INTERNATIONAL JOURNAL OF ENHANCED RESEARCH IN MANAGEMENT AND COMPUTER APPLICATIONS
VOL. 2, ISSUE 2, FEB. 2013 ISSN NO: 2319-7471

Fuzzy Clustering Techniques for Optimum Currency Area Variables

Dr. Vijaynt Agarwal[1], Aastha Chhabra[2], Krit Gupta[3]
[1,2,3]Netaji Subhas Institute of Technology, Dwarka, New Delhi
[1]vijayantonly@yahoo.com

Abstract: Through this paper, we attempt to cluster the various Economic and Monetary Union (EMU) according to the different Optimum Currency Area (OCA) variables and demonstrate their similarity to various other countries. We utilize three different fuzzy clustering techniques for the process; thus, determining the different clusters these countries should be grouped in as suggested by the OCA variables. The reason for taking fuzzy clustering is that fuzzy clustering allows the data to be part of one or more clusters as opposed to hard clustering which limits the data to only one cluster. This study extends the studies of Itir Ozer and Ibrahim Ozkan (2008) .We use the same data for the analysis. We also carried out the cluster validity utilizing four different indices to get the accurate number of clusters and Principal Component Analysis (PCA) to get a more comprehensive view of the data.

Keywords: Clustering, Cluster Validity, Optimum Currency Area, Principal Component Analysis.

1. INTRODUCTION

The appropriate area of a common currency is given by the Optimum Currency Area (OCA) theory and it tries to find out the unified currency's suitable realm. The OCA variables give a respectable idea about the concept of a unified currency. The OCA parameters that we use in this study are synchronization in business cycles, volatility in the real exchange rates, synchronization in the real interest rates, degree of trade integration and convergence of inflation.

This study extends the studies of Itir Ozer and Ibrahim Ozkan (2008) [1] .We use the same data for the analysis (Table I). As per them, Industrial production series and the real interest rates have been de-trended with the Hodrick-Prescott (H-P) filter with the smoothing parameter set at 50,000 in the calculation of synchronization in business cycles and synchronization in the real interest rates. While their analysis was done using only one clustering technique, namely, Fuzzy C-Means, we extend the analysis by applying three different clustering techniques namely, Fuzzy C-Means (FCM) clustering [2], Gustafson–Kessel (GK) clustering [3] and Gath–Geva (GG) [4] clustering on the data collected keeping in mind that the clustering techniques are non-linear and probabilistic in nature [8].

We also carried out the cluster validity to get the accurate number of clusters. While Ozer and Ozkan grouped their data into 4 clusters, the optimum number of clusters as calculated by us came out to be 3 for FCM and GK clustering and 4 for GG clustering.
Ozer and Ozkan (2008) [1] tried to gain various insights from the OCA variables by studying the performance of the countries for different OCA variables. For this, they separately took combinations of different number of variables and provided the best results.
We, on the other hand, take all the five variables in our analysis and cluster the countries according to all the variables, thus, getting an overall picture and present the results.

Assuming Germany as the center country, the study uses five different OCA variables. The analysis was done on fifteen EMU countries, eight non-EMU European Union (EU) countries, Croatia, which is set to become a member of the EU in 2013, Canada (North America) and Japan (Asia). Three clustering techniques were employed on the dataset calculated by Ozer and Ozkan in their study (2008) [1] and clusters were obtained based on the similarity these countries have with each other according to the selected variables.

We further analyze by carrying out the Principal Component Analysis (PCA) on the data. PCA defines various principal components based on the set of the variables of the data.
To group the data in a meaningful way, we also carry out the cluster validity. We cluster data using different values of c, that is, the different number of clusters and utilize different validity measures to predict the efficiency of the obtained partition.
Since none of the validity measure is complete and accurate in itself, we use a combination of four different indices, namely, Partition Coefficient (PC), representing the overlapping between clusters; Classification Entropy (CE), measuring the fuzziness of the partition; Dunn's Index (DI), identifying compact and well separated clusters, and Alternative Dunn Index (ADI), modifying the original Dunn's Index.

