Notes on Intuitionistic Fuzzy Sets
Print ISSN 1310–4926, Online ISSN 2367–8283
2022, Volume 28, Number 3, Pages 343–352
DOI: 10.7546/nifs.2022.28.3.343-352

Proceedings of the 25th Jubilee Edition
of the International Conference on
Intuitionistic Fuzzy Sets (ICIFS'2022)
9–10 September 2022, Sofia, Bulgaria

Clustering stock price volatility using intuitionistic fuzzy sets

Georgy Urumov[1] **and Panagiotis Chountas**[2]

[1] School of Computer Science and Engineering, University of Westminster
115 New Cavendish Street, London W1W 6UW
e-mail: w1767944@westminster.ac.uk

[2] School of Computer Science and Engineering, University of Westminster
115 New Cavendish Street, London W1W 6UW
e-mail: p.i.chountas@westminster.ac.uk

Received: 28 June 2022 **Revised:** 31 July 2022 **Accepted:** 1 August 2022

Abstract: Clustering involves gathering a collection of objects into homogeneous groups or clusters, such that objects in the same cluster are more similar when compared to objects present in other groups. Clustering algorithms that generate a tree of clusters called dendrogram which can be either divisive or agglomerative. The partitional clustering gives a single partition of objects, with a predefined *K* number of clusters. The most popular partition clustering approaches are: *k*-means and fuzzy *C*-means (FCM). In *k*-means clustering, data are divided into a number of clusters where data elements belong to exactly one cluster. The *k*-means clustering works well when data elements are well separable. To overcome the problem of non-separability, FCM and IFCM clustering algorithm were proposed. Here we review the use of FCM/IFCM with reference to the problem of market volatility.

Keywords: K-Means, FCM, IFCM, Intuitionistic fuzzy sets, Volatility of Volatility.
2020 Mathematics Subject Classification: 03E72, 68T20.

1 Introduction

In order to understand the financial markets, it is critical to study the volatility of the market. This research report will help the reader in understanding the stock volatility and possible ways of predicting it. Volatility is a metric of the dispersion of returns for a given security or market index. In most cases, the higher the volatility, the riskier the security. Volatility is measured as either the standard deviation or variance between returns from that same security or market index.

In the securities markets, volatility is often associated with big swings in either direction. An asset's volatility is a key factor when pricing options contracts. ‘Vega’ or the sensitivity of the options’ prices with respect to volatility is normally the largest component (or partial derivative in math parlance) of the options’ price.

Volatility often refers to the amount of uncertainty or risk related to the size of changes in a security's value. This means that the price of the security can change dramatically over a short time period in either direction. A lower volatility means that a security's value does not fluctuate dramatically, and tends to be steadier.

Market volatility can also be seen through the VIX or Volatility Index. The VIX was created by the Chicago Board Options Exchange as a metric to capture the 30-day expected volatility of the U.S. stock market derived from real-time quote prices of S&P 500 call and put options. It is effectively a instrument of future stakes investors and traders are making on the direction of the markets or individual securities. A high reading on the VIX implies a riskier market.

Also referred to as statistical volatility, historical volatility (HV) measures the fluctuations of underlying securities by measuring price changes over certain periods of time. It is the less widespread metric compared to implied volatility because it isn't forward-looking.

When there is a rise in historical volatility, a stock’s price will also move more than normal. At this time, there is an expectation that something has changed. If the historical volatility is dropping, on the other hand, it means any uncertainty has been reduced significantly, so things return to the way they were.

CBOE Volatility Index (VIX) is a real-time measure of the market’s expectations of the near-term changes in the S&P 500 index. The index is derived from the prices of at the money and out of the money SPX index options with near-term expiration dates. The calculation derives a 30-day forward projection of volatility. In financial markets, volatility is a measure of how fast prices change and measures the degree of fear or greed amongst the market participants. The formula for the VIX calculation [5] is:

$$\sigma^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T}\left[\frac{F}{K_0} - 1\right]^2$$

where:

- $VIX = \sigma * 100$,
- T = Time to expiration,
- F = Forward index level derived from index option prices,
- K_0 = First strike below the forward index level, F
- K_i = Strike price of ith out of the money option; a call if $K_i > K_0$ and a put if $K_i < K_0$; both put and call if $K_i = K_0$.

$$\Delta K_i = \frac{K_{i+1} - K_{i-1}}{2}$$

- R = Risk-free interest rate to expiration
- $Q(K_i)$ = The midpoint of the bid-ask spread for each option with strike K_i.

