is well established in finance.

	SEPD-based clustering	Correlation-based clustering
AAL	2	1
ADM	2	1
AHT	1	1
AUTO	2	2
AZN	2	3
BHP	2	4
BME	2	3
CPG	2	5
CRH	2	1
DGE	2	3
EVR	2	1
FERG	1	6
GSK	2	3
IAG	2	4
IHG	2	1
III	1	1
MNG	2	5
MRO	2	5
NWG	2	1
PRU	2	1
RIO	2	4
SVT	1	7
TSCO	2	3
VOD	2	3
WPP	2	1

Table 4: Differences in the classification between the entropy weighted distribution-based and the correlation-based clustering approaches – FTSE100 data.

3.2. S&P500 stocks: industrial sector

As additional experiment we also select the stock prices of the companies belonging to the industrial sector that are included in the S&P500 Index. More in details, we downloaded the last 10 years of daily observations for all the 74 stocks quoted, specifically from the 1-st January 2011 to 1-st January 2021.

The considered stocks have different lengths because some of them have been quoted later. Differently from the previous expertiment, we now decide to consider in the sample also the stocks with different lengths, thus containing missing values.

Indeed, as the proposed approach is of model-based type, we are able to cluster two time series with different length as far they share a similar distribution. Indeed, in the sample there are also stocks with a length $T = 200$ as in the case of CARR and OTIS.

The entire list of the stocks considered in the sample, with their length, is shown in the Tab. 13. Particularly, for each time series we consider the logarithmic returns (Fig. 7).
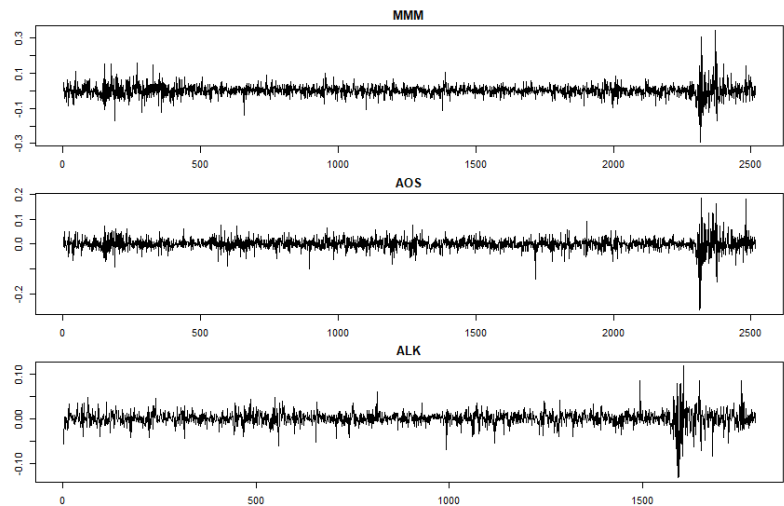
Figure 7. Sample of stock returns time series included in the dastaset under consideration (S&P500 data).

As in the previous experiment, in order to empirically show the aforementioned stock returns characteristics (i.e. heavy tails and skewness) in the Fig. 8 are reported the empirical densities for the sample of stock returns showed also in the Fig. 7.
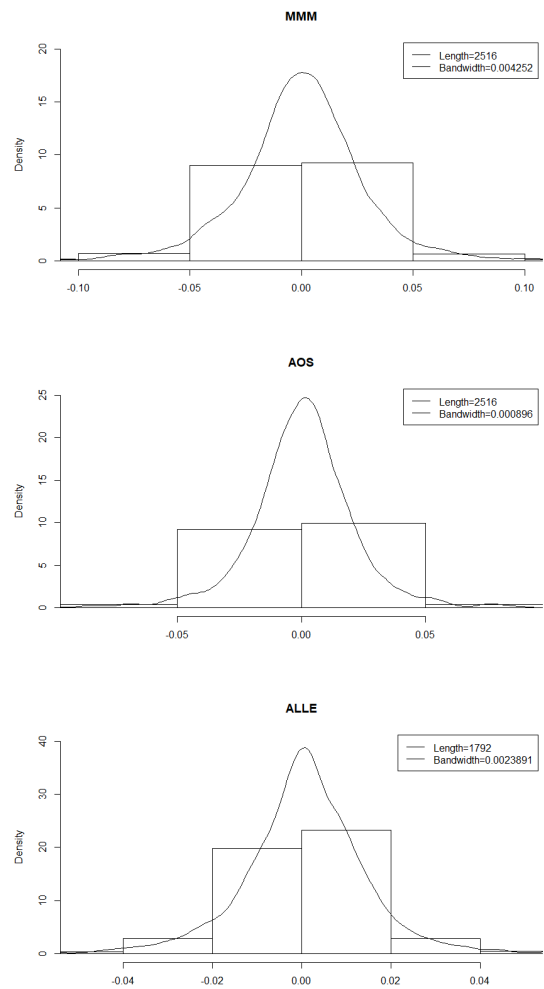
Figure 8. Empirical densities of the stocks shown in Fig. 7.