The theories are explained in the section 2, with the results in section 3. The explanation of the results is given in section 4.

VOL. 2, ISSUE 2, FEB. 2013 ISSN NO: 2319-7471

2. DATA AND METHODOLOGY

2.1 OCA CRITERIA

Following the data presented by the studies done by Itir Ozer and Ibrahim Ozkan [1] and assuming Germany, the strongest economy in the EMU, as the center country, we take five different OCA variables, as defined in [1]:

1. Synchronization in business cycles - The cross-correlation of the cyclical component of the deseasonalised industrial production series, de-trended with an application of Hodrick-Prescott (H-P) filter with the smoothing parameter set at 50,000 have been calculated with respect to Germany and made to fall between 0 and 2.
2. Volatility in the real exchange rates - Standard Deviation of the log-difference of real bilateral exchange rates before 1999 are taken. After 1999, Euro has been used.
3. Synchronization in the real interest rates - The cross-correlation of the cyclical components of the real interest rate cycle of a country with respect to Germany have been calculated and then de-trended with H-P filter.
4. Degree of trade integration - This has been measured by

   $$(x_i^{EU\text{-}25} + m_i^{EU\text{-}25}) / (x_i + m_i)$$

   where x_i and m_i are exports and imports (of goods) respectively, of country i, and superscript EU-25 represents European Union countries as of May 2004.
5. Convergence of inflation – This has been measured by e_i-e_g, where e_i and e_g are the rates of inflation in country i and Germany, respectively.

The data as given is presented in Table 1.

TABLE I: Countries v/s the 5 OCA Variables [1]

	Synchronization in Business Cycles	Volatility in the Real Exchange Rates	Synchronization in the Real Interest Rates	Degree of Trade Integration	Convergence of Inflation
Austria	0.0965	0.0046	0.3633	76.38	0.3436
Belgium	0.2821	0.0121	0.5183	75.52	0.8296
Croatia	0.9736	0.0253	1.6042	67.49	1.3846
Cyprus	0.8874	0.00470	0.4384	63.82	0.6046
Czech Republic	0.9351	0.0129	1.2068	80.03	-0.108
Denmark	0.4276	0.0046	0.0497	70.49	-0.1454
Finland	0.2459	0.0044	0.5648	62	-1.0923
France	0.4427	0.0028	0.5049	66.83	-0.2098
Germany	0	0	0	62.96	0
Greece	0.3882	0.0047	1.1608	57.33	1.6073
Hungary	0.1536	0.0206	1.487	75.27	1.5975
Ireland	0.6647	0.0046	0.6415	2.750	0.4617
Italy	0.4642	0.00310	0.5366	59.610	0.0313
Luxembourg	0.5957	0.0111	0.262	81.54	0.536
Netherlands	0.6107	0.0042	0.4927	66.98	-0.2906
Norway	0.7397	0.0342	0.2538	75.46	-0.4319
Poland	0.3994	0.0261	0.3448	76.36	0.1528
Portugal	1.111	0.0053	0.5307	78.20	0.3397
Romania	0.9328	0.0338	0.5271	71.61	7.0354
Slovak Republic	0.6833	0.0145	1.3264	83.13	0.7549
Slovenia	0.3025	0.0067	0.8753	74.16	0.525
Spain	0.5056	0.0033	0.2794	69.25	1.4138
Sweden	0.4373	0.0123	0.3722	67.49	-1.5007
Turkey	0.5966	0.0672	0.4547	49.81	6.2252
United Kingdom	0.317	0.0161	1.0504	53.380	0.8768
Canada	0.3371	0.0241	0.4975	8.380	0.2802
Japan	0.3931	0.0252	1.0198	14.43	-2.2271