In 2012, CBOE introduced a new index called VVIX. The VVIX index is a volatility of volatility measure as it represents the expected volatility of the 30-day forward price of the CBOE

Volatility Index (the VIX). It is this expected volatility that drives the price of the VIX options. VVIX is calculated from the price of a portfolio of liquid at and out of the money VIX options.

As one of the parameters that influences the movement in stock prices, volatility is a crucial input to numerous financial market operations and it becomes an important tool to assess the risk of ruin for portfolio managers, investors and other interested parties [4, 6–9]. The modelling of volatility is a complex task, because it cannot be observed directly. Indeed, it can only be measured by looking at the extent of the movement of the option prices and deriving it mathematically from that.

The mathematical estimators are complicated also by volatility clusters, fat tails, non-normality of the distribution and structural breaks in the distribution of the returns. These features cannot be captured by simple classical models such as the autoregressive moving average (ARMA) process.

The increasing amount of data being generated and collected at various operational domains has made unsupervised data mining tasks, and clustering a preferable data analysis technique. Clustering is the organisation of clusters that contains proximate feature vectors according to some distance measure.

2 Clustering volatility with FS and IFS

A major challenge posed by big data clustering applications is dealing with uncertainty in the formation of the feature vectors. Considering that feature values may be subject to uncertainty owing to imprecise measurements and noise, the distances that determine the membership of a feature vector to a cluster will also be subject to uncertainty. Therefore, the possibility of erroneous membership assignments in the clustering process is evident.

Current fuzzy clustering approaches do not utilise any information about uncertainty at the feature level. This paper accepts the challenge to deal with such kind of information and introduces some thoughts about a modification to the FCM. Features are represented by intuitionistic fuzzy values, i.e., elements of an intuitionistic fuzzy set. Intuitionistic fuzzy sets [6-8] that can be useful in coping with the hesitancy originating from imperfect or imprecise information

3 Results and discussion

The computation is coded in Google Colab, Python. We study daily dataset for VIX from January 1990 to April 2022. Our primary data source is Bloomberg Data services. In our view daily data for this study is preferable to other time frames as more short-term data such as minute or hourly data is far too noisy with little predictive value whereas data from longer timeframes is less useful. There are many measures of volatility. One such measure is called the CBOE Volatility Index or VIX. This index was created by Menachem Brenner and Dan Galai in a series of papers starting in 1989. VIX measures 30-day expectation of volatility given by a weighted portfolio of at and out-the-money European options on the S&P 500 Index. In other words, it is a weighted average of implied volatilities as measured from the call and put prices.

Volatility moves in financial markets are rare and sporadic i.e. periods of low volatility typically follow each other until the regime change due to technical and/or fundamental exogenous factors. Subsequently higher volatility tends to lead to even higher volatility and vice versa.

As one of the parameters that influences the movement in stock prices, volatility is a crucial input to numerous financial market operations and it becomes an important tool to assess the risk of ruin for portfolio managers, investors and other interested parties. The modelling of volatility is a complex task because it cannot be observed directly. Indeed, it can only be measured by looking at the extent of the movement of the option prices and deriving it mathematically from that.

The mathematical estimators are complicated also by volatility clusters, fat tails, non-normality of the distribution and structural breaks in the distribution of the returns.

We are going to use K-means and Fuzzy C-Means clustering techniques on the VIX. VIX is a real time index representing expected volatility over the coming 30 days in percentage terms based on S&P 500 index.

4 K-means

K-means clustering partitions data into K-clusters that minimise squared errors inside clusters using '*Euclidean'* distances i.e. it minimises distances from centroids to data points inside clusters.

To find the number of centroids we use the '*Elbow*' method and '*Silhouette*' analysis.