| Stock	Mean	St. Dev.	Skewness	Kurtosis	JB test	Length
MMM	0.0002	0.0326	0.4403	13.4706	19141.2841***	2517
AOS	0.0006	0.0242	-0.8093	17.4330	32194.0594***	2517
ALK	0.0005	0.0164	-0.5533	9.9428	7494.6041***	1793
ALLE	0.0006	0.0163	-0.7401	16.1544	27639.5014***	2517
AAL	0.0008	0.0169	-0.0805	3.4418	1248.2562***	2517
AME	0.0006	0.0227	-0.6764	27.5441	79867.2387***	2517
BA	0.0058	0.0393	0.9665	7.8917	562.3908***	200
CHRW	0.0004	0.0183	-0.4986	5.0412	2775.3550***	2517
CARR	0.0001	0.0154	-1.3744	11.4410	14543.0106***	2517
CAT	0.0004	0.0186	-0.0322	8.3402	7308.2827***	2517
CTAS	0.0010	0.0159	-0.3440	14.9755	23605.0368***	2517
CPRT	0.0006	0.0183	0.1817	14.9407	23459.7916***	2517
CSX	0.0011	0.0162	-0.5401	19.4430	39825.4771***	2517
CMI	0.0005	0.0260	-0.7426	14.9902	23833.2396***	2517
DE	0.0006	0.0172	-0.1867	8.2728	7204.9810***	2517
DAL	0.0005	0.0173	-0.6308	10.5860	11939.0318***	2517
DOV	0.0007	0.0159	-1.2111	15.0746	24484.0007***	2517
ETN	0.0003	0.0173	-0.9108	17.9497	34187.4551***	2517
EMR	0.0004	0.0181	0.0802	12.4383	16253.7474***	2517
EFX	0.0003	0.0150	-0.3772	6.7968	4913.7515***	2517
EXPD	0.0006	0.0173	0.1549	7.6491	6157.3116***	2517
FAST	0.0009	0.0203	-0.4309	8.8030	7638.9534***	2339
FDX	0.0004	0.0182	-0.6436	10.6534	12096.2927***	2517
FTV	0.0005	0.0178	-0.3642	17.6651	14804.7770***	1133
FBHS	0.0004	0.0144	-0.4477	6.3062	4262.9559***	2517
GNRC	-0.0001	0.0202	-0.0884	8.9715	8458.9914***	2517
GD	0.0012	0.0241	0.5907	10.1009	10864.4486***	2517
GE	0.0005	0.0173	0.0860	13.3704	18780.3863***	2517
GWW	0.0007	0.0175	-0.4450	6.4335	4337.4524***	2463
HON	0.0006	0.0148	-0.1974	11.3107	13454.9980***	2517
HWM	-0.0001	0.0264	-0.4662	11.5068	13999.6075***	2517
HII	0.0007	0.0150	-0.5132	6.5607	4633.3047***	2517
IEX	0.0007	0.0160	1.9067	46.0650	146908.9322***	1647
INFO	0.0008	0.0251	-0.3388	6.9870	1892.4783***	917
ITW	0.0006	0.0150	-0.1942	11.5515	14032.6104***	2517
IR	0.0003	0.0180	-0.2123	5.1921	2852.1589***	2517
JBHT	0.0005	0.0158	-0.2285	8.1095	6931.2264***	2517
J	0.0004	0.0158	-0.4979	7.3329	5753.7195***	2517
JCI	0.0006	0.0196	-0.5279	12.0699	15419.7615***	2517
KSU	0.0006	0.0179	-1.2609	19.1160	39046.5207***	2517
LHX	0.0007	0.0161	-0.3588	10.6848	12046.8015***	2517
LDOS	0.0008	0.0134	-0.6141	14.4475	22082.3240***	2517
LMT	0.0005	0.0208	-0.4192	7.3590	5763.8034***	2517
MAS	0.0007	0.0207	-0.2718	5.2009	2873.8198***	2517
NLSN	0.0004	0.0137	-0.8574	11.6144	14478.3288***	2517
NSC	0.0001	0.0198	-1.6786	27.7429	81459.1085***	2500
NOC	0.0007	0.0143	-0.1255	8.0477	6811.0182***	2517
ODFL	0.0006	0.0175	-0.1828	10.7789	12218.9127***	2517
OTIS	0.0010	0.0181	-0.0697	4.9432	2570.1778***	2517
PCAR	0.0021	0.0255	0.3171	5.3143	245.0319***	200
PH	0.0003	0.0167	-0.1098	5.1097	2749.0581***	2517
PNR	0.0005	0.0195	-0.5206	12.0102	15265.5206***	2517
PWR	0.0004	0.0181	-0.5505	13.2218	18489.7124***	2517
RTX	0.0005	0.0207	-1.9917	34.1081	123835.8960***	2517
RSG	0.0004	0.0196	-0.1265	9.3393	9169.7468***	2517
RHI	0.0006	0.0189	-0.2785	11.0507	12860.5887***	2517
ROK	0.0008	0.0153	-0.4364	8.5542	7767.5581***	2517
ROL	0.0007	0.0148	-0.5568	8.8381	8336.3830***	2517
ROP	0.0006	0.0121	-1.7053	22.4745	54267.9783***	2517
SNA	0.0002	0.0164	-0.4683	16.5103	28722.5064***	2517
LUV	0.0005	0.0167	-0.1462	7.3703	5716.4328***	2517
SWK	0.0005	0.0198	-0.8209	22.7150	54471.1200***	2517
TDY	0.0011	0.0205	-0.6358	26.5648	74280.1975***	2517
TXT	0.0009	0.0173	-1.4434	25.9249	71458.1550***	2517
TT	0.0007	0.0175	-0.4868	6.2835	4248.3077***	2517
TDG	0.0003	0.0223	-0.3101	10.5591	11752.7020***	2517
UNP	0.0002	0.0307	-0.7200	16.6715	29409.5251***	2517
UAL	0.0007	0.0161	-0.4693	8.5580	7786.9958***	2517
UPS	0.0005	0.0137	0.0987	12.3985	16151.3421***	2517
URI	0.0009	0.0296	-0.5009	6.2084	4155.5368***	2517
VRSK	0.0007	0.0137	-0.1206	13.4321	18957.1646***	2517
WM	0.0004	0.0200	-0.5989	9.3161	9268.1082***	2517
WAB	0.0006	0.0120	-0.6692	14.2371	21478.1235***	2517
XYL	0.0007	0.0166	-0.1198	8.1649	6462.2884***	2320

Note: *** means significance at 1% confidence level.

Table 5: Descriptive statistics and normality test of [51] for the S&P500 stocks.

From Fig. 8 it is possible to note that the considered time series show very different distributions as well as a strong deviation from Gaussianity. Moreover, we also report in the Tab. 5 the main descriptive statistics as well the [51] test of normality. In general, from these simple considerations appear clearly the need for the specification of a very flexible distribution able to accurately capture these differences.

As previously described, the first step of the proposed clustering procedure involves the estimation of the Skewed Exponential Power Distribution parameters (i.e. location, scale, skewness and shape) by means of maximum likelihood method. Then, as usual, the second step of the procedure involves the decision about the number of clusters C.

As previously specified, we take advantage of the Silhouette Width Cirterion (SWC), whose results are shown in the Fig. 9.
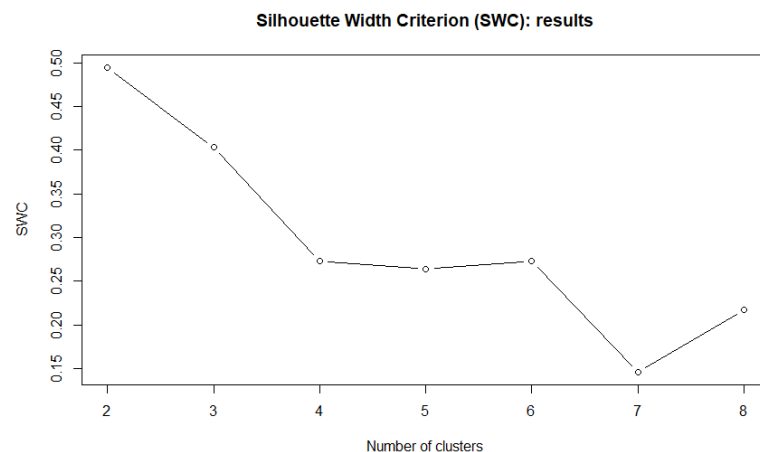
Figure 9. Silhouette Width Criterion for different number of clusters C (distribution-based clustering) – S&P500 stocks.

The highest value of the silhouette is obtained with $C = 2$ clusters. Then, from the distribution(SEPD)-based Entropy Weighting k-Means (SEPD-EWKM) algorithm we obtain the hard partition showed in the Tab. 6.

As in the previous experiment, the two resulting clusters are not balanced since the second cluster contains most of the stocks in the sample. Moreover, it appears clearly that the two clusters differentiate each other in terms of shape. Indeed, the first cluster contains all the stocks with shape parameter p lower than $p = 0.9$, while on the other side in the second one we have all the stocks with higher shape's parameters.

However, also the skweness λ allows a remarkable distinction among the two clusters since in the first group we find most of the stocks with $\lambda \geq 1$ while in the second one the stocks with lower degree of skewness. Nevertheless, the heterogeneity in terms of skewness in the first cluster appear considerable.