2.2 FUZZY CLUSTERING TECHNIQUES

2.2.1 FUZZY C-MEANS (FCM) CLUSTERING

FCM Clustering is a form of fuzzy clustering. In fuzzy clustering, each point has a degree of belonging to numerous clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Hence, the points at the edge of the cluster may have a lesser degree than points in the center of the cluster. FCM Clustering minimizes the function called as C-Means functional (1) as given in [11]:

$$J(X;U,V) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^{m}||x_k - v_i||_A^2 \quad (1)$$

where

$$V = [v_1, v_2, \ldots\ldots, v_c], \quad v_i \in R^n \quad (2)$$

Equation (2) is a vector of cluster prototypes, which have to be determined, μ_{ik} gives the elements of the partition matrix U, x_k represents the data set, m the weighting exponent and A gives the subset.

$$D_{ikA}^2 = ||x_k - v_i||_A^2 = (x_k - v_i)^T A (x_k - v_i) \quad (3)$$

Equation (3) is a squared inner product distance norm.

The weighted mean, v_i, of the N data items (4) that belong to a cluster [11], where the weights are the membership degrees, is given by

$$v_i = \frac{\sum_{k=1}^{N}\mu_{ik}^{m}x_k}{\sum_{k=1}^{N}\mu_{i,k}^{m}}, \quad 1 \le i \le c \quad (4)$$

The disadvantage of FCM Clustering is that it can only identify circular clusters, generically, clusters with same shape and alignment.

2.2.2 THE GUSTAFSON-KESSEL (GK) CLUSTERING

The GK Clustering is an updated version of the FCM Clustering technique. It uses an adaptive distance norm, which enables it to detect clusters of different geometrical shapes [9]. Hence, it overcomes the shortcoming of the FCM Clustering technique.
The inner-product norm [11] used in GK Clustering is given in (5) as:

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i), \quad 1 \le i \le c, \; 1 \le k \le N \quad (5)$$

Therefore, in GK Clustering technique, each cluster has its own matrix A_i.

The objective function [11] of the GK Clustering is rewritten in (6):

$$J(X;U,V,A) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^{m}D_{ikA_i}^2 \quad (6)$$

and hence, the fuzzy covariance matrix of the i-th cluster F_i [11] is defined by (7) as:

$$F_i = \frac{\sum_{k=1}^{N}(\mu_{ik})^{m}(x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^{N}(\mu_{ik})^{m}} \quad (7)$$

The GK Clustering technique utilizes altering distance norms and hence, is capable of detecting clusters of dissimilar geometrical shapes, and so, provides more reality to the analysis.

2.2.3 THE GATH-GEVA (GG) CLUSTERING

The GG Clustering technique is a modified version of the GK Clustering technique [10]. F_{wi}, that is, the fuzzy covariance matrix of the i-th cluster [11], is given by (8):

$$F_{wi} = \frac{\sum_{k=1}^{N}(\mu_{ik})^{w}(x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^{N}(\mu_{ik})^{w}}, \quad 1 \le i \le c \quad (8)$$

Here, w is the weighting exponent, which enables the generalization of the expression. We use w = 2, so that partition becomes more fuzzy to compensate the exponential term of the distance norm [11]. The benefit of this approach is that clusters are not limited by

VOL. 2, ISSUE 2, FEB. 2013 ISSN NO: 2319-7471

volume. So, GG Clustering technique can detect clusters of not only different shapes, sizes but also of varying densities. However, it is less robust and needs a good commencing to avoid it to converge to a local optimum.

2.3 PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is a mathematical technique that converts a number of correlated variables into a number of uncorrelated variables called principal components. What is often seen is that a group of variables move together in a set of data with various variables. Hence, what is done is that a group of similarly moving variables is clubbed together and replaced by a new variable.

PCA creates a vector space with each variable as a vector and each principal component as a linear combination of the original variables [11].

Therefore, The number of useful principal components is less than the original variables. So, PCA fulfills the function of reducing the scale of the data set as well as identifies novel underlying variables.