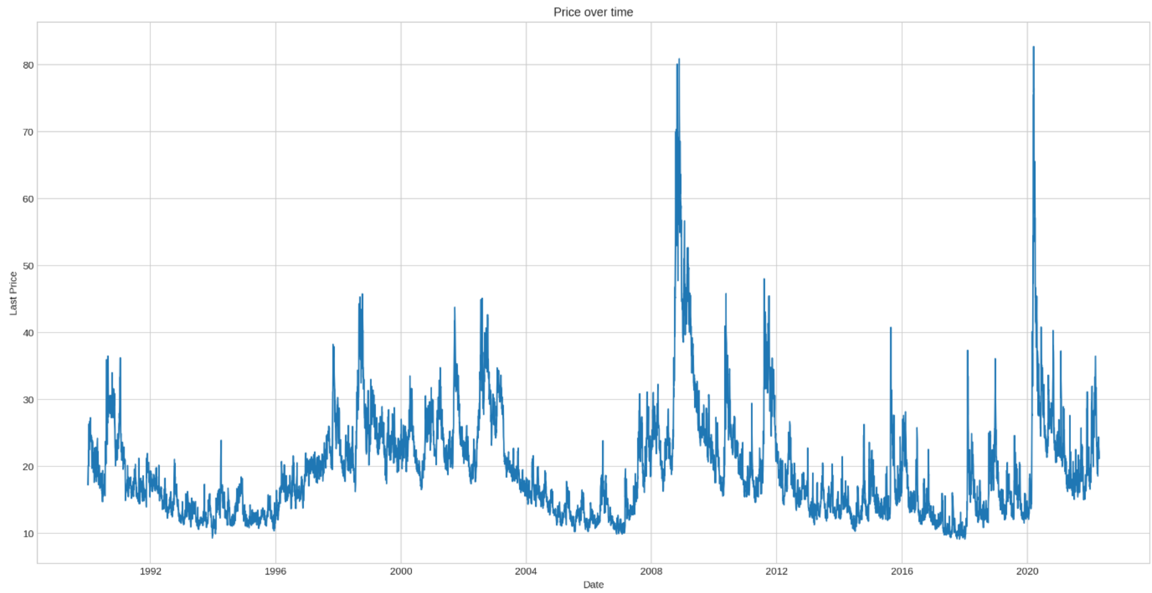

Figure 1. Plot of VIX from 1990 to 2022.

In order to normalise the data, we use '*StandardScaler*' technique. We create a normalised data set with mean 0 and standard deviation equal to 1.

As we do not know how many clusters there are, we use *'for'* loop with K in range from 1 to 10. This gives us a visual representation using '*Elbow*' method.

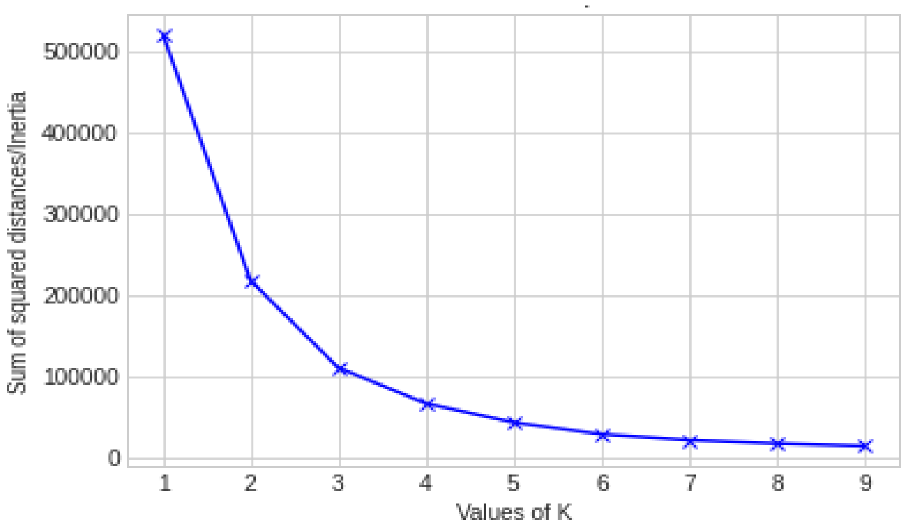

Figure 2. Elbow Method for optimal *k*

As the number of clusters increase, the error is minimised more and more. Visually, we are looking for an elbow of this curve. It appears in the region of 2–4 clusters.