	Location μ	Scale σ	Shape p	Skeweness λ	Cluster
BA	0.000529	0.020121	0.809235	0.961483	1
CARR	0.005634	0.037575	0.869097	1.022203	1
CTAS	0.001057	0.014841	0.828099	0.972663	1
EFX	0.000743	0.015000	0.910557	0.976399	1
FTV	0.000516	0.016727	0.880068	1.010577	1
GE	-0.000323	0.019545	0.819186	0.987702	1
GNRC	0.001210	0.023105	0.909975	1.013188	1
HON	0.000665	0.014166	0.898461	0.999329	1
INFO	0.000744	0.014293	0.862185	1.000896	1
LDOS	0.000719	0.016602	0.907270	0.955210	1
MMM	0.000428	0.013126	0.894845	0.965327	1
NLSN	0.000161	0.018310	0.886294	1.006444	1
OTIS	0.003455	0.024927	0.768527	1.186048	1
RTX	0.000246	0.015165	0.830205	0.979384	1
SWK	-0.000646	0.018981	0.798267	0.880760	1
TDG	0.001156	0.018331	0.820059	0.992412	1
TXT	0.000295	0.021221	0.865774	1.001103	1
UAL	0.000242	0.028890	0.904769	1.006104	1
VRSK	0.000705	0.012904	0.881891	0.973572	1
WM	0.000407	0.011229	0.862004	0.943475	1
AAL	0.000173	0.031175	0.970261	0.978034	2
ALK	0.000561	0.022819	0.941247	0.978997	2
ALLE	0.000522	0.015682	0.980738	0.931036	2
AME	0.000644	0.015499	0.929582	0.973628	2
AOS	0.000757	0.016787	1.040460	0.949007	2
CAT	0.000356	0.018032	1.043622	0.999351	2
CHRW	0.000153	0.014709	0.959506	0.930355	2
CMI	0.000387	0.018176	0.977557	0.988523	2
CPRT	0.001047	0.014979	0.927105	0.992372	2
CSX	0.000646	0.017462	1.043010	0.955829	2
DAL	0.000517	0.024814	0.952124	0.980727	2
DE	0.000556	0.016703	0.951225	0.983644	2
DOV	0.000546	0.016845	0.988254	0.968131	2
EMR	0.000267	0.016463	0.930394	0.967078	2
ETN	0.000450	0.017423	0.995600	0.974408	2
EXPD	0.000267	0.014679	1.011972	0.957943	2
FAST	0.000560	0.016874	1.017518	0.999670	2
FBHS	0.000868	0.019631	0.965633	0.978979	2
FDX	0.000456	0.017386	0.963798	0.981288	2
GD	0.000396	0.014028	1.047922	0.940179	2
GWW	0.000490	0.016469	0.960258	0.995514	2
HII	0.000664	0.017031	1.015763	0.935556	2
HWM	-0.000064	0.025139	0.909516	0.939999	2
IEX	0.000699	0.014599	1.031668	0.962203	2
IR	0.000843	0.024407	1.015252	0.942748	2
ITW	0.000639	0.014417	0.922543	0.969491	2
J	0.000343	0.017644	1.045491	0.976824	2
JBHT	0.000498	0.015420	1.081486	1.014217	2
JCI	0.000431	0.015401	1.046638	0.968509	2
KSU	0.000603	0.018785	1.023660	0.994975	2
LHX	0.000620	0.015306	0.974821	0.932664	2
LMT	0.000783	0.012718	0.962104	0.975880	2
LUV	0.000532	0.020178	0.986733	0.964936	2
MAS	0.000680	0.020248	1.019500	0.984060	2
NOC	0.000740	0.013834	1.019146	0.947251	2
NSC	0.000618	0.016801	1.014521	0.964525	2
ODFL	0.001056	0.017740	1.116417	0.950527	2
PCAR	0.000295	0.016396	1.074364	0.986350	2
PH	0.000523	0.018614	0.968202	0.976833	2
PNR	0.000388	0.017448	1.044595	0.924451	2
PWR	0.000523	0.019236	0.945752	0.953866	2
RHI	0.000365	0.018797	0.971600	0.962834	2
ROK	0.000568	0.018245	0.959031	0.993951	2
ROL	0.000815	0.014646	0.979196	0.991513	2
ROP	0.000716	0.014215	0.935390	0.964640	2
RSG	0.000593	0.011185	0.925639	0.977828	2
SNA	0.000507	0.016137	0.944876	0.962180	2
TDY	0.000861	0.016373	1.002064	0.966281	2
TT	0.000704	0.017061	0.988580	0.965844	2
UNP	0.000661	0.015663	1.073618	0.996020	2
UPS	0.000453	0.012969	0.923857	0.967459	2
URI	0.000914	0.028867	1.027572	0.951293	2
WAB	0.000425	0.019372	0.976249	0.969278	2
XYL	0.000673	0.016063	1.013724	0.981268	2

Table 6: MLE estimates of a Skewed Exponential Power Distribution and the entropy weighting clustering results – S&P500 data.

Heterogeneity can be analysed also by means of the features' weights that show at the same time the relative importance of each estimated parameter in determining the cluster's composition. The optimal weights for this experiment are reported in the Tab. 7.

	Location (w_1)	Scale (w_2)	Shape (w_3)	Skewness (w_4)
Cluster 1	0.255808	0.255629	0.247364	0.241199
Cluster 2	0.258694	0.258503	0.229691	0.253112

Table 7: Distribution-based EWKM for S&P500 stocks: resulting weights

The weights in Tab. 7 highlight that the important information in determining clusters' differences are the distribution's shapes. Indeed, while the other parameters have almost the same weights, very close to an equally weighting scheme, in the cluster 2 the shape is less weighted. According to the weights interpretation we have seen so far, the lower weight assigned to w_p depends by the greater distances among the stocks within the second cluster in terms of shape.

While in the previous experiment we compared the clusters obtained with the proposed distribution-based approach with those obtained by a correlation-based one, in this case this is not possible. Indeed, not all the clustering procedures can handle time series with different lengths.

In the next Section, we propose a possible use of this clustering approach in the real world. An immediate example is, once it is applied to financial data, represented by the portfolio selection. Therefore, in the Section 4 we provide the results about the financial performance of the portfolios build by means of the proposed clustering model.

In this context, since we will work only with time series of equal length, we will be able to compare the proposed clustering approach with a correlation-based one also for the S&P500 Industrial data.

4. Portfolio analysis

The clusters obtained in the previous Section by the proposed approach can be seen as possible portfolios from an asset allocation perspective.

Financial literature provided various approaches to portfolio selection. In what follows, we consider the Global Minimum Variance (GMV) strategy [52]. Assuming to have N time series of stock returns collected into a matrix R_t, the portfolio problem can we written as [53,54]:

$$\min_{w} w'\Sigma w \qquad (14)$$

under the constraint:

$$\sum_{n=1}^{N} w_m = 1 \qquad (15)$$

The optimal global minimum variance weights w, as solution of the minimization problem (14), are:

$$w = \frac{\Sigma^{-1}1_N}{1'_N\Sigma^{-1}1_N} \qquad (16)$$

Note that the elements of the vector w can be negative, so we allow for short sales. Then, by replacing Σ^{-1} with $\hat{\Sigma}^{-1}$ we get the optimal *estimated* GMV portfolio weights that we

call $\hat{w}$. In this paper, since we don't have the problem of dimensionality[3], we estimate the covariance matrix Σ by means of the sample covariance estimator:

$$\hat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T} (R_t - \hat{\mu})' (R_t - \hat{\mu}) \tag{17}$$

with $\hat{\mu}$ is the vector containing the sample averages over the time of the stocks in R_t.