In this work, we have used the second objective, where, the covariance matrix $\boldsymbol{F}$ [11] is given in (9) as:

$$\boldsymbol{F} = \frac{1}{N}(\boldsymbol{x_k} - \boldsymbol{v})(\boldsymbol{x_k} - \boldsymbol{v})^T \tag{9}$$

where $\boldsymbol{v}$ is the mean of the data and $\boldsymbol{N}$ is the number of objects in the data set. The eigenvector having the largest value is in the same direction as the first principal component, and so, the first principal component accounts for as much of the variability in the data as possible.

The eigenvector associated with the second largest eigenvalue determines the direction of the second principal component. Principal Components are necessarily independent when the data set is jointly normally distributed.

In our analysis, we have used the Fuzzy Sammon mapping [5], which takes every cluster to be a point, independently to the form of the original cluster prototype.

It provides a highly comprehensive view of data having more than 3 variables. Projecting the variables on the 1st component v/s the 2nd component graph (fig.6) does further analysis. Percentage of variance produced is also plotted as a function of the principal components (fig.7).

2.4 CLUSTER VALIDITY

Calculating validity is the problem, which analyzes whether a given fuzzy break-up appropriately fits to the entire data. Many times, it so happens that the number of fixed clusters is wrongly attributed or the cluster shape might not correspond to the groups of data. Hence, Cluster validity determines the correct number of clusters required for a particular data and gives the grouping a meaning. Numerous scalar validity measures have been proposed in the theory. Since, none of them is complete in themself, hence, we use four different indices [11] as given below:

1. Partition Coefficient (PC), representing the overlapping between clusters;
2. Classification Entropy (CE), measuring the fuzziness of the partition;
3. Dunn’s Index (DI), identifying compact and well separated clusters, and
4. Alternative Dunn Index (ADI), modifying the original Dunn’s Index.

3. RESULTS
3.1 CLUSTERING RESULTS
3.1.1 FCM CLUSTERING RESULTS

As determined by the validity indices (PC and CE) in fig.4, the optimum number of clusters for FCM is 3. We set the number of clusters as 3. The results of the clustering are shown in Table I, with the corresponding plot in fig.1. In the plot, the blue dots indicate the various countries while the red dots show the cluster centers. Similarly, for GK Clustering, the optimum number of clusters came out to be 3. The clusters are as shown in Table III, with the corresponding plot in fig.3. For GK Clustering, the validity analysis gave us the number of clusters as 4. The results are shown in Table IV, with the corresponding plot in fig.4.

TABLE II: Fuzzy C Means Clustering Results (c =3)

Cluster I	Cluster II	Cluster III
Croatia	Austria	Ireland
Czech Republic	Belgium	Turkey
Greece	Cyprus	Canada
Hungary	Denmark	Japan
Portugal	Finland	
Romania	France	
Slovak Republic	Germany	
United Kingdom	Italy	
	Luxembourg	
	Netherlands	
	Norway	
	Poland	
	Slovenia	
	Spain	
	Sweden	

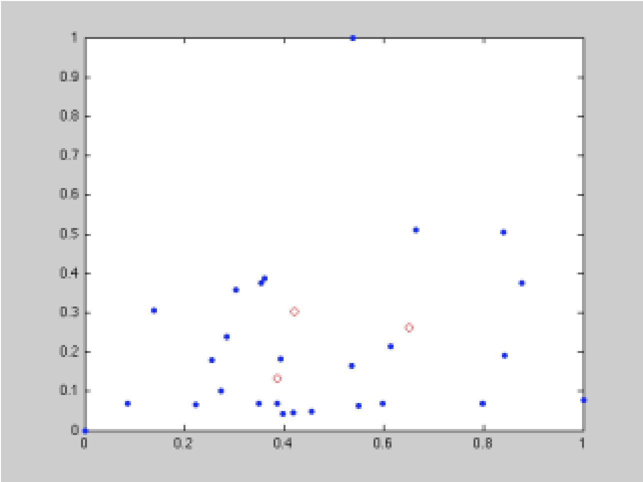

Figure 1. Fuzzy C-Means Clustering Results

3.1.2 GK CLUSTERING RESULTS

TABLE III: Gustafson-Kessel Clustering Results (c=3)

