

Using Hierarchical Clustering Algorithms for Turkish Residential Market

Ali HEPŞEN (Corresponding Author)

Department of Finance, Faculty of Business Administration

Istanbul University, Istanbul, 34000, Turkey

Tel: 90-212-4737070 ext. 18321 E-mail: alihepsen@yahoo.com

Metin VATANSEVER

Department of Math and Stats, Faculty of Arts and Science

Yildiz Technical University, Istanbul, 34000, Turkey

E-mail: vatansevermetin@gmail.com

Received: October 8, 2011

Accepted: October 20, 2011

Published: January 1, 2012

doi:10.5539/ijef.v4n1p138

URL: <http://dx.doi.org/10.5539/ijef.v4n1p138>

Abstract

Clustering has a potentially important contribution to real estate portfolio analysis. In this study several hierarchical clustering algorithms are applied to rental returns for seventy-one metropolitan residential markets in Turkey. The aim is to develop homogeneous groupings for real estate portfolios. The historical clustering algorithms documented in this study provides a useful guideline for real estate investors to select appropriate market areas and formulate efficiently diversified investment portfolios. The empirical findings support the three-cluster partition of the districts that reveals a clear rental return distinction of residential markets in Turkey. Cluster 1 is composed of twenty nine districts, which have the lowest rental return levels over the period of 2007:M6 to 2011:M6. Thirty four districts are grouped in Cluster 2. The cities in this group have relatively higher rental returns. The rest eight cities belong to Cluster 3. Rental return levels are distinctively higher than the other two groups. On the other hand, high rental returns are associated with higher levels of risk (standard deviation), and vice versa.

Keywords: Clustering algorithm, Time series data mining, Residential market, Rental returns

JEL Classification: E30, E44, L74, R10, R20, R30

1. Introduction

The term real estate can be defined from three different perspectives: the physical, the legal and the financial economic view. The physical concept defines real estate as a three dimensional structure of walls, ceilings, and floors. From the legal point of view, real estate can only be regarded as a building, stationary and fixed at a certain location, in combination with a parcel and the assigned rights. From the financial economic view, real estate describes a considerable investment vehicle for private, commercial and institutional investors (Geltner et.al, 2007). Investments in real estate reveal different features compared to conventional assets like stocks and bonds. In particular, this applies to long term investment horizons and is recognizable by low correlations and a distinctive risk/ return structure, which in turn is accountable for being classified as an alternative asset. With respect to issues of asset allocation, investments in real estate provide remarkable potential for diversifying an investor's portfolio.

Real estate market can be classified into various sub-markets using various variables including housing type, housing tenure, social-economic status, location and so on. However, price capitalizes most of the characteristics. For this reason, price is an important variable to classify the whole market into sub-one. Most of the studies focused on discovering the relationship between house prices in different geographic areas and demonstrating how portfolio risk can be reduced by diversifying across various geographic categories. A few studies developed groupings by applying cluster analysis to a set of rental returns. The main purpose of this study is to develop homogeneous sub-markets for 71 metropolitan residential markets in Turkey by using several hierarchical clustering algorithms (average, centroid, complete, median, single, ward and weighted) that are applied to rental returns. The remainder of this paper proceeds as follows. Section two reviews the literature related to cluster algorithm in real estate. Section three introduces the selected data and outlines the theoretical framework. Section four describes the development of the clustering model and the results of this study. The final section is the conclusion.

2. Literature Review

Time series data mining has attracted great attention in math, economics, finance and other application domains (Elton and Gruber, 1971; Bradley, Fayyad and Reina, 1998; Aggarwal, Hinneburg and Keim, 2001; Chis and Dumitrescu, 2002; Keogh and Kasetty, 2002; Chis, 2004; Yang et.al, 2005; Jorge, Nuno and Daniel, 2006) where time series datasets are very common. Time series classification and clustering in real estate are also becoming important since 1990s'.

Grissom, Wang and Webb (1991) examine intercity diversification using data on capitalization rates for the office sector in Texas over the period 1983–1986. They find spatial variation in property markets on an intercity level, offering the potential for more sophisticated diversification strategies than the traditional geographical groupings. Malizia and Simons (1991) examine existing diversification strategies and, subsequently, develop an alternative. Examining the traditional geographical classification, they find greater within-group heterogeneity for some classes, than when the US is considered as a single unit. The authors develop an alternative, new classification, using employment growth data (as a proxy for return) for US counties. Using a limited dataset comprising five observations between 1969 and 1987, the resulting classification was found to have superior diversification benefits compared with existing geographical and geographical/economic classifications. Goetzmann (1993) compares the risk and expected return of investment portfolios with and without single family homes. He concludes that spreading investment in residential assets across regions substantially reduces the risk.

Abraham, Goetzmann and Wachter (1994) explore interrelationship of housing market returns using the 1977-1992 returns to housing price indices data in 30 metropolitan areas. They emphasize the role that the interrelationship of housing market returns play on the purposes of equity investment, portfolio diversification and risk hedging. They apply the K-means clustering algorithm, and several grouping outcomes are identified. In addition, the bootstrapping testing is conducted to examine the robustness of the clustering algorithm, and the test outcome supports the results of the clustering analysis. The study verifies that the structural differences in housing markets exist between cities, so that housing market partition is not an effect of random association. The structural features of housing returns play an important role in diversifying debt and equity portfolio as well as hedging the housing market risks.

Goetzmann and Wachter (1995a) use cluster analysis to develop two classifications of 21 and 22 local office markets, using rent and vacancy data respectively. They find that cities form distinct groups, despite being geographically diverse, and that these groups do not necessarily conform to the existing geographical categories used for diversification purposes. Palmon, Simth and Sopranzetti (2004) examine the existence of price clustering in real estate listing prices and its consequences on transaction outcomes. The findings of their research indicate that the tendency to use even-ending (000-ending) prices is negatively related to the precision of the price estimates of the traded item and to the rounding cost. In contrast, the tendency to use just-below-even ending (900-ending) prices is negatively related to the rounding cost, but not to the precision of the price estimates of the traded item; furthermore, this tendency is also positively related to the number of properties listed by the listing real estate broker.

Although there are more studies on U.S. housing markets, the cluster analytical techniques are widely used in other countries. For instance, Goetzmann and Wachter (1995b) generate the K-means cluster analysis to investigate the real estate returns in the office markets across countries. They find that the global market can be disaggregated into three groups, European, Scandinavian, Iberian and Asian markets. The fluctuation of U.S. real estate market is part of the global market trend, so that there exists a strong cross-sectional relationship in the world office market.