Nevertheless, [59] showed that empirically the naive or *Talmudic*[4] diversification rule returns the highest performances in out-of-sample analysis with respect most of alternatives. This result highlights the relevance of the estimation error in portfolio selection, coming from the fact that the investors estimates unknown quantities. Indeed, the equally weighted strategy $(1/N)$ is the only diversification strategy with zero estimation error, since nothing is estimated.

In what follows we consider each cluster as a possible set of stock and we use both the naive $1/N$ and the Global Minimum Variance (GMV) approaches to build C-th different portfolios.

First of all, we use the first 5 years of observations to generate the clusters according to the distribution-based procedure discussed above. Then, the proposed clustering approach is compared from the point of view of asset allocation also with a correlation-based clustering, commonly used in finance to form portfolio of stocks.

In order to evaluate the out-of-sample performances of each portfolio, we follow the empirical procedure of [59], based on a "rolling-sample" approach.

Specifically, given a T daily observation of the securities returns, we choose an estimation widow of one year, $M = 252$, to estimate the covariance structure across the asset needed for the implementation of the GMV strategy.

Then, in order to avoid a costly daily portfolio rebalancing, we suppose a monthly rebalance, such that with a window of $M = 252$ observations the investor update the portfolio structure each $m = 20$ trading days.

This process is recursively repeated by adding the return for the next period in the dataset and dropping the earliest one until the end of the dataset is reached. The result is, therefore, a time series of length $(T - M)/m$ of returns[5].

Given the time series of monthly out-of-sample returns, we compute the out-of-sample Sharpe ratio of the portfolio c, SR_c, defined as the sample mean of out-of-sample portfolio returns divided by its standard deviation:

$$SR_c = \frac{\hat{\mu}_c}{\hat{\sigma}_c} \tag{18}$$

where $\hat{\mu}_c$ is the average of the $(T - M)/m$ out of sample returns for the c-th portfolio and $\hat{\sigma}_c$ its standard deviation. Moreover, to account for the amount of trading required to implement the GMV strategy, we compute the portfolio turnover, defined as follows:

$$TOV_c = \frac{1}{\tilde{T}} \sum_{t=1}^{\tilde{T}} \sum_{n=1}^{N} (|\hat{w}_{n,t+1} - \hat{w}_{n,t}|) \tag{19}$$

with $\tilde{T} = (T - M)/m$ and $\hat{w}_{n,t}$ be the portfolio GMV weight assigned to the n-th asset at time t with the covariance matrix across the assets estimated with the last M observations.

---

[3] In the *large dimensional setting*, where $N > T$, the sample covariance estimator results in an ill-conditioned covariance matrix that cannot be inverted (for example see [55–58]). However in both the considered applications presented in this paper we have that $T > N$ (actually M, the estimation window, is always greater than the number of assets N).

[4] The *Talmud* is the central text of Rabbinic Judaism that provides the following investment advice: "let every man divide his money into three parts, and invest a third in land, a third in business and a third let him keep by him in reserve".

[5] Supposing a daily portfolio rebalancing the final length would be $T - M$. In presence of trading costs, a daily rebalace is intuitively more expensive than a monthly one.
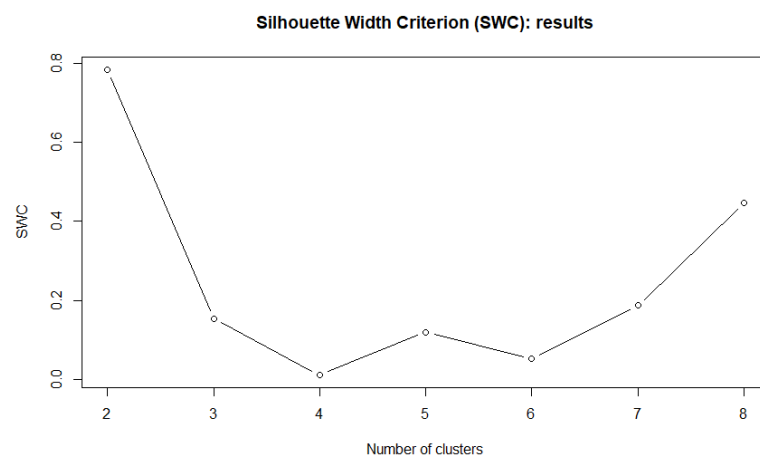
4.1. FTSE100

We consider first how the clustering approaches can be used to form portfolios of stocks (e.g. [1,3]) in the case of the first analysed dataset containing the $N = 25$ stocks without missing values included in the FTSE100 Index.

First of all, in order to backtest the profitability of the trading strategies based on the clustering approaches, we consider only the first 5 years of daily observation as a dataset to perform cluster analysis. Clearly, since we are using half of the sample of the analysis conducted in the previous Section, we could expect different stocks' classification.

As in the previous Section, we compare the proposed distribution-based clustering approach with another common clustering model used in finance to build portfolio of stocks. The alternative clustering approach uses the assets' correlations instead of their distribution to build the clusters (e.g. [1,2]).

As shown by Fig. 10, according to the SEPD-based EWKM algorithm we select $C = 2$ clusters with an high average silhouette, that is equal to 0,8.

Figure 10. SWC for the first 5-year observations (distribution-based clustering) – FTSE100 data.

On the other hand, following the same approach, the correlation-based clustering approach suggests the presence of $C = 6$ clusters (see Fig. 11) and an average silhouette equal to 0,08, ten times lower than the one shown in Fig. 10.
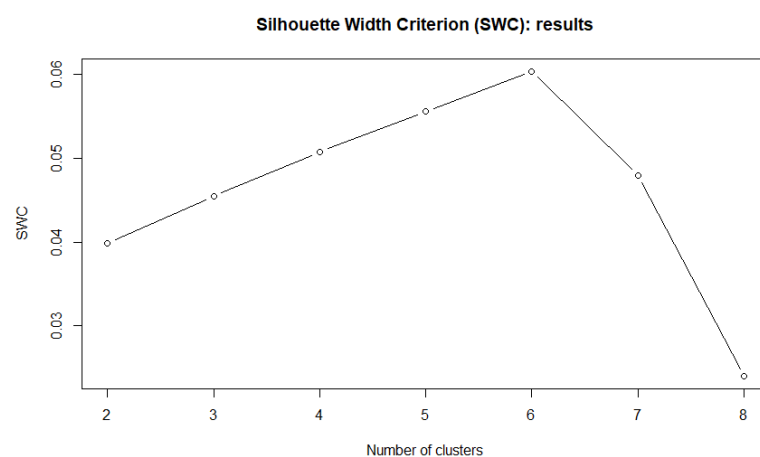
Figure 11. SWC for the first 5-year observations (correlation-based clustering) – FTSE100 data.

In other words, on the basis of some in-sample arguments we can argue that the clustering resulting from the application of a distribution-based approach is much more accurate than

another one based on correlation. The clusters composition for both approaches is shown in Tab. 8.