Cluster I	Cluster II	Cluster III
Canada	Finland	Austria
Japan	France	Belgium
	Italy	Croatia
	Luxembourg	Cyprus
	Norway	Czech Republic
	Portugal	Denmark
	Turkey	Germany
	Sweden	Greece
	Canada	Hungary
	Japan	Ireland
		Netherlands
		Poland
		Romania
		Slovak Republic
		Slovenia
		United Kingdom
		Spain

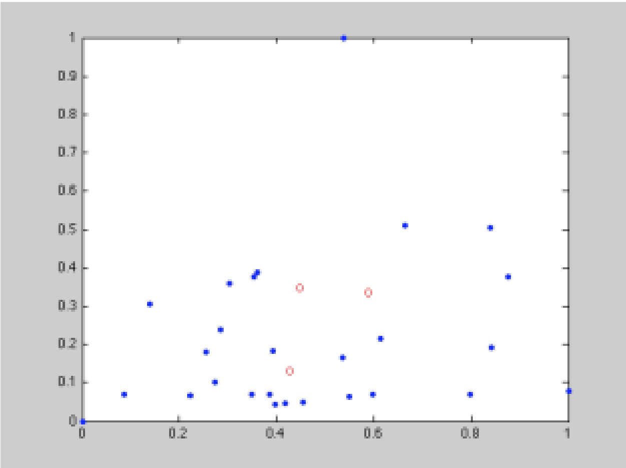

Figure 2. Gustafson-Kessel Clustering Results

3.1.3 GG CLUSTERING RESULTS

TABLE IV: Gath-Geva Clustering Results (c=4)

Cluster I	Cluster II	Cluster III	Cluster IV
Croatia	Cyprus	Ireland	Austria
Czech Republic	Luxembourg	Canada	Belgium
Hungary	Norway	Japan	Denmark
Slovak Republic	Poland		Finland
	Portugal		France
	Romania		Germany
	Spain		Greece
	Sweden		Italy
	Turkey		Netherlands
			Slovenia
			United Kingdom

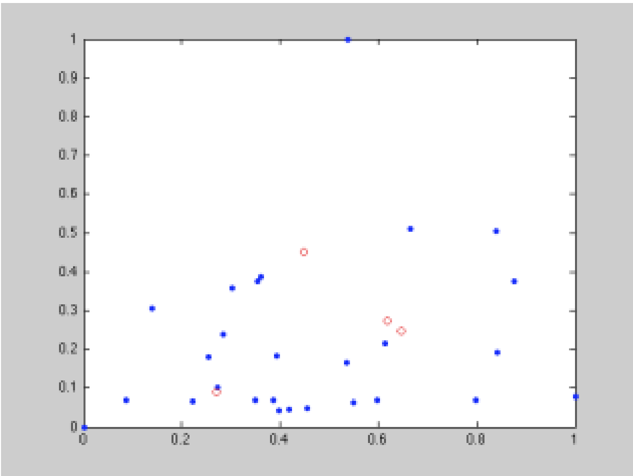

Figure 3. Gath-Geva Clustering Results

3.2 VALIDITY RESULTS

3.2.1 VALIDITY RESULTS WITH RESPECT TO VARIOUS CLUSTERING TECHNIQUES

Table V gives the results of the values of the various coefficients namely, PC, CE, DI and ADI for all the clustering techniques used.

TABLE V: The 4 validity indices v/s the clustering techniques

	PC	CE	DI	ADI
FCM	0.4473	1.0211	0.1213	0.0113
GK	0.9998	0.0010	0.1683	0.0435
GG	0.9752	NAN	0.1705	0.0025

3.2.2 VALIDITY MEASUREMENTS

Table VI gives the various values of the validity coefficients for the different values of the number of clusters.