Hoesli, Lizieri and Macgregor (1997) conducts a cluster analysis to the United Kingdom commercial property markets with a dataset containing the property returns for 156 property markets, and the dataset include three types of properties: retailing locations, office locations and industrial locations. For the result, the study does not identify a distinct regional clustering, and instead it claims that the property type plays a critical role in differentiating housing market behavior. The paper also uses the discriminant analysis and the test of the stability of the cluster structures to examine the study, and the results are supportive to the findings.

Lee (2001) presents an elegant and simple approach to the decomposition of property type and regional influences on property returns. By using data on retail, office and industrial properties spread across 326 real estate locations in the United Kingdom, over the period 1981 to 1995. He implies that the property type composition of the real estate fund should be the first level of analysis in constructing and managing the real estate portfolio.

A study, by Kim (2000) classifies Seoul housing market through cluster analysis by size of condominium, price and rent. Kang (1995) analyzes the factors making price difference by cluster analysis using housing price. Kim and Park (2003) show housing price change is different according to districts by Hedonic Price Model. In this manner, most South Korean studies use housing price for clustering housing market.

On the other hand, there is no any previous study analyzing residential rental returns in Turkey, an emerging market. At this point, our paper is the first academic study to establish clustering algorithms for Turkish real estate market.

3. Data and the Theoretical Framework

We firstly provide an overview of data and then give a brief explanation of the theoretical framework of time series clustering models and cluster procedure which are important for understanding the methodology of this study.

3.1 Overview of Data

The primary purpose of this study is to develop homogeneous groupings for 71 metropolitan residential markets in Turkey by using several hierarchical clustering algorithms (average, centroid, complete, median, single, ward and weighted). To cluster the sub-markets, rental returns over the period 2007:M6 to 2011:M6 from Reidin.com are utilized. (Note 1). Table 1 shows the names of 71 metropolitan residential markets and their annual rental returns.

3.2 Theoretical Framework

The goal of clustering is to find similarities and differences among unlabeled data objects to classify them into a small number of homogeneous groups where within-group-object similarity is maximized and where between-group-object dissimilarity is maximized. When the objects are time series data, such a classification might be useful to detect a few representative patterns, forecast future performances, quantify the affinity, conduct a survey, and etc (Liau, 2005; Xu and Wunsch II, 2005; Xu and Wunsch II, 2009; Han and Kamber, 2006).

If all data objects do not change with time, those data sets are called static. There are several cluster methods for static data in the literature. Han and Kamber (2001) classified clustering methods developed for handling various static data into five major categories: partitioning methods, hierarchical methods, density based methods, grid-based methods, and model-based methods. On the other hand, various algorithms have been developed to cluster different types of time series data. According to Liau (2005) there are three main approaches in time series clustering: raw-data-based, feature-based and model-based clustering. Clustering static data algorithms in such a way that time series data can be handled or to convert time series data into the form of static data so that the existing algorithms for clustering static data can be directly used. This approach usually works directly with raw time series data, thus called raw-data-based approach, and the major modification lies in replacing the distance/similarity measure for static data with an appropriate one for time series. The latter approach first converts a raw time series data either into a feature vector of lower dimension or a number of model parameters, and then applies a conventional clustering algorithm to the extracted feature vectors or model parameters, thus called feature-based and model-based approach, respectively. Figure 1 shows those three approaches (Liau, 2005).

One of the most widely used clustering methods is hierarchical clustering, due to the great visualization power it offers. One of the advantages of this method is its generality since the user does not need to provide any parameters such as the number of the cluster (Xu R. and Wunsch II, 2009; Xu R. and Wunsch II, 2005). On the other hand hierarchical clustering is not restricted to cluster time series with equal length (Liao, 2005).

A hierarchical clustering method works by grouping data objects into a tree of clusters according to the proximity matrix. A tree structure called a dendrogram is commonly used to represent the process of hierarchical clustering. It shows how objects are grouped together step by step. Clustering results can be obtained by cutting the dendrogram at different levels. This representation provides very informative descriptions and visualization for the potential data clustering structures (Xu R. and Wunsch II, 2009; Xu R. and Wunsch II, 2005).

There are generally two types of hierarchical clustering methods: agglomerative and divisive. Agglomerative method starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters, until all objects are in a single cluster or until certain termination conditions such as the desired number of clusters are satisfied. Divisive method does just the reverse of agglomerative hierarchical clustering by starting with all objects in one cluster. It subdivides the cluster into smaller and smaller pieces (Han and Kamber, 2001). Figure 2 shows a dendrogram of divisive hierarchical clustering method for 7 time series (Keogh and Kasetty, 2002).

4. Development of the Clustering Model and Evaluation of Study Results

The aim of this paper is to investigate the development of a classification and to reduce the overall risk of the investor's portfolio for 71 metropolitan residential markets in Turkey. The clustering of data employs the hierarchical algorithm, to which each of the objects stands out as its own cluster initially and they are combined into a hierarchy or treelike structure based on the similarity of objects. Such a classification could provide insights into the development of refined property portfolio diversification strategies. Euclidean distance approach is the representative distance measurement to quantify the inter-object similarity in the hierarchical algorithm, and it is defined as the straight-line distance between objects in n-dimensional space. It focuses on the magnitude of the distances, and group objects that

are close to each other. In this study, the 71x71 Euclidean distance matrix is used as an input of clustering, to maximize the distances between heterogeneous markets. After getting the distance matrix, we can classify the 71 returns into 10 clusters by using average, centroid, complete, median, single, ward and weighted hierarchical clustering methods. Then, Dunn Index and Silhouette Index are employed to decide the number of optimum clusters. The details of these algorithms and indices can be obtained from the Matlab Statistics Toolbox.

The validity indices of the Turkish residential market rental return data with seven hierarchical clusters are shown in Table 2. Dunn index suggests different cluster numbers and appear systematically to overestimate the number, whereas Silhouette index show two clusters across all the algorithms except centroid (3) and ward (3). The frequency of 2 and 3 highlights the true number of clusters in these data based on the indices. On the other hand, choosing an appropriate cluster number is a demanding problem (Xu R. and Wunsch II, 2009; Xu and Wunsch II, 2005). To overcome this demanding problem, visual approaches (visual cluster validity) can be used (Hathaway and Bezdek, 2003). The visual cluster validity (VCV) is a technique for visualizing high dimensional data as if they comprised an image. The basic idea is to map the data into an image framework, using the grey scale values or colors to indicate the magnitude of each variable for each observation.

The VCV method reorders rows and columns of the dissimilarity (or similarity) matrix using the cluster labels after some clustering methods have been applied. In other words, the original sequence of observations has been arranged such that the members of every cluster lie in consecutive rows and columns of the permuted dissimilarity matrix. Clearly defined light (dark, depending on the grey scale) squares along the diagonal indicate compact clusters that are well separated from neighboring points. If the data do not contain significant clusters, then this is readily seen in the image.