	SEPD-based clustering	Correlation-based clustering
AAL	2	1
ADM	2	1
AHT	2	1
AUTO	2	2
AZN	2	1
BHP	2	3
BME	2	1
CPG	2	3
CRH	2	1
DGE	1	1
EVR	2	1
FERG	2	4
GSK	1	1
IAG	2	3
IHG	2	1
III	2	5
MNG	2	3
MRO	2	3
NWG	2	3
PRU	2	1
RIO	2	3
SVT	2	6
TSCO	2	1
VOD	1	1
WPP	1	1

Table 8: Final group assignment of the two alternative clustering approaches. The first column shows the results of the distribution-based approach, while the second column shows those of the correlation-based clustering (FTSE100 data).

Nevertheless, in this Section we are interested in the out-of-sample performances in terms of portfolio selection. In the case of the distribution-based clustering, following [3], we consider the two clusters as two possible different portfolios. As stated before, we consider two alternative diversification rules, the *naive* ($1/N$) and the Global Minimum Variance (GMV). Therefore we have four possible trading strategies.

In the case of the correlation-based clustering, from Tab. 8 it clearly appears that the stocks AUTO, SVT, III and FERG from single clusters. Therefore, we exclude these stocks and consider the clusters 1 and 3 as alternative portfolios, constructed with both naive and GMV diversification rules.

We compare the resulting portfolio in terms of return-risk trade-off represented by the Sharpe Ratio, the amount of risk in worst scenarios computed by means of the Value at Risk (VaR) and the Expected Shortfall (ES) and the trading expenses trough the Turnover. The results are shown in Tab. 9.

	Sharpe Ratio	VaR	ES	Turnover
Naive (SEPD) - Cluster 1	0.262155	-0.014396	-0.018755	
GMV (SEPD) - Cluster 1	0.278905	-0.012489	-0.016318	0.022555
Naive (SEPD) - Cluster 2	0.129284	-0.018028	-0.023003	
GMV (SEPD) - Cluster 2	0.123989	-0.009490	-0.012100	0.013587
Naive (corr) - Cluster 1	0.216579	-0.016056	-0.020760	
GMV (corr) - Cluster 1	0.290720	-0.010547	-0.013808	0.018170
Naive (corr) - Cluster 3	0.020610	-0.026068	-0.032775	
GMV (corr) - Cluster 3	0.054591	-0.016854	-0.021284	0.026502

Table 9: Portfolio performance measures – experiment with FTSE100 data

In general, following a naive diversification approach, all the portfolios build with the distribution-based clustering approach show much superior performances than those constructed with the alternative approach. Indeed, the two SEPD-based portfolios have a Sharpe Ratio equal to 26,2% and 12,9% respectively while the alternative portfolios have lower Sharpe Ratios equal to 21,6% and 2%.

In terms of VaR and Expected Shortfall the two SEPD-based portfolios build under naive diversification rule show similar risk profiles with respect the Cluster 1 portfolio build trogh the correlation-based clustering, while the Cluster 2 portfolio (correlation-based) has a very high values compared to the others. Therefore, the SEPD-based clustered portfolios show a better return-risk profile, also in adverse scenarios.

In the end, since the weights' structure don't change over time the turnover any naive portfolio is set to be zero.

On the side of the GMV diversification rules, the benefit of the distribution-based clustered portfolios are still evident. Indeed, despite the best portfolio in terms of Sharpe Ratio is the first cluster obtained by the correlation-based approach (SR equal to 29%), the portfolio build with the Cluster 3 (correlation-based) shows a very poor Sharpe Ratio performance equal to 5%.

The GMV portfolio build on the Cluster 1 (SEPD-based) has a Sharpe Ratio equal to 27,8%, while the one build on the Cluster 2 has a Sharpe Ratio of 12,3%. Clearly, once the cluster analysis is conducted, the investor don't know which portfolio will perform better in out-of-sample. Therefore, let's suppose that ex ante we invest equally across the two clustered portfolios. The overall return of this investment strategy is higher if the investor chooses to invest in the SEPD-based clustered portfolios than in the case of correlation-based.

In terms of VaR and Expected Shortfall the results are even better. Indeed, in both the cases the two portfolios with the lowest VaR and ES are the SEPD-based clustered portfolios.

In terms of turnover, the SEPD-based Cluster 2 shows the lowest value among the alternative and in general the SEPD-based trading rules have a much lower cost in aggergate.

Therefore, we can conclude that the SEPD-based Entropy Weighted algorithm proposed in the Section 2, that aims to cluster stocks according to their distribution, shows good performances also from a portfolio selection perspective. The correlation-based algorithm, that discard data distribution instead of correlations, performs poorer.

4.2. S&P500 industrials

In this sub-Section we provide the portfolio analysis for the second experiment with S&P500 Industrial real data. Nevethless, in this case an important preliminary step to facilitate the analysis under consideration is to exclude from the sample the S&P500 industrial stocks showing missing values. Hence. from an initial sample of $N = 74$, we obtain a thinner sample of $N = 65$ stocks.

As previously, we compare the distribution-based clustering approach presented in the Section 2 with the correlation-based clustering, commonly used to form portfolio of stocks. The Fig. 12 shows the SWC criterion according to different number of clusters C.
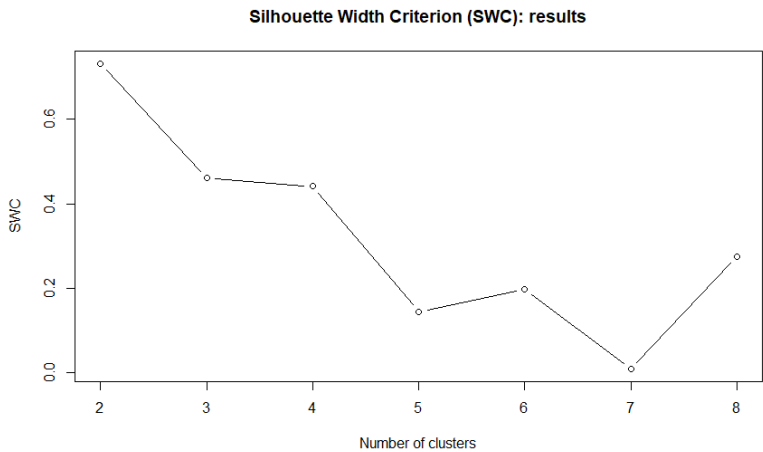
Figure 12. SWC for the first 5-year observations (distribution-based clustering) – S&P500 data.

With a SWC greater than 0,8 we select $C = 2$. In the Fig. 13, instead, is reported the same criterion in the case of the correlation-based clustering algorithm.
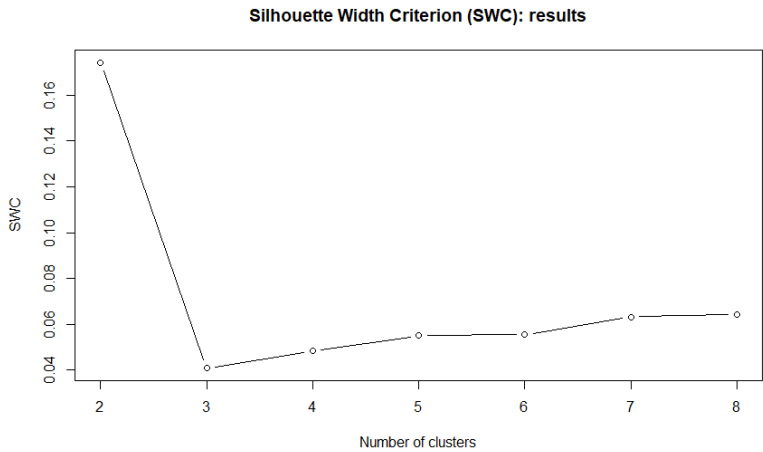
Figure 13. SWC for the first 5-year observations (correlation-based clustering) – S&P500 data.