VOL. 2, ISSUE 2, FEB. 2013 ISSN NO: 2319-7471

TABLE VI: The number of clusters (c) v/s the 4 validity indices

c	2	3	4	5	6
PC	0.7554	0.74403	0.99976	0.870381	0.9999
CE	0.3887	0.40563	0.00103	0.233486	1.97350
DI	0.07017	0.11005	0.16830	0.07017	0.15815
ADI	0.0205	0.1731	0.04354	0.00210	4.0180e-07

c	7	8	9	10
PC	0.8888	0.9998	0.999	0.9681
CE	0.212624	0.00062	0.00278	0.066019
DI	0.1547	0.08150	0.17972	0.07172
ADI	0.01165	5.38263e-08	1.96404e-08	2.54229e-08

c	11	12	13	14
PC	0.9896	0.98096	0.8946	0.9608
CE	0.02694	0.028190	0.2244	0.067787
DI	0.19231	0.244991	0.079873	0.0843712
ADI	6.2828e-10	2.66408e-09	2.37315e-07	3.050820e-09

Fig.4 shows the optimum number of clusters as given by the validity coefficients, PC and CE [6]. As can be seen, the first minimum in the plot of PC comes out at 3. Hence, the optimum number of clusters is 3. CE is the mirror image of PC. Since, none of the index is reliable in itself, we check our analysis by taking two more coefficients, namely DI and ADI. The values of DI and ADI verify our results as shown in fig.5.

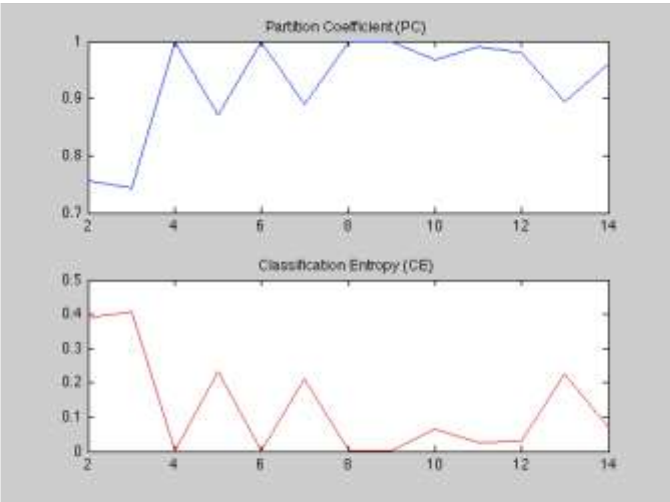

Figure 4. Validity Measurement (PC and CE)

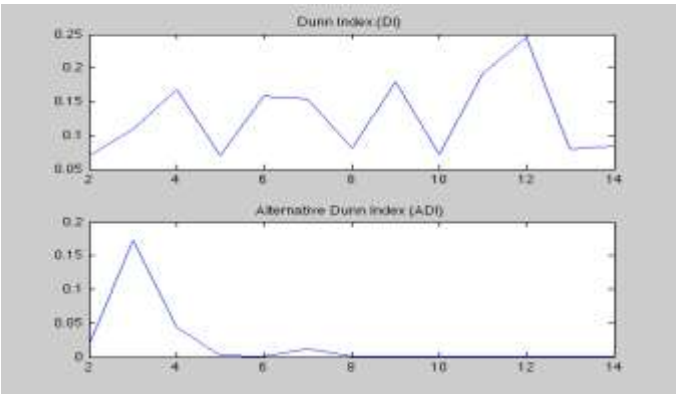

Figure 5. Validity Measurement (DI and ADI)

VOL. 2, ISSUE 2, FEB. 2013 ISSN NO: 2319-7471

3.3 PCA RESULTS

Fig 6. plots the values of the first and the second principal components, whereas, fig.7 expresses the variability as a percentage.