In this study the VCV method is then used to assess the cluster validity of this data. The input data is directly calculated from the data as Euclidean distance. The images related to the results of seven hierarchical clustering algorithms are shown in below figures (Figure 3, Figure 4, Figure 5, Figure 6, Figure 7, Figure 8, and Figure 9).

It should be noticed that, while the data are grouped by 2, 3, 4, 5, 6, 7, 8, 9 and 10; this will not necessarily be reflected in unsupervised clustering, e.g. there may be insufficient features to permit the separation. We can see vague area in the large dark block with inconspicuous boundaries which implies the cluster may include two or more overlapped clusters in it with very close relationship to each other in Figure 3, 4, 5, 6 and 7. It is also seen that there are three cluster blocks in Figure 8 and 9, one cluster size can be identified around 30, 30 and the other one is 10. On the other hand, the diagonal dark blocks are clearer in Figure 8. That is, "Ward" hierarchical cluster algorithms with the three clusters gives the best cluster performance on this data set. Figure 10 shows the dendrogram of "Ward" hierarchical cluster algorithm and each color indicates each cluster set for rental returns in Turkish residential market. The names of districts in each cluster sets with similar rental returns are also listed in Table 3.

The three-cluster partition of the districts reveals a clear rental return distinction of residential markets shown by Table 3. Cluster 1 is composed of twenty nine districts, which have the lowest rental return levels over the period of 2007:M6 to 2011:M6. Thirty four districts are grouped in Cluster 2. The cities in this group have relatively higher rental returns. The rest eight cities belong to Cluster 3. Rental return levels are distinctively higher than the other two groups, so that they are specified as "hot" housing markets. In the investment viewpoint for return maximization, sub-markets divided by rental returns have little correlation between themselves, so it can be the appropriate standard for investors to make housing investment portfolio to diversify risk.

Table 4 shows the descriptive statistics for the rental return data over the period 2007:M6 to 2011:M6. As can be seen, the best performing districts are all in cluster three (average mean 8.32%), while the first cluster performs the worst (average mean 6.70%). On the other hand, high rental returns are associated with higher levels of risk (standard deviation). The highest level of risk is in cluster three (average standard deviation 0.60%) and the lowest level of risk is in cluster one (average standard deviation 0.36%).

5. Conclusion

The objective of investors is to maximize expected returns, although they are subject to constraints, primarily risk. Return is the motivating force in the investment process. It is the reward for undertaking the investment. Rental returns from real estate investing are crucial to investors; they are what the game of investments is all about. The measurement of rental returns is necessary for investors to assess how well they have done or how well investment managers have done on their behalf. From that point of view, the aim of this paper is to develop homogeneous groupings for 71 metropolitan residential markets in Turkey by using several hierarchical clustering algorithms.

This study contributes to the literature in two aspects. First, it produces new information on Turkey's residential markets in the context of portfolio diversification. The historical clustering algorithms documented in this study

provides a useful guideline for real estate investors to select appropriate market areas and formulate efficiently diversified investment portfolios. The empirical findings of this study support the three-cluster partition of the districts that reveals a clear rental return distinction of residential markets in Turkey. Cluster 1 is composed of twenty nine districts, which have the lowest rental return levels (average mean 5.65%) over the period of 2007:M6 to 2011:M6. Thirty four districts are grouped in Cluster 2. The cities in this group have relatively higher rental returns (average mean 6.70%). The rest eight cities belong to Cluster 3. Rental return levels (average mean 8.32%) are distinctively higher than the other two groups, so that they are specified as “hot” housing markets for investment. On the other hand, high rental returns are associated with higher levels of risk (standard deviation), and vice versa. Secondly, the results of this paper could be useful not only for understanding the historical real estate market behavior, but also for investors to make rational investment decisions based on more accurate and realistic expectations of the future. There is no any previous study analyzing residential rental returns in Turkey. At this point, our paper is the first academic study to establish clustering algorithms for Turkish residential market.

References

- Abraham, J.M., Goetzmann W.N., & Wachter S.M. (1994). Homogeneous Groupings of Metropolitan Housing Markets. *Journal of Housing Economics*, 3 (3). 186–206. <http://dx.doi.org/10.1006/jhec.1994.1008>
- Aggarwal, C., Hinneburg A., & Keim D.A. (2001). On the Surprising Behavior of Distance Metrics in High Dimensional Space. In Proceedings of the 8th International Conference on Database Theory. London, UK.
- Bradley P.S., Fayyad U.M., & Reina C. (1998). Scaling Clustering Algorithm to Large Databases, In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining. New York, USA.
- Chis, M. (2004). Hierarchical Clustering using Evolutionary Algorithms. Paper Presented at the International Conference on Mathematical Methods for Learning and Advances in Data Mining and Knowledge Discovery. Como, Italy.
- Chis, M., & Dumitrescu D. (2002). Evolutionary Hierarchical Clustering for Data Mining, In Proceedings of the Symposium “Zilele Academice Clujene”. Cluj-Napoca, Romania.
- Elton, E.J., & Gruber M.J. (1970). Homogeneous Groups and the Testing of Economic Hypotheses. *Journal of Financial and Quantitative Analysis*, 4 (5), 581-602. <http://dx.doi.org/10.2307/2330115>
- Geltner, D.M., et.al. (2007). *Commercial Real Estate Analysis and Investments*, 2nd Edition. Thomson South-Western, Cincinnati, OH, USA.
- Grissom, T.V., Wang K., & Webb J.R. (1991). The Spatial Equilibrium of Intra-Regional Rates of Return and The Implications for Real Estate Portfolio Diversification. *Journal of Real Estate Research*, 7 (1). 59–71. [Online] Available from: <http://ideas.repec.org/s/jre/issued.html> (September 1, 2011).
- Goetzmann, W.N. (1993). The Single Family Home in the Investment Portfolio. *Journal of Real Estate Finance and Economics*, 6 (3). 201-222. <http://dx.doi.org/10.1007/BF01096958>
- Goetzmann, W.N., & Wachter S.M. (1995a). Clustering Methods for Real Estate Portfolios. *Real Estate Economics*, 23 (3), 271–310. <http://dx.doi.org/10.1111/1540-6229.00666>
- Goetzmann, W.N., & Wachter, S.M. (1995b). *The Global Real Estate Crash: Evidence from an International Database in A Global Perspective on Real Estate Cycles* by Stephen J. Brown and Crocker H. Liu, Kluwer Academic Publishers. Boston, USA.
- Hathaway, R.J., & Bezdek J.C. (2003). Visual Cluster Validity Displays for Prototype Generator Clustering Methods. *Pattern Recognition Letters*, 24 (9), 1563-1569. [http://dx.doi.org/10.1016/S0167-8655\(02\)00395-1](http://dx.doi.org/10.1016/S0167-8655(02)00395-1)
- Han, J., & Kamber M. (2006). *Data Mining: Concepts and Techniques*, Morgan Kaufmann. San Francisco, USA.
- Hoesli, M., Lizieri C., & Macgregor B.D. (1997). The Spatial Dimensions of the Investment Performance of UK Commercial Property. *Urban Studies*, 34 (9), 1475–1494. <http://dx.doi.org/10.1080/0042098975529>
- Jorge, C., Nuno C., & Daniel P. (2006). A Periodogram-Based Metric for Time Series Classification. *Computational Statistics and Data Analysis*, 50. 2668-2684. <http://dx.doi.org/10.1016/j.csda.2005.04.012>
- Kang, J.M. (1995). A Study on Market Segmentation of Apartment in Seoul. Master D. diss. University of KonKuk.
- Keogh, E., & Kasetty S. (2002). On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Alberta, Canada.