The '*Silhouette*' method is another method of finding the optimal number of clusters by computing the silhouette coefficients of each point that measures how much a point is similar to its own cluster compared to other clusters. Similar to '*Elbow*' method, we train K-means clustering for each of the values of *k*. Plot of the graph shows the silhouette score on *y*-axis and the number of clusters on the *x*-axis.

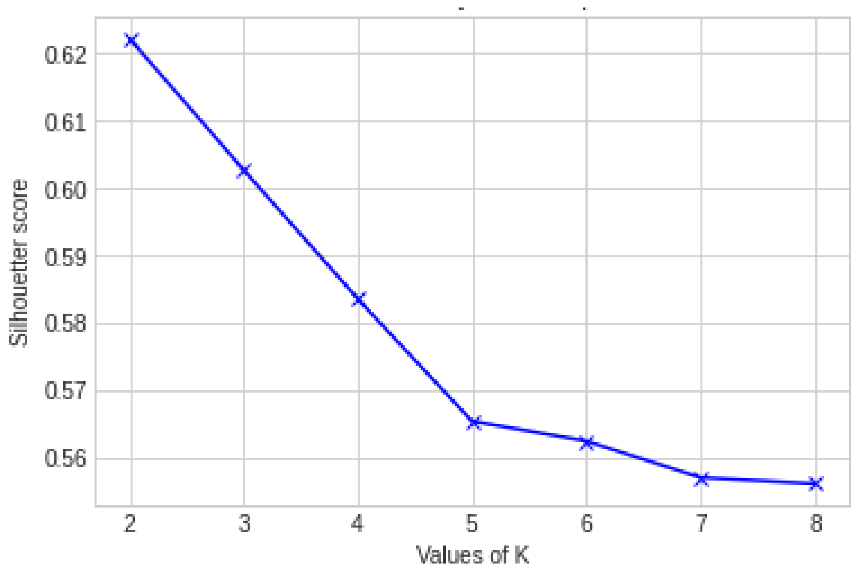

Figure 3. Silhouette analysis for optimal *k*

Prior to plotting the clusters, we find the centroids of the clusters: [26.81939959], [19.69309811], [13.48929072], [62.16890411], [38.19894207].

Next we plot the graph with 2 and 5 clusters (Figure 4). The top graph shows two clusters identified by horizontal lines and the bottom graph with 5 clusters identified also by the horizontal lines. The top graph simply splits the data into periods of low and high volatility. The graph shows when the volatility is low it tends to stay low and when it is high it tends to remain high.

The bottom graph has 5 centroids which are represented as horizontal lines. The graph splits the data into periods of very low volatility, low volatility, medium volatility, high volatility, and extremely high volatility.