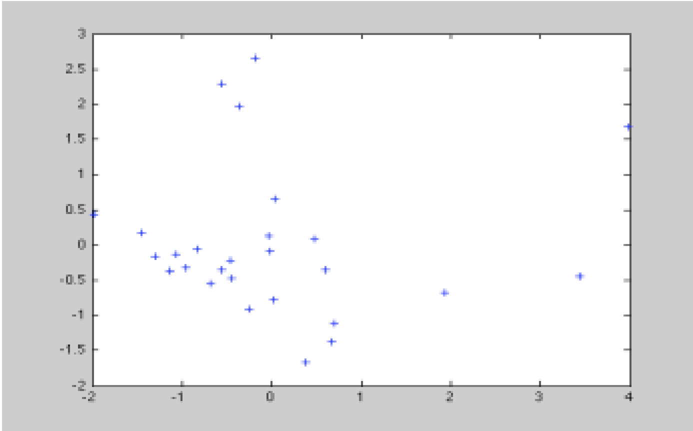

Figure 6. First Component (X-axis) v/s Second Component (Y-axis)

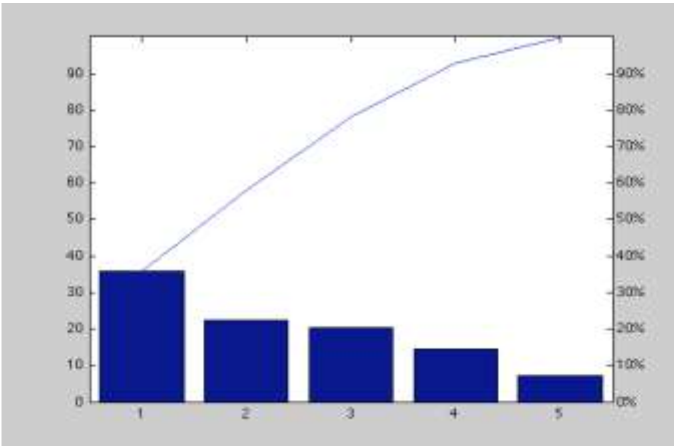

Figure 7. Percentage Variability explained by each component

Fig.8 shows the 5 OCA variables as a function of the first and second principal components in the 2D vector space [7]. Fig. 9 shows the same variables in the 3D vector space as plotted against the first 3 principal components. Red dots depict the various countries, whereas the OCA variables are shown in blue.

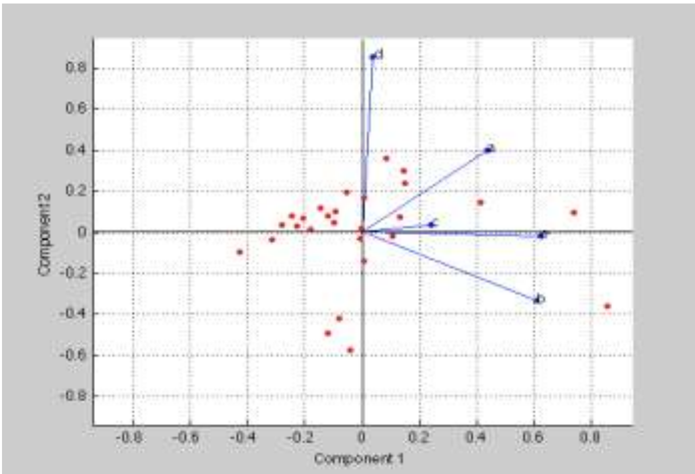

Figure 8. The vector space with all the 5 OCA variables (2D)

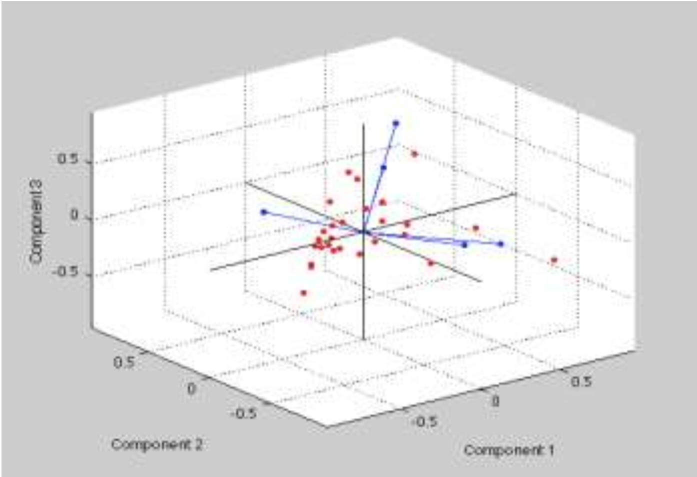

Figure 9. The vector space with all the 5 OCA variables (3D)

4. EXPLANATION OF RESULTS

The results showed that there were various discrepancies in the countries, not only between the countries from the different backgrounds, but even countries in the EMU did not form a common cluster.