- Kim, B.T. (2000). Determination of Spatial Boundaries of Seoul Housing Sib-market. Master D. diss. University of YonSei.
- Kim, G.S., & Park J.Y. (2003). The Spatial Pattern of Housing Prices: Seoul and New Town. *Journal of the KRSA*, 19, 47-61.
- Lee, S.L. (2001). The Relative Importance of Property Type and Regional Factors in Real Estate Returns. *Journal of Real Estate Portfolio Management*, 7 (2), 159-167. [Online] Available from: <http://business.fullerton.edu/finance/jrepm/past/Default.htm> (September 1, 2011).
- Liao, T.W. (2005). Clustering of Time Series Data—A Survey, *Pattern Recognition*, 38, 1857-1874. <http://dx.doi.org/10.1016/j.patcog.2005.01.025>
- Malizia, E.E., & Simons R.A. (1991). Comparing Regional Classifications for Real Estate Portfolio Diversification, *Journal of Real Estate Research*, 6 (1), 53–78. [Online] Available from: <http://ideas.repec.org/s/jre/issued.html> (September 1, 2011).
- Palmon, O., Barton A.S., & Sopranzetti B.J. (2004). Clustering in Real Estate Prices: Determinants and Consequences. *Journal of Real Estate Research*, 26 (2), 115-136. [Online] Available from: <http://ideas.repec.org/s/jre/issued.html> (September 1, 2011).
- Reidin.com. (2011). Real Estate Investment and Development Information Network. [Online] Available from: <http://www.reidin.com> (September 1, 2011).
- Xu, R. & Wunsch II D.C. (2005). Survey of Clustering Algorithms, *IEEE Transactions on Neural Networks*, 16 (3), 645-678. <http://dx.doi.org/10.1109/TNN.2005.845141>
- Xu, R., & Wunsch II D.C. (2009). *Hierarchical Clustering*, in Clustering, John Wiley and Sons, Inc., Hoboken, New Jersey, USA.
- Yang, K., et.al. (2005). Clever: A Feature Subset Selection Technique for Multivariate Time Series, Technical Report, University of Southern California, USA.

Notes

Note 1. REIDIN.com is the world's first business information service, that is designed to meet the unique requirements of real estate market professionals and academics who are interested in the emerging countries, develops residential property price indices for number of countries including Turkey. Reidin.com is calculating rental returns, that are the total yearly rent divided by the house price and expressed in percentage, both for 7 major cities and their 71 district.

Table 1. Clusters in Cities and Annual Rental Returns*

City Name	District Name	Rental Return	City Name	District Name	Rental Return
Adana	Seyhan	5.92%	Istanbul	Fatih	7.04%
Adana	Yuregir	7.44%	Istanbul	Gaziosmanpasa	6.51%
Ankara	Altindag	8.06%	Istanbul	Gungoren	7.18%
Ankara	Cankaya	6.23%	Istanbul	Kadikoy	4.98%
Ankara	Etimesgut	6.34%	Istanbul	Kagithane	7.01%
Ankara	Golbasi	4.51%	Istanbul	Kartal	5.91%
Ankara	Kecioren	5.88%	Istanbul	Kucukcekmece	5.97%
Ankara	Mamak	6.18%	Istanbul	Maltepe	6.19%
Ankara	Pursaklar	5.01%	Istanbul	Pendik	6.05%
Ankara	Sincan	5.80%	Istanbul	Sancaktepe	5.98%
Ankara	Yenimahalle	4.92%	Istanbul	Sariyer	5.24%
Antalya	Kepez	6.52%	Istanbul	Silivri	6.35%
Antalya	Konyaalti	5.31%	Istanbul	Sultanbeyli	6.32%
Antalya	Muratpasa	6.32%	Istanbul	Sultangazi	6.15%
Bursa	Nilufer	5.37%	Istanbul	Sile	6.20%
Bursa	Osmangazi	6.35%	Istanbul	Sisli	6.73%
Bursa	Yildirim	6.28%	Istanbul	Tuzla	5.15%
Istanbul	Adalar	5.67%	Istanbul	Umraniye	5.59%
Istanbul	Arnavutkoy	6.09%	Istanbul	Uskudar	5.96%
Istanbul	Atasehir	8.67%	Istanbul	Zeytinburnu	6.91%
Istanbul	Avcilar	6.58%	Izmir	Balcova	5.49%
Istanbul	Bagcilar	6.53%	Izmir	Bayrakli	4.79%
Istanbul	Bahcelievler	6.88%	Izmir	Bornova	5.25%
Istanbul	Bakirkoy	6.30%	Izmir	Buca	5.44%
Istanbul	Basaksehir	5.05%	Izmir	Cigli	5.71%
Istanbul	Bayrampasa	6.52%	Izmir	Gaziemir	5.16%
Istanbul	Besiktas	6.27%	Izmir	Guzelbahce	5.30%
Istanbul	Beykoz	5.87%	Izmir	Karabaglar	6.29%
Istanbul	Beylikduzu	6.73%	Izmir	Karsiyaka	5.45%
Istanbul	Beyoglu	10.76%	Izmir	Konak	6.36%
Istanbul	Buyukcekmece	5.93%	Izmir	Narlidere	4.21%
Istanbul	Catalca	5.16%	Kocaeli	Gebze	5.64%
Istanbul	Cekmekoy	7.15%	Kocaeli	Golcuk	6.30%
Istanbul	Esenler	7.01%	Kocaeli	Izmit	6.65%
Istanbul	Esenyurt	9.42%	Kocaeli	Korfez	6.29%
Istanbul	Eyup	10.47%			

*Rental returns are as of June 2011.