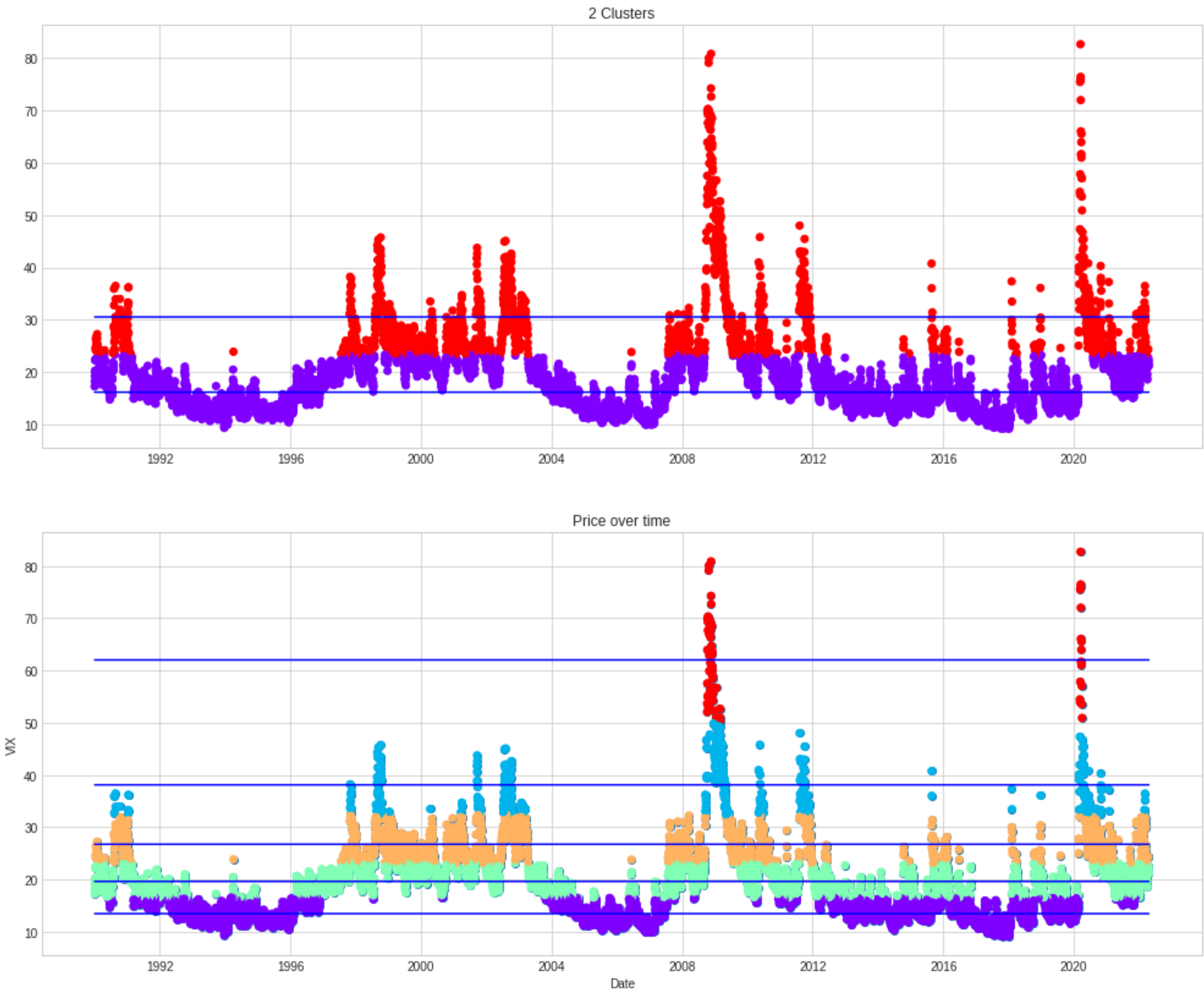

Figure 4. The graph with two clusters (top) and five clusters (bottom)

We find the silhouette score to check how well we fit the data into clusters. The silhouette score falls within the range [–1, 1]. The silhouette score of 1 means that the clusters are very dense and nicely separated. The score of 0 means that clusters are overlapping.

Our silhouette score is 0.554.

Lastly, we provide the transition matrix for the clusters. Table 1 shows how many times the VIX moved from one cluster to another. It is done in absolute terms where rows are starting cluster and columns are final cluster. For example, VIX was 3326 in Cluster 0 and stayed there. It moved 219 times in Cluster 3 from lower cluster to upper cluster. If there were no jumps, the matrix would be symmetric but it is not.

Table 1. Transition matrix for the clusters

Index		Final cluster				
		0	**1**	**2**	**3**	**4**
Starting culster	**0**	3326	0	0	219	0
	1	1	1231	67	171	0
	2	0	65	324	2	9
	3	218	174	0	2254	0
	4	0	0	9	0	65

5 Fuzzy C-Means (FCM)

We are now going to repeat the same exercise using Fuzzy C-Means (FCM) clustering. This algorithm is considered to be better than K-means because unlike K-means where the data points exclusively belong to one cluster, in FCM algorithm, the data point can belong to more than one cluster which is where *Fuzzy* methods come in. FCM assigns membership grades which indicate the degree to which data points belong to each cluster. Plotting the Fuzzy C-Means Clusters 0–4 gives us the following results (Figure 5).