In this second experiment, the correlation-based clustering model suggests the same groups as the distribution-based one. However, the silhouette is again very low compared to the one showed in Fig. 12, meaning that the quality of the resulting classification is much lower. The different clustering results are reported in the Tab. 10

	SEPD-based clustering	Correlation-based clustering
AAL	2	1
ALK	1	1
AME	2	2
AOS	2	2
BA	1	2
CAT	2	2
CHRW	2	2
CMI	2	2
CPRT	2	2
CSX	2	2
CTAS	2	2
DAL	1	1
DE	2	2
DOV	2	2
EFX	2	2
EMR	2	2
ETN	2	2
EXPD	2	2
FAST	2	2
FDX	2	2
GD	1	2
GE	2	2
GNRC	2	2
GWW	2	2
HON	2	2
HWM	2	2
IEX	2	2
ITW	2	2
J	2	2
JBHT	2	2
JCI	2	2
KSU	2	2
LDOS	2	2
LHX	2	2
LMT	1	2
LUV	1	1
MAS	2	2
MMM	2	2
NOC	1	2
NSC	2	2
ODFL	2	2
PCAR	2	2
PH	2	2
PNR	2	2
PWR	2	2
RHI	2	2
ROK	2	2
ROL	2	2
ROP	2	2
RSG	2	2
RTX	2	2
SNA	2	2
SWK	2	2
TDG	2	2
TDY	2	2
TT	2	2
TXT	2	2
UAL	1	1
UNP	1	2
UPS	2	2
URI	2	2
VRSK	2	2
WAB	2	2
WM	2	2

Table 10: Final group assignment of the two alternative clustering approaches. The first column shows the results of the distribution-based approach, while the second column shows those of the correlation-based clustering (S&P500 data).

The portfolio performances of the proposed approaches, assuming both naive and GMV diversification rules, are reported in the Tab. 11.

	Sharpe Ratio	VaR	ES	Turnover
Naive (SEPD) - Cluster 1	0.201330	-0.018444	-0.023791	
GMV (SEPD) - Cluster 1	0.300182	-0.014314	-0.018771	0.027739
Naive (SEPD) - Cluster 2	0.147101	-0.019142	-0.024487	
GMV (SEPD) - Cluster 2	0.098290	-0.009808	-0.012460	0.020254
Naive (corr) - Cluster 1	0.131469	-0.029350	-0.037461	
GMV (corr) - Cluster 1	0.196732	-0.023842	-0.030730	0.037102
Naive (corr) - Cluster 2	0.155000	-0.018541	-0.023746	
GMV (corr) - Cluster 2	0.097379	-0.009428	-0.011975	0.026502

Table 11: Portfolio performance measures – experiment with S&P500 data

In the case of naive diversification rule, the Tab. 11 shows that the best portfolio in terms of out-of-sample Sharpe Ratio is the one based on the Cluster 1 resulting from the distribution-based clustering approach, with a value of 20%. Moreover, in terms of VaR and ES the two distribution-based clustered portfolios share similar risk than the correlation-based Cluster 2 portfolio, while the correlation-based Cluster 1 portfolio shows much higher values being, therefore, much more risky in adverse scenarios.

The construction of GMV portfolios starting from the identified clusters shows similarly interesting results. In particular, the distribution-based Cluster 1 portfolio still is the most performing with a Sharpe Ratio of 30%, while the correlation-based GMV Cluster 1 portfolio has a peformance lower than 20%. On the other side, both portfolios constructed on Cluster 2 shows similar Shape Ratio but still the distribution-based allows a little overperformance of 10 basis points.

In terms of risk, looking at the VaR and ES, the distribution-based Cluster 1 portfolio has a much lower amount of risk compared to the correlation-based Cluster 1 portfolio and at the same time has a much higher Sharpe Ratio. The other two portfolios constructed according to the the Cluster 2 are again very similar.

In the end, we compare the portfolio performances with respect the turnover. The distribution-based Cluster 2 portfolio has the lower turnover, while the correlation-based Cluster 1 the highest. Moreover, the distribution-based Cluster 1 portfolio has a similar turnover than the correlation-based Cluster 2 but with a Sharpe Ratio higher than 11%.

Therefore, also in this case we can conclude that the SEPD-based Entropy Weighted K-means approach developed in the Section 2 allows the construction of high performance clustered portfolios, regardless the diversification rule used for their construction.

5. Conclusions

In this paper, we propose a new model-based clustering approach for classifying skewed and heavy tailed time series, by means of an entropy weighting clustering algorithm.

Clustering techniques are useful tools for exploratory data analysis in the way they identify common structures in an unlabeled dataset.

For example, a possible application of financial time series clustering concerns the asset allocation, where groups of similar stocks could be seen as portfolios of asset that shares similar characteristics.

Many recent papers aim to improve the existing clustering techniques for time series data. This article proposes a model clustering model that refers to data based on a very important family of Asymmetric functions: the Skewed Exponential Power Distribution (SEPD) also called in literature as the Skewed Generalized Error Distribution (SGED). This distribution is very useful for classifying time series in presence of fat-tailed and asymmetric time series.

The clustering algorithm, which represents the innovative aspect of this paper, applies the idea of entropy weighting clustering of [7,45] to the parameters estimated by a flexible

probability distribution as in [23].

The criterion is that time series with similar parameter estimates are placed in the same group. Therefore, with a k-means clustering algorithm, the measure of dissimilarity is determined on the basis of these estimates. In this paper we therefore propose to combine all the information in an optimal way to form clusters.

Finally, to demonstrate the effectiveness of the proposed clustering approach, in this paper we propose two different applications to stock market data. Financial market data lend themselves well to adhering to our methodological proposal. In fact, the empirical densities of daily stock returns time series are proved to be non-Gaussian, asymmetric and heavy.

Ours wants to be a fairly innovative research address and certainly many can be there financial applications that benefit from modeling equity returns via exponential power distribution and its extensions for skewness.

Indeed a final important result allows us to conclude that the new clustering algorithm we described in the paper can be used to form equity portfolios. Indeed, we compared the performances of the distribution-based clustering model proposed in this paper with a correlation-based clustering algorithm that is commonly used by financial practitioners to form portfolio of stocks. According to several measures, such as the Sharpe Ratio, the Value at Risk, the Expected Shortfall and the Turnover we demonstrated the superior performances of the proposed clustering approach also from an asset allocation perspective.

A first possible future research can be devoted to the application of the proposed underlying idea to different probability distributions. For example, the Asymmetric Power Distribution of [40] represents an interesting possibility for modeling situations where we suppose two different behaviors in the distribution's tails.