Table 2. Number of Optimum Clusters of Seven Hierarchical Clustering Algorithms Based on Dunn and Silhouette Indices

Hierarchical Algorithms	Dunn Index	Silhouette Index
Single	3	2
Complete	10	2
Average	9	2
Centroid	2	3
Median	5	2
Ward	10	3
Weighted	9	2

Table 3. "Ward" Hierarchical Cluster Algorithm with Three Groups Specified

Clusters	Name of Districts
Cluster 1	Golbasi, Pursaklar, Yenimahalle, Kepez, Konyaalti, Muratpasa, Nilufer, Adalar, Besiktas, Beykoz, Buyukcekmece, Catalca, Kadikoy, Kartal, Sariyer, Sile, Sultanbeyli, Tuzla, Umraniye, Balcova, Bayrakli, Bornova, Buca, Cigli, Gaziemir, Guzelbahce, Karsiyaka, Narlidere, Gebze
Cluster 2	Seyhan, Cankaya, Etimesgut, Kecioren, Mamak, Sincan, Osmangazi, Yildirim, Arnavutkoy, Atasehir, Avcilar, Bagcilar, Bakirkoy, Basaksehir, Bayrampasa, Beylikduzu, Esenler, Eyup, Fatih, Gaziosmanpasa, Kagithane, Kucukcekmece, Maltepe, Pendik, Sancaktepe, Silivri, Sisli, Sultangazi, Uskudar, Karabaglar, Konak, Golcuk, Izmit, Korfez
Cluster 3	Yuregir, Altindag, Bahcelievler, Beyoglu, Cekmekoy, Esenyurt, Gungoren, Zeytinburnu

Table 4. Descriptive Statistics for Rental Returns from 2007:M6 to 2011:M6

Cluster 1			Cluster 2			Cluster 3		
District	Mean	St.Dev.	District	Mean	St.Dev.	District	Mean	St.Dev.
Golbasi	5.51%	0.97%	Seyhan	6.19%	0.46%	Yuregir	7.93%	0.56%
Pursaklar	5.78%	0.56%	Cankaya	6.28%	0.21%	Altindag	8.46%	0.87%
Yenimahalle	5.24%	0.18%	Etimesgut	6.29%	0.41%	Bahcelievler	7.59%	0.44%
Kepez	5.76%	0.35%	Kecioren	6.09%	0.18%	Beyoglu	9.62%	0.75%
Konyaalti	5.32%	0.13%	Mamak	6.55%	0.52%	Cekmekoy	8.30%	0.74%
Muratpasa	5.96%	0.16%	Sincan	6.57%	0.69%	Esenyurt	9.01%	0.30%
Nilufer	5.47%	0.23%	Osmangazi	6.37%	0.29%	Gungoren	7.92%	0.56%
Adalar	5.97%	0.23%	Yildirim	6.13%	0.28%	Zeytinburnu	7.72%	0.59%
Besiktas	6.18%	0.19%	Arnavutkoy	7.06%	0.69%	Average	8.32%	0.60%
Beykoz	5.86%	0.30%	Atasehir	7.31%	1.00%			
Buyukcekmece	6.18%	0.46%	Avcilar	7.16%	0.34%			
Catalca	5.74%	0.65%	Bagcilar	7.03%	0.37%			
Kadikoy	5.89%	0.62%	Bakirkoy	6.33%	0.39%			
Kartal	5.95%	0.26%	Basaksehir	6.49%	0.93%			
Sariyer	5.67%	0.25%	Bayrampasa	7.18%	1.08%			
Sile	5.62%	0.41%	Beylikduzu	6.58%	0.28%			
Sultanbeyli	6.29%	0.49%	Esenler	7.26%	0.82%			
Tuzla	5.62%	0.45%	Eyup	7.10%	1.40%			
Umraniye	5.92%	0.26%	Fatih	7.46%	0.49%			
Balcova	5.11%	0.47%	Gaziosmanpasa	6.57%	0.40%			
Bayraklı	5.22%	0.38%	Kagithane	7.24%	0.33%			
Bornova	5.61%	0.22%	Kucukcekmece	6.64%	0.50%			
Buca	5.64%	0.22%	Maltepe	6.40%	0.39%			
Cigli	5.70%	0.34%	Pendik	6.42%	0.88%			
Gaziemir	5.30%	0.24%	Sancaktepe	6.50%	0.53%			
Guzelbahce	5.40%	0.45%	Silivri	6.77%	0.59%			
Karsiyaka	5.80%	0.31%	Sisli	6.50%	0.37%			
Narlidere	4.23%	0.21%	Sultangazi	6.63%	0.34%			
Gebze	5.89%	0.44%	Uskudar	6.39%	0.28%			
Average	5.65%	0.36%	Karabaglar	6.87%	0.41%			
			Konak	6.56%	0.28%			
			Golcuk	7.07%	0.73%			
			Izmit	6.95%	0.49%			
			Korfez	6.76%	0.47%			
			Average	6.70%	0.52%			

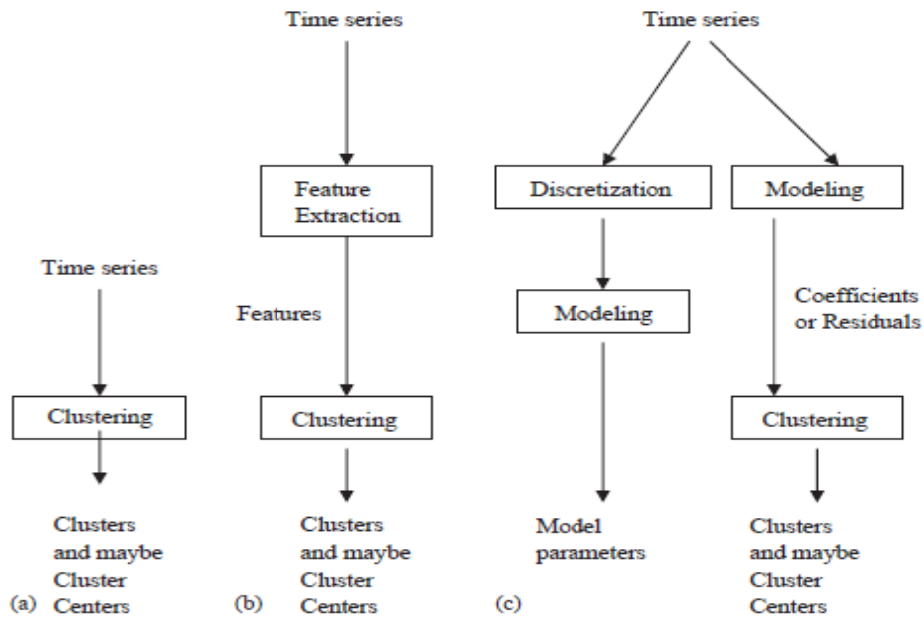


Figure 1. Three Time Series Clustering Approaches: (a) Raw-Data-Based, (b) Feature-Based, (c) Model-Based