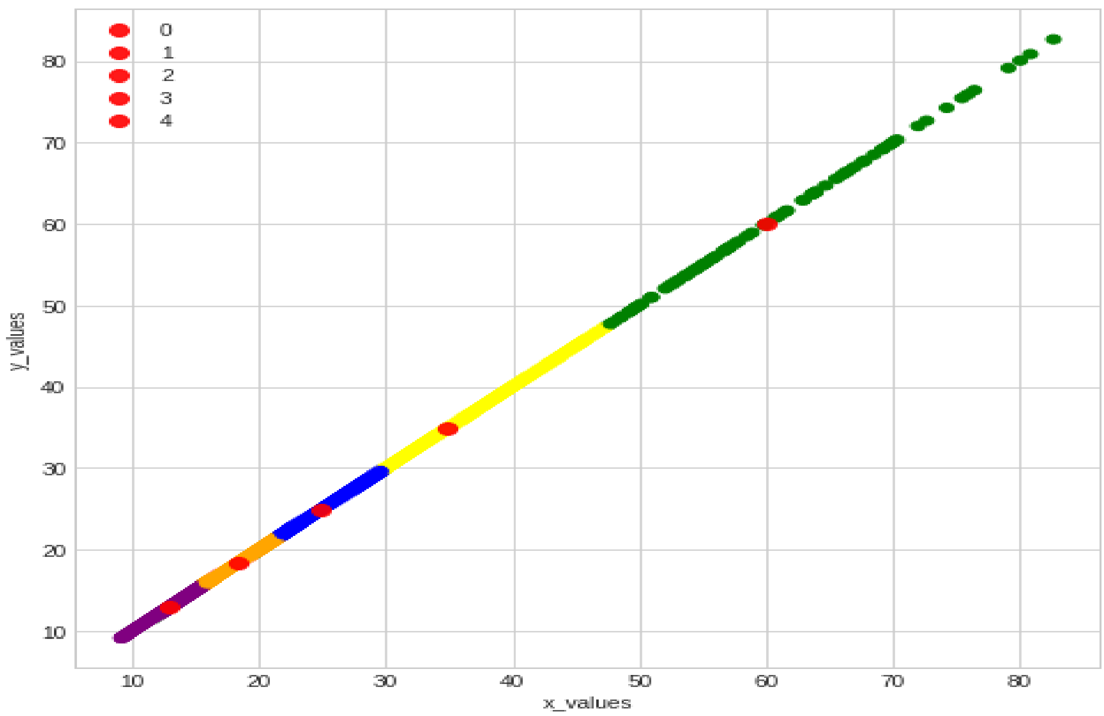

Figure 5. Plotting the Fuzzy C-Means. The red dots are centroids with clusters around them.

Plotting the best Cluster 5 gives 5 centroids which are represented as horizontal lines (Figure 6). The graph splits the data into periods of very low volatility, low volatility, medium volatility, high volatility, and extremely high volatility according to FCM algorithm.

The Silhouette Score is 0.555 which in this case is not very different to K-Means and is probably due to random state initialisation.

.