In the FCM clustering technique, eleven EMU countries made a cluster, whereas three EMU countries were identified in a separate cluster with Croatia, Czech Republic, Hungary, Romania, UK. Ireland, Canada, Japan and Turkey made their separate cluster.

In the GK clustering technique, the maximum EMU countries in a cluster were ten, whereas, the other EMU countries were distributed among the different clusters. Japan and Canada formed a separate cluster.

Five EMU countries made a cluster with Norway, Turkey and Sweden also having some resemblance to Canada as well as Japan.

In the GG clustering technique, nine EMU countries were spotted in a cluster, whereas four EMU countries formed a separate cluster showing more similarity with countries of the European Union. Japan, Canada and Ireland formed a separate cluster, showing an improvement over the result of the GK clustering and Croatia, Czech Republic, Hungary and Slovak Republic formed a separate cluster.

In the PCA analysis, the 35.89% of variability is explained by the 1st principal component, the 2nd principal component explains 22.23% of variability, while the percentages for the 3rd, 4th, and 5th principal components were 20.21%, 14.37% and 7.28% respectively.

Hence, we conclude that the differences defined by the OCA variables in the different countries of the world are too extensive to adopt a common currency for the world. So, it would be impractical to adopt a common global currency. True single-currency candidates require perfect likeness and strong economic ties that are not found amongst the real market economies.

However, the countries as suggested by the clustering analysis show more likeliness and have a higher tendency of being grouped as countries having a common currency.

REFERENCES

[1]. Itir Ozer, Ibrahim Ozkan, "Practical insights from OCA variable combinations", Economic Annals, vol. 53, 2008.
[2]. J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Algorithms", Plenum Press, 1981.
[3]. R. Babuska, P.J. van der Veen, U. Kaymak, "Improved co-variance estimation for Gustafson-Kessel clustering", in Proceedings of 2002 IEEE International conference on Fuzzy Systems, pp. 1081-1085, 2002.
[4]. I. Gath, A. B. Geva, "Unsupervised Optimal Fuzzy Clustering" in IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 7, pp 773-781, 1989.
[5]. A Kovacs. J Abonyi,"Vizualization of fuzzy clustering results by modified sammon mapping", in Proceedings of the 3rd International Symposium of Hungarian Researchers on Computational Intelligence, pp. 177-188, 2002.
[6]. A. M. Bensais, L. O. Hall, J. C. Bezdek, L. P. Clarke, M. L. Silbiger, J. A. Arrington, R. F. Murtagh, "Validity-guided (re)clustering with applications to image segmentation", in IEEE Tansactions on Fuzzy Systems, pp. 112-123, 1996.
[7]. Juha Vesanto, "Neural Network Tool for Data Mining: SOM Toolbox", in Proceedings of Symposium on Tool Environments and Development Methods for Intelligent Systems (TOOLMET2000), pp. 184-196, 2000.
[8]. F. Hoppner, F. Klawonn, R. Kruze, T. Runkler, "Fuzzy Cluster Analysis", Wiley, Chichester, 1999.
[9]. D. E. Gustafson, W. C. Kessel, "Fuzzy clustering with fuzzy covariance matrix", in The Proceedings of the IEEE CDC, San Diego, pp. 761-766, 1979.
[10]. J. C. Bezdek, J. C. Dunn, "Optimal fuzzy partitions: A heuristic for estimating the parameters in a normal distributions", in IEEE Transactions on Computers, pp. 835-838, 1975.
[11]. Balazs Balasko, Janos Abonyi, Balazs Feil, "Fuzzy Clustering and Data Analysis Toolbox".