Moreover, another interesting research direction can be devoted to the developments of a new distribution-based clustering approach where also the time varying parameters estimated from the Skewed Exponential Power Distribution (or others) are considered.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A List of stocks

ID	Name	Symbol	Sector
1	3i	III	Financial Services
2	Admiral Group	ADM	Nonlife Insurance
3	Anglo American plc	AAL	Mining
5	Ashtead Group	AHT	Support Services
7	AstraZeneca	AZN	Pharmaceuticals & Biotechnology
8	Auto Trader Group	AUTO	Media
12	B&M	BME	Retailers
13	BAE Systems	BA.	Aerospace & Defence
17	BHP	BHP	Mining
25	Compass Group	CPG	Support Services
26	CRH plc	CRH	Construction & Materials
29	Diageo	DGE	Beverages
31	Evraz	EVR	Industrial Metals & Mining
33	Ferguson plc	FERG	Support Services
36	GlaxoSmithKline	GSK	Pharmaceuticals & Biotechnology
42	IHG Hotels & Resorts	IHG	Travel & Leisure
46	International Airlines Group	IAG	Travel & Leisure
56	M&G	MNG	Asset Managers
57	Melrose Industries	MRO	Automobiles & Parts
60	NatWest Group	NWG	Banks
68	Prudential plc	PRU	Life Insurance
74	Rio Tinto	RIO	Mining
83	Severn Trent	SVT	Gas, Water & Multi-utilities
94	Tesco	TSCO	Food & Drug Retailers
97	Vodafone Group	VOD	Mobile Telecommunications
100	WPP plc	WPP	Media

Table 12: List of stocks (FTSE100) considered in the application with real data

ID	Symbol	Name
1	MMM	3M Company
2	AOS	A.O. Smith Corp
16	ALK	Alaska Air Group
21	ALLE	Allegion
30	AAL	American Airlines Group
38	AME	Ametek
70	BA	Boeing Company
79	CHRW	C. H. Robinson Worldwide
88	CARR	Carrier Global
90	CAT	Caterpillar Inc.
107	CTAS	Cintas Corporation
123	CPRT	Copart Inc
128	CSX	CSX Corp.
129	CMI	Cummins Inc.
135	DE	Deere & Co.
136	DAL	Delta Air Lines Inc.
150	DOV	Dover Corporation
158	ETN	Eaton Corporation
164	EMR	Emerson Electric Company
168	EFX	Equifax Inc.
179	EXPD	Expeditors
184	FAST	Fastenal Co
186	FDX	FedEx Corporation
197	FTV	Fortive Corp
198	FBHS	Fortune Brands Home & Security
206	GNRC	Generac Holdings
207	GD	General Dynamics
208	GE	General Electric
216	GWW	Grainger (W.W.) Inc.
230	HON	Honeywell Int'l Inc.
233	HWM	Howmet Aerospace
237	HII	Huntington Ingalls Industries
238	IEX	IDEX Corporation
240	INFO	IHS Markit
241	ITW	Illinois Tool Works
244	IR	Ingersoll-Rand
257	JBHT	J. B. Hunt Transport Services
259	J	Jacobs Engineering Group
262	JCI	Johnson Controls International
265	KSU	Kansas City Southern
276	LHX	L3Harris Technologies
282	LDOS	Leidos Holdings
289	LMT	Lockheed Martin Corp.
301	MAS	Masco Corp.
333	NLSN	Nielsen Holdings
336	NSC	Norfolk Southern Corp.
338	NOC	Northrop Grumman
349	ODFL	Old Dominion Freight Line
353	OTIS	Otis Worldwide
354	PCAR	Paccar
356	PH	Parker-Hannifin
361	PNR	Pentair plc
387	PWR	Quanta Services Inc.
391	RTX	Raytheon Technologies
396	RSG	Republic Services Inc
398	RHI	Robert Half International
399	ROK	Rockwell Automation Inc.
400	ROL	Rollins Inc.
401	ROP	Roper Technologies
415	SNA	Snap-on
417	LUV	Southwest Airlines
418	SWK	Stanley Black & Decker
433	TDY	Teledyne Technologies
438	TXT	Textron Inc.
449	TT	Trane Technologies plc
450	TDG	TransDigm Group
461	UNP	Union Pacific Corp
462	UAL	United Airlines Holdings
463	UPS	United Parcel Service
464	URI	United Rentals Inc.
471	VRSK	Verisk Analytics
483	WM	Waste Management Inc.
491	WAB	Westinghouse Air Brake Technologies Corp
500	XYL	Xylem Inc.

Table 13: List of stocks (FTSE100) considered in the application with real data

References

1. Mantegna, R.N. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems* **1999**, *11*, 193–197.
2. Tola, V.; Lillo, F.; Gallegati, M.; Mantegna, R.N. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control* **2008**, *32*, 235–258.
3. Iorio, C.; Frasso, G.; D'Ambrosio, A.; Siciliano, R. A P-spline based clustering approach for portfolio selection. *Expert Systems with Applications* **2018**, *95*, 88–103.
4. Liao, T.W. Clustering of time series data—a survey. *Pattern recognition* **2005**, *38*, 1857–1874.
5. D'Urso, P. Dissimilarity measures for time trajectories. *Statistical Methods & Applications* **2000**, *1*, 53–83.
6. D'urso, P. Fuzzy C-means clustering models for multivariate time-varying data: different approaches. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **2004**, *12*, 287–326.
7. Coppi, R.; D'Urso, P. Fuzzy unsupervised classification of multivariate time trajectories with the Shannon entropy regularization. *Computational statistics & data analysis* **2006**, *50*, 1452–1477.
8. Coppi, R.; D'Urso, P.; Giordani, P. A fuzzy clustering model for multivariate spatial time series. *Journal of Classification* **2010**, *27*, 54–88.
9. D'Urso, P.; De Giovanni, L.; Massari, R. Robust fuzzy clustering of multivariate time trajectories. *International Journal of Approximate Reasoning* **2018**, *99*, 12–38.
10. Caiado, J.; Crato, N.; Peña, D. Comparison of times series with unequal length in the frequency domain. *Communications in Statistics—Simulation and Computation®* **2009**, *38*, 527–540.
11. Alonso, A.M.; Maharaj, E.A. Comparison of time series using subsampling. *Computational statistics & data analysis* **2006**, *50*, 2589–2599.
12. D'Urso, P.; Maharaj, E.A. Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems* **2009**, *160*, 3565–3589.
13. Caiado, J.; Crato, N.; Peña, D. A periodogram-based metric for time series classification. *Computational Statistics & Data Analysis* **2006**, *50*, 2668–2684.
14. D'Urso, P.; Maharaj, E.A. Wavelets-based clustering of multivariate time series. *Fuzzy Sets and Systems* **2012**, *193*, 33–61.
15. Maharaj, E.A.; D'Urso, P.; Galagedera, D.U. Wavelet-based fuzzy clustering of time series. *Journal of classification* **2010**, *27*, 231–275.
16. Maharaj, E.A.; D'Urso, P. Fuzzy clustering of time series in the frequency domain. *Information Sciences* **2011**, *181*, 1187–1211.
17. D'Urso, P.; De Giovanni, L.; Massari, R.; D'Ecclesia, R.L.; Maharaj, E.A. Cepstral-based clustering of financial time series. *Expert Systems with Applications* **2020**, *161*, 113705.
18. Piccolo, D. A distance measure for classifying ARIMA models. *Journal of Time Series Analysis* **1990**, *11*, 153–164.
19. Otranto, E. Clustering heteroskedastic time series by model-based procedures. *Computational Statistics & Data Analysis* **2008**, *52*, 4685–4698.
20. D'Urso, P.; De Giovanni, L.; Massari, R. GARCH-based robust clustering of time series. *Fuzzy Sets and Systems* **2016**, *305*, 1–28.
21. Iorio, C.; Frasso, G.; D'Ambrosio, A.; Siciliano, R. Parsimonious time series clustering using p-splines. *Expert Systems with Applications* **2016**, *52*, 26–38.
22. Maharaj, E.A.; Alonso, A.M.; D'Urso, P. Clustering seasonal time series using extreme value analysis: An application to Spanish temperature time series. *Communications in Statistics: Case Studies, Data Analysis and Applications* **2015**, *1*, 175–191.
23. D'Urso, P.; Maharaj, E.A.; Alonso, A.M. Fuzzy clustering of time series using extremes. *Fuzzy Sets and Systems* **2017**, *318*, 56–79.
24. Corduas, M.; Piccolo, D. Time series clustering and classification by the autoregressive metric. *Computational statistics & data analysis* **2008**, *52*, 1860–1872.
25. D'Urso, P.; Cappelli, C.; Di Lallo, D.; Massari, R. Clustering of financial time series. *Physica A: Statistical Mechanics and its Applications* **2013**, *392*, 2114–2129.
26. Cerqueti, R.; Giacalone, M.; Mattera, R. Model-based fuzzy time series clustering of conditional higher moments. *International Journal of Approximate Reasoning* **2021**, *134*, 34–52.
27. Fernandez, C.; Osiewalski, J.; Steel, M.F. Modeling and inference with v-spherical distributions. *Journal of the American Statistical Association* **1995**, *90*, 1331–1340.
28. Fernández, C.; Steel, M.F. On Bayesian modeling of fat tails and skewness. *Journal of the american statistical association* **1998**, *93*, 359–371.