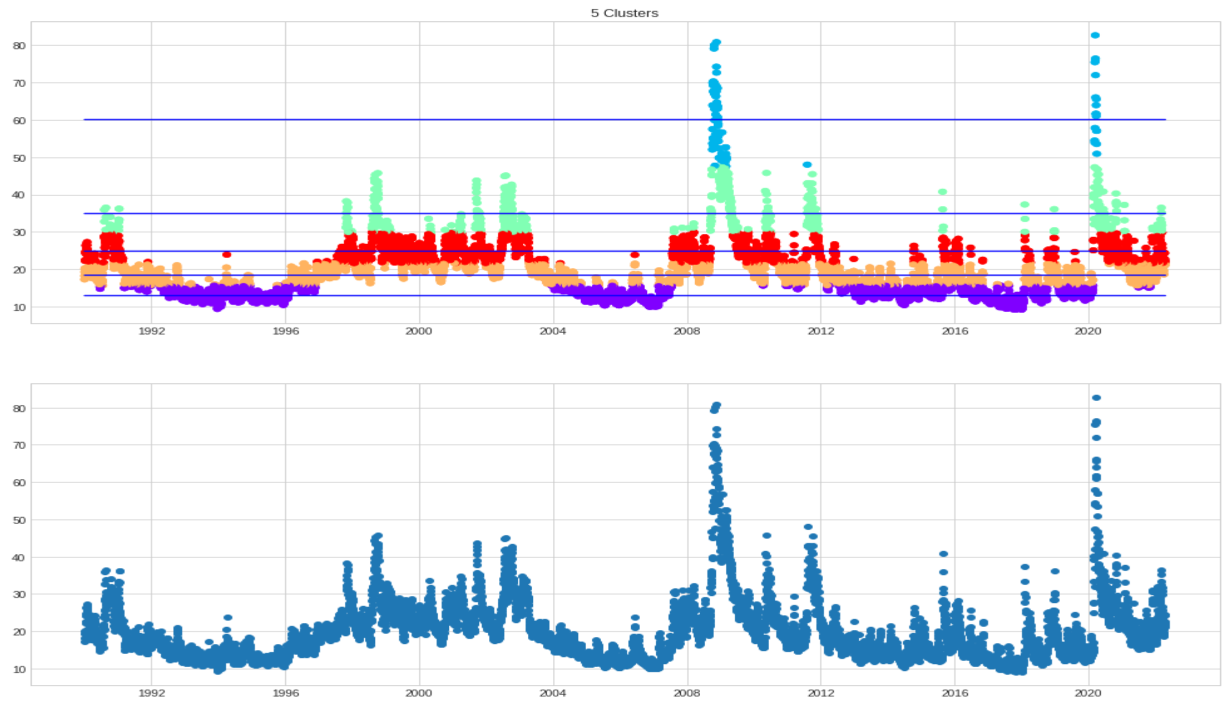

Figure 6. Volatility of Volatility based on Fuzzy C Means

6 Intuitionistic Fuzzy Sets and IFCM

The IFS non membership value is calculated as a complement of the membership value to 1. However, in reality because of uncertainty, the non-membership is not always equal to one minus the membership value. To deal with this uncertainty, Atanassov proposed another higher order fuzzy set called IFS [1–3]. An IFS $\tilde{A}$ in X is given by:

$$\tilde{A} = \left\{ \left\langle x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x) \mid x \in X \right\rangle \right\},$$

where X is a universe of discourse and $\mu_{\tilde{A}}(x) : X \longrightarrow [0, 1]$, $\nu_{\tilde{A}}(x) : X \longrightarrow [0, 1]$ with the condition $0 \le \mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) \le 1$; $\forall x \in X$ and $\mu_{\tilde{A}}(x)$, $\nu_{\tilde{A}}(x)$ denote membership and non membership degree, respectively.

For each IFS $\tilde{A}$ in X, the hesitation degree should be considered. The hesitation degree of an element $x \in X$ is defined as

$$\pi_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) - \nu_{\tilde{A}}(x),$$

where $\pi_{\tilde{A}}(x)$ is hesitation degree and should satisfy the elementary condition of intuitionism, i.e., $0 \le \pi_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) - \nu_{\tilde{A}}(x) \le 1$. In the literature, two fuzzy complements or IFS generators are used to construct intuitionistic fuzzy set: Sugeno's and Yager's [10]. The fuzzy complement function is defined as:

$$N(\mu(x)) = g^{-1}(g(1) - g(\mu(x)),$$

where $g(.)$ is an increasing function and $g : [0, 1] \to [0, 1]$.

Yager's class can be generated by using the following function:

$$g(x) = x^{\lambda}.$$

Non-membership values are calculated from Yager's intuitionistic fuzzy complement $N(x)$. The IFSs using Yager's intuitionistic fuzzy complement become

$$A_{\lambda}^{IFS} = \left\{ x, \mu_A(x), (1 - \mu_A(x))^{\frac{1}{\lambda}} \ \middle| \ x \in X \right\}.$$

Non-membership values are calculated from Sugeno's intuitionistic fuzzy complement $N(x)$. IFS constructed using Sugeno's fuzzy complement is as follows:

$$A_{\lambda}^{IFS} = \left\{ x, \mu_A(x), \frac{1 - \mu_A(x)}{1 + \lambda \mu_A(x)} \ \middle| \ x \in X \right\}.$$

In soft clustering methods, the membership value is computed based on a distance function [11–17]. So distance metric plays an important role. In the literature, many distance metrics are proposed developed similarity measures of IFSs based on Hausdorff distance. Povisiona results show that Hausdorff distance is simple and works better than other distance metrics. Hence, there is a need to take advantage of IFS and Hausdorff distance to increase the cluster's density and thus seperability.