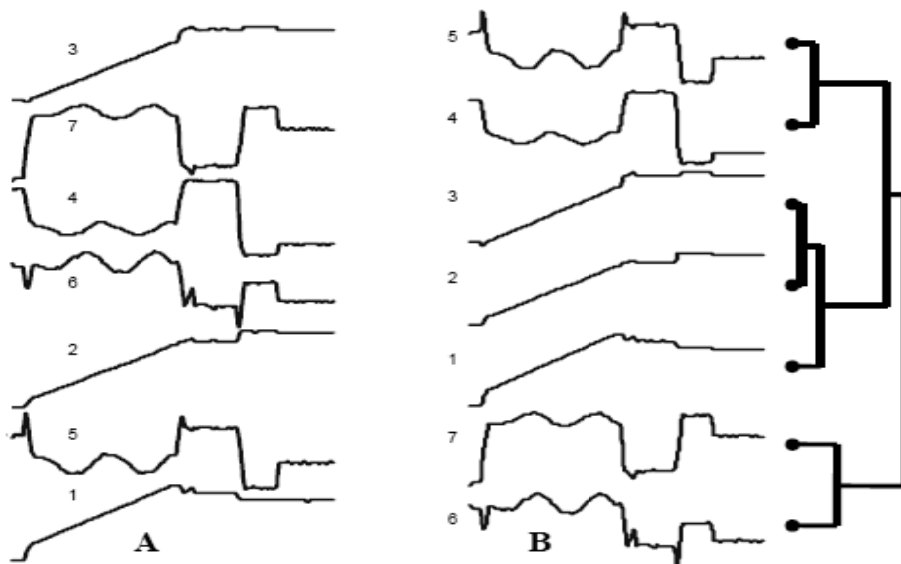
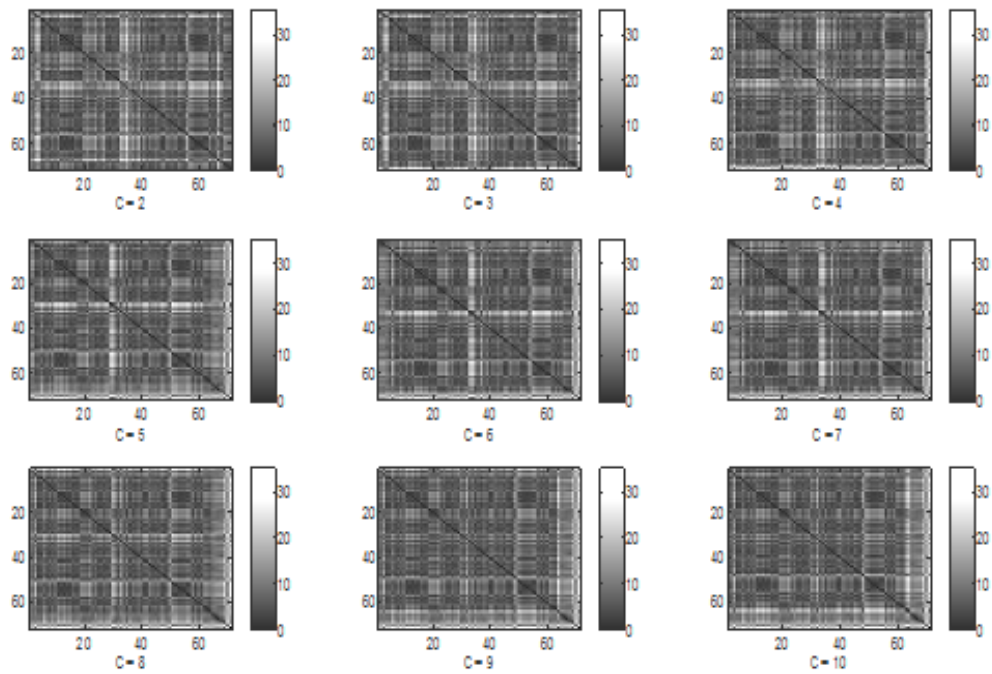
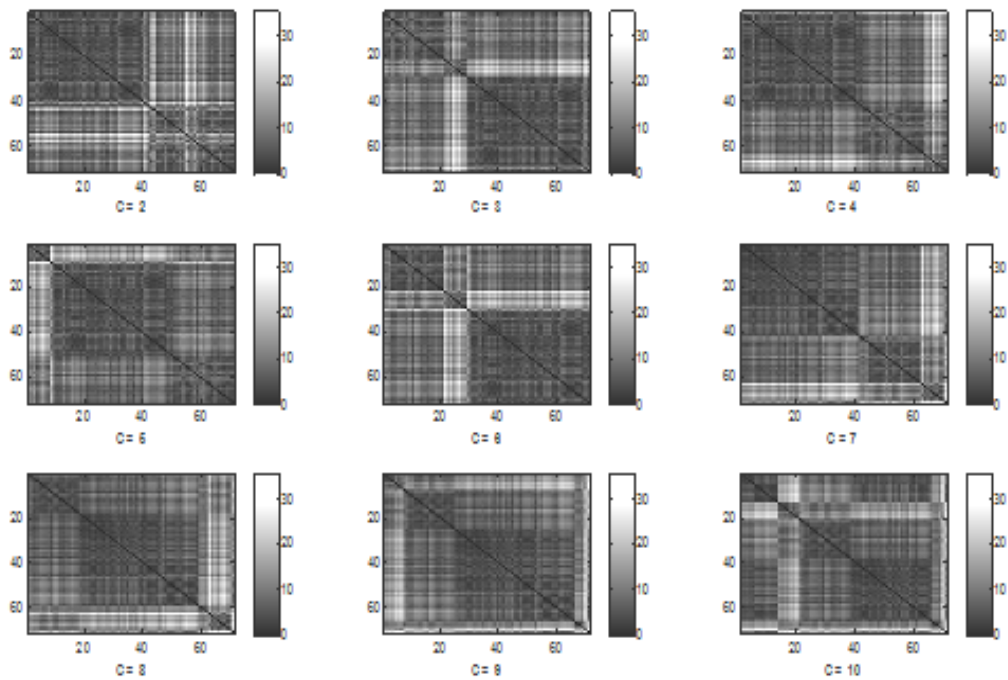