29. Theodossiou, P. Skewed generalized error distribution of financial assets and option pricing. *Multinational Finance Journal* **2015**, *19*, 223–266.
30. Komunjer, I. Asymmetric power distribution: Theory and applications to risk measurement. *Journal of applied econometrics* **2007**, *22*, 891–921.
31. Hsieh, D.A. Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business & Economic Statistics* **1989**, *7*, 307–317.
32. Nelson, D.B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society* **1991**, pp. 347–370.
33. Duan, J.C. Conditionally fat-tailed distributions and the volatility smile in options. *Rotman School of Management, University of Toronto, Working Paper* **1999**.
34. Ayebo, A.; Kozubowski, T.J. An asymmetric generalization of Gaussian and Laplace laws. *Journal of Probability and Statistical Science* **2003**, *1*, 187–210.
35. Christoffersen, P.; Dorion, C.; Jacobs, K.; Wang, Y. Volatility components, affine restrictions, and nonnormal innovations. *Journal of Business & Economic Statistics* **2010**, *28*, 483–502.
36. Mattera, R.; Giacalone, M. Alternative distribution based garch models for bitcoin volatility estimation. *The Empirical Economics Letters* **2018**, *17*, 1283–1288.
37. Cerqueti, R.; Giacalone, M.; Panarello, D. A Generalized Error Distribution Copula-based method for portfolios risk assessment. *Physica A: Statistical Mechanics and its Applications* **2019**, *524*, 687–695.
38. Cerqueti, R.; Giacalone, M.; Mattera, R. Skewed non-Gaussian GARCH models for cryptocurrencies volatility modelling. *Information Sciences* **2020**.
39. Trucíos, C. Forecasting Bitcoin risk measures: A robust approach. *International Journal of Forecasting* **2019**, *35*, 836–847.
40. Zhu, D.; Zinde-Walsh, V. Properties and estimation of asymmetric exponential power distribution. *Journal of econometrics* **2009**, *148*, 86–99.
41. Kotz, S.; Kozubowski, T.; Podgorski, K. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*; Springer Science & Business Media, 2012.
42. Giacalone, M.; Panarello, D.; Mattera, R. Multicollinearity in regression: an efficiency comparison between Lp-norm and least squares estimators. *Quality & Quantity* **2018**, *52*, 1831–1859.
43. Giacalone, M. A combined method based on kurtosis indexes for estimating p in non-linear Lp-norm regression. *Sustainable Futures* **2020**, *2*, 100008.
44. Mudholkar, G.S.; Hutson, A.D. The epsilon–skew–normal distribution for analyzing near-normal data. *Journal of statistical planning and inference* **2000**, *83*, 291–309.
45. Jing, L.; Ng, M.K.; Huang, J.Z. An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data. *IEEE Transactions on knowledge and data engineering* **2007**, *19*, 1026–1041.
46. Huang, J.Z.; Ng, M.K.; Rong, H.; Li, Z. Automated variable weighting in k-means type clustering. *IEEE transactions on pattern analysis and machine intelligence* **2005**, *27*, 657–668.
47. Eshima, N. *Statistical Data Analysis and Entropy*; Springer, 2020.
48. Rousseeuw, P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* **1987**, *20*, 53–65.
49. Adcock, C.; Eling, M.; Loperfido, N. Skewed distributions in finance and actuarial science: a review. *The European Journal of Finance* **2015**, *21*, 1253–1281.
50. Cont, R. Empirical properties of asset returns: stylized facts and statistical issues **2001**.
51. Jarque, C.M.; Bera, A.K. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique* **1987**, pp. 163–172.
52. Bodnar, T.; Mazur, S.; Okhrin, Y. Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research* **2017**, *256*, 292–307.
53. Bodnar, T.; Gupta, A.K. Robustness of the inference procedures for the global minimum variance portfolio weights in a skew-normal model. *The European Journal of Finance* **2015**, *21*, 1176–1194.
54. Bodnar, T.; Mazur, S.; Podgórski, K. A test for the global minimum variance portfolio for small sample and singular covariance. *AStA Advances in Statistical Analysis* **2017**, *101*, 253–265.
55. Pappas, D.; Kiriakopoulos, K.; Kaimakamis, G. Optimal portfolio selection with singular covariance matrix. International Mathematical Forum, 2010, Vol. 5, pp. 2305–2318.
56. Bodnar, T.; Mazur, S.; Podgórski, K. Singular inverse Wishart distribution and its application to portfolio theory. *Journal of Multivariate Analysis* **2016**, *143*, 314–326.

57. Bodnar, T.; Mazur, S.; Podgórski, K.; Tyrcha, J. Tangency portfolio weights for singular covariance matrix in small and large dimensions: estimation and test theory. *Journal of Statistical Planning and Inference* **2019**, *201*, 40–57.

58. Gulliksson, M.; Mazur, S. An iterative approach to ill-conditioned optimal portfolio selection. *Computational Economics* **2019**, pp. 1–22.

59. De Miguel, V.; Garlappi, L.; Uppal, R. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies* **2007**, 22, 1915–1953.