7 Conclusions

Central Banks in particular pay very close attention to volatility and volatility of volatility. In fact, we know that most Central bank models include volatility of volatility as a factor of how loose or tight the financial conditions are. Hence, it has direct impact on things that most people care about such as their mortgage rates, inflation and unemployment. With a very few exceptions, unlike pure sciences and math, systems in economics exhibit inter-dependencies, hesitation and reflexivity i.e. participation in experiments influences its outcomes and vice versa. To this extent IFS based unsupervised techniques for data analysis seem to have an advantage.

References

[1] Atanassov, K. T. (1983). Intuitionistic fuzzy sets, *VII ITKR's Session*, Sofia, 1983 (Deposed in Central Science–Technology Library of Bulgaria Academy of Science –1697/84).

[2] Atanassov, K. T. (1999). Intuitionistic Fuzzy Sets: Past, Present and Future. In: Wagenknecht, M., & Hampel, R. (eds) *3rd Conference of the European Society for Fuzzy Logic and Technology*, Zittau, Germany (10.09.2013–12.09.2003), 12–19.

[3] Atanassov, K. T. (1999). Intuitionistic Fuzzy Sets: Theory and Applications. Physica-Verlag, New York.

[4] Bucci, A. (2020). Realized Volatility Forecasting with Neural Networks. *Journal of Financial Econometrics*, 18(3), 502–531.

[5] Demeterfi, K., Derman, E., Kamal, M., & Zou, J. (1999). *More Than You Ever Wanted to Know about Volatility Swaps*. Goldman Sachs Quantitative Strategies Research Notes, March 1999.

[6] Francq, C., & Zakoian, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. 1st edition ed. Chichester: John Wiley & Sons.

[7] Poon, S.-H., & Granger, C. W. J. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, 41(2), 478–539.

[8] Pradeepkumar, D., & Ravi, V. (2017). Forecasting financial time series volatility using Particle Swarm Optimisation trained Quantile Regression Neural Network. *Applied Soft Computing*, 58, 35–52.

[9] Site, A., Birant, D., & Isik, Z. (2019). Stock Market Forecasting Using Machine Learning Models. *IEEE 2019 Innovations in Intelligent Systems and Applications Conference (ASYU)* - Izmir, Turkey (31.10.2019 – 02.11.2019), 318–323.

[10] Sugeno, M., & Terano, T. (1977). A model of learning based on fuzzy information, *Kybernetes*, 6, 157–166.

[11] Szmidt, E., & Kacprzyk, J. (1997). On measuring distances between intuitionistic fuzzy sets. *Notes on Intuitionistic Fuzzy Sets*, 3(4), 1–3.

[12] Szmidt, E., & Kacprzyk, J. (2000). Distances between intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 114(3), 505–518.

[13] Szmidt, E., & Kacprzyk, J. (2001). Intuitionistic fuzzy sets in some medical applications. *Notes on Intuitionistic Fuzzy Sets,* 7(4), 58–64.

[14] Van Lung, H., & Kim, J.-M. (2009). A generalized spatial fuzzy *C*-means algorithm for medical image segmentation. In: FUZZ-IEEE'09: Proceedings of the 18th international conference on Fuzzy Systems, Jeju Island, Korea, 20.08.2009 – 24.08.2009, 409–414.

[15] Wang, W., & Zhang, Y. (2007). On fuzzy cluster validity indices. *Fuzzy Sets and Systems*, 158(19), 2095–2117.

[16] Wang, Z., Xu, Z., Liu, S., & Tang, J. (2011). A netting clustering analysis method under intuitionistic fuzzy environment. *Applied Soft Computing*, 11(8), 5558–5564.

[17] Wang, Z., Xu, Z., Liu, S., & Yao, Z. (2014). Direct clustering analysis based on intuitionistic fuzzy implication. *Applied Soft Computing*, 23, 1–8.