Figure 2. A Hierarchical Clustering of 7 Time Series



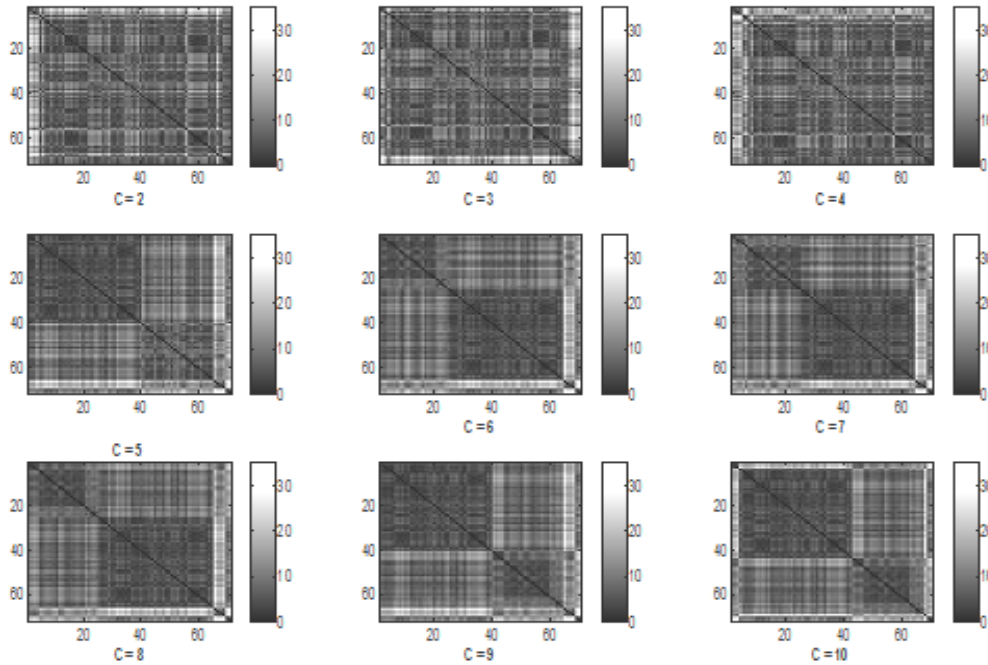
* C indicates number of clusters.

Figure 3. Results of “Single” Hierarchical Clustering Algorithm



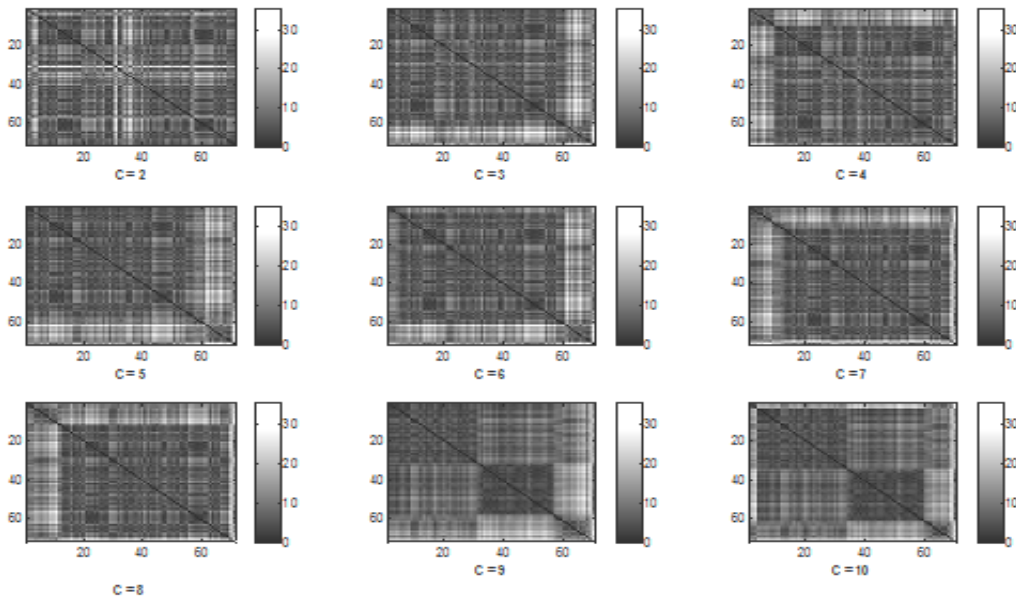
* C indicates number of clusters.

Figure 4. Results of “Complete” Hierarchical Clustering Algorithm



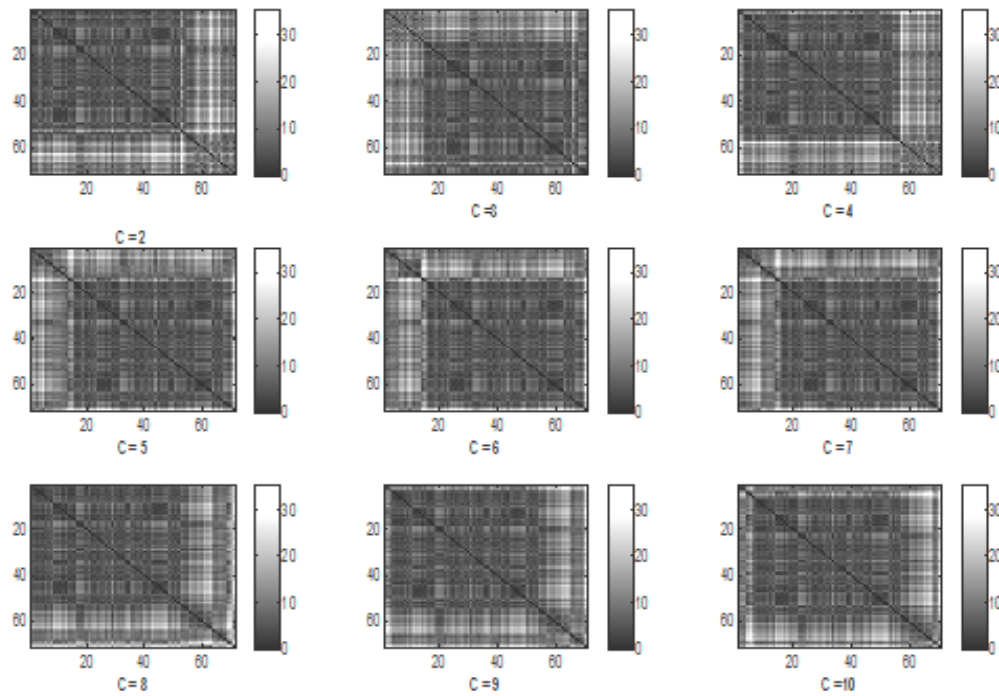
* C indicates number of clusters.

Figure 5. Results of “Average” Hierarchical Clustering Algorithm



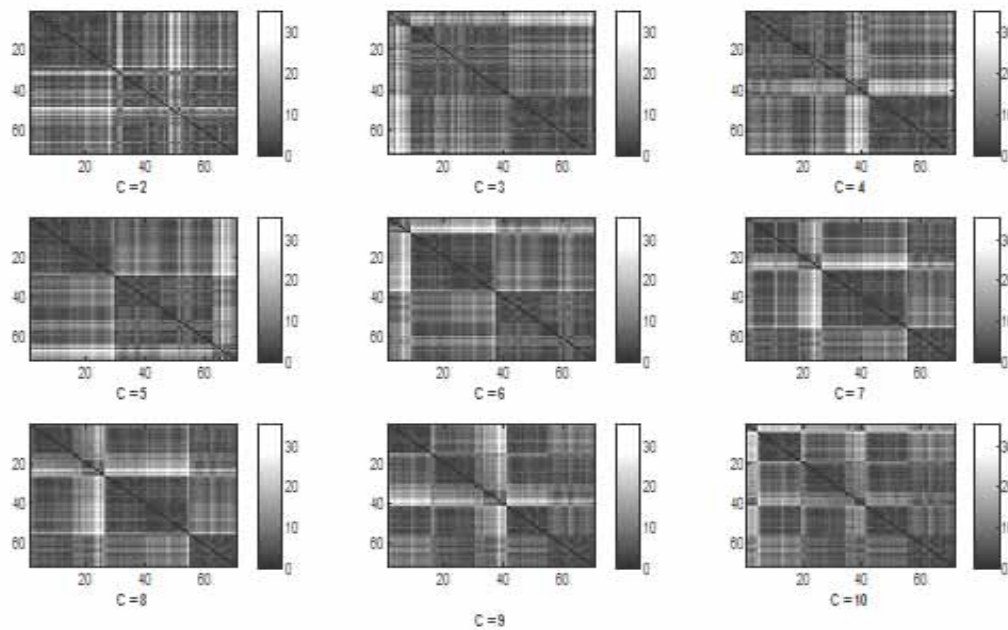
* C indicates number of clusters.

Figure 6. Results of “Centroid” Hierarchical Clustering Algorithm



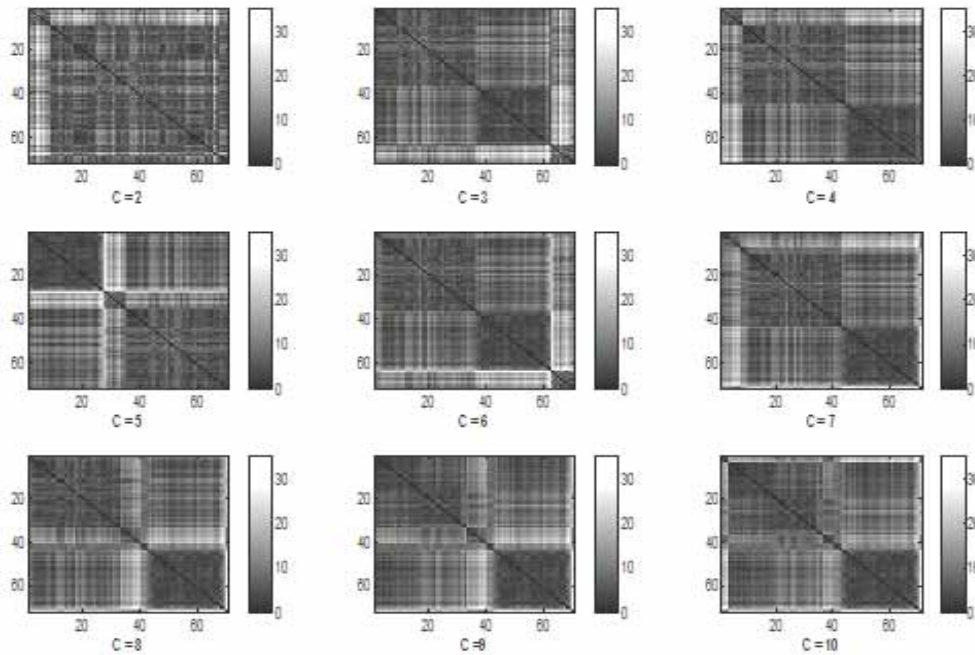
* C indicates number of clusters.

Figure 7. Results of "Median" Hierarchical Clustering Algorithm



* C indicates number of clusters.

Figure 8. Results of "Ward" Hierarchical Clustering Algorithm



* C indicates number of clusters.

Figure 9. Results of “Weighted” Hierarchical Clustering Algorithm

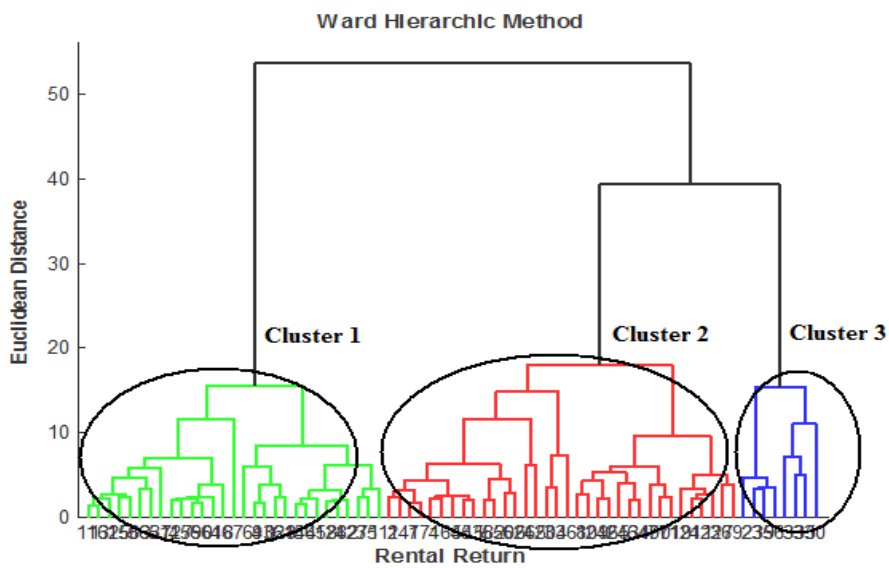


Figure 10. Dendrogram of “Ward” Hierarchical Cluster Algorithm