Regional Employment in Greece by means of Cluster Analysis

Michaelides, Panayotis G. and Economakis, George and Lagos, Dimitrios

2005

Regional Employment in Greece by means of Cluster Analysis

PANAYOTIS MICHAELIDES
School of Applied Mathematics and Physics
National Technical University of Athens
9 Heroon Polytechneiou
Zografou Campus, 157.80
GREECE
pmichael@central.ntua.gr

GEORGE ECONOMAKIS
School of Business Administration
University of the Aegean, Chios
Mechaleion Building, Mechalon 8, 82100
GREECE
geconomakis@ba.aegean.gr

DIMITRIOS LAGOS
School of Business Administration
University of the Aegean, Chios
Mechaleion Building, Mechalon 8, 82100
GREECE
d.lagos@aegean.gr

Abstract: This article attempts to group the regions of Greece into clusters with similar characteristics, by means of Cluster Analysis. The results demonstrate that the thirteen regions of Greece can be grouped in four major clusters, which are, in general terms, consistent with the findings of another research.

Keywords: Strategic Regional Planning, Sustainable Development, Employment, Cluster Analysis, Greece

1. Introduction

The purpose of the present paper is to group the thirteen regional departments of Greece, into clusters of regions sharing similar characteristics, in the period 1991-2001. The paper uses the clustering analysis methodology which offers a reliable quantitative framework.

More precisely, with the aid of the clustering analysis methodology and the use of statistical data concerning regional employment, useful information is gathered about the three sectors of economic activity. Continuously, the regions of Greece are grouped according to their basic employment characteristics.

The outline of the paper is as follows: section 2 discusses briefly the regional problem in Greece; section 3 provides the methodological framework of Clustering Analysis; section 4 presents the data used and the classification results; finally, section 5 compares the findings with the results of a previous study using a different methodology and concludes the paper.

2. Regional Development in Greece

The regional structure of economic activity in Greece consists of the thirteen (13) regional departments, belonging to the Greek territory, presented and number - for reasons of convenience - in Table 1, below.

Table 1: The Regional Departments of Greece

Number	Regional Department
1	Eastern Macedonia and Thrace
2	Central Macedonia
3	Western Macedonia
4	Epirus
5	Thessaly
6	Ionian Islands
7	Western Greece,
8	Sterea Ellada
9	Attica
10	Peloponnesus
11	Northern Aegean
12	Southern Aegean
13	Crete

A significant characteristic of the thirteen regions of Greece is the over-accumulation of population in Attica (34.31%), and in Central Macedonia (9.64%) adding up to, approximately, a very high 44% of the total population of the country, in 2001. Meanwhile, these two regions produce about 54% of the Gross National Product (G.N.P.) and accumulate 55% of aggregate employment, i.e. 38% and 17%, respectively [1].

More precisely, over the 1991-2001 period, the population of Greece increased by 6.8%, while the Athens Metropolitan Area grew also at 6.8%, thus maintaining its share of about one third of the total Greek population. However, population increase in the remainder of Attica was very high (27.4%) and is mainly due to suburbanization driven by new infrastructure projects in outer areas confirming its dominant position as the main urban center of Greece.

These are associated with the location of the new Athens International Airport, which provides significant employment opportunities and a further boost to the development of the area at the expense of the other regional departments of Greece.

Consequently, Attica as well as Central Macedonia, constitute two very significant poles inside the Greek territory.

Therefore, apart from the fact that each and every one of the above mentioned regions contributes to the development of the country as a whole, it seems that the country faces a continuously expanding regional problem. In the words of Kazakos [2]: "The main socioeconomic problem […] is that […] the geographical accumulation of economic activities contributes more to their development than to their diffusion".

Subsequently, a major characteristic of the problem is that the already existing disparities are keep growing. The irrational allocation of the population and the economic activities in the countryside generate an increase in the economy's social cost [3]. In addition, the total economy's rate of growth did not foster, to a large extent, the economic and social cohesion with the other countries of the European Community; neither did it contribute decisively to the reduction of disparities among the regions [4].

In general, a regional problem arises when economic and demographic accumulation is present in certain urban centers, such us Athens or Thessalonica, while the countryside is weakening in terms of economic development and population growth. The problem is rooted in the country's historical context (e.g. Ottoman Empire, postwar immigration, lack of infrastructure in the countryside, urbanization process observed in the 1970s, immigration in Greece, geographic formation of the country, ineffectiveness of the implemented regional policy, etc.) ([2], [3], [5], [6]). Needless to say, the dominance of the urban center over the rest of the country becomes, day by day, increasingly powerful.

It is evident that a new and dynamic regional planning policy is needed. The methodology of Clustering Analysis will help identify the groups of regional departments that share the same characteristics and, thus, a common regional policy planning aimed at groups of regional departments could be implemented leading to sustainable development.

In the next section, the methodological framework is presented.

3. Methodology

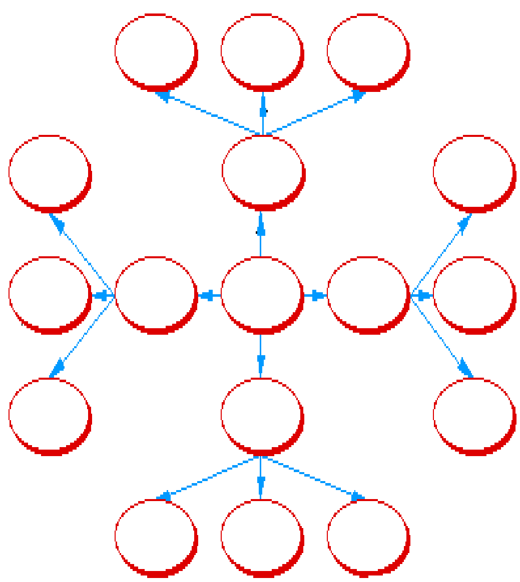
3.1 Cluster Analysis

A general question in applied economics is how to organize observed data into meaningful structures. Clustering has been used since long for grouping together entities and data with similar characteristics, but nowadays it has acquired new dimensions as a solution to the complexity related to voluminous datasets and information repositories. The reason for its increased significance and convenience is that it relies on creating natural groups in the existing data rather than classifying them on the basis of some externally imposed criteria [7].

Clustering techniques are applied when we need to divide the data into clusters. These clusters presumably reflect some mechanism at work in the domain from which data are drawn; the mechanism causes some units of the cluster to bear a stronger resemblance to one another than they do to the remaining units [7].

The term *cluster analysis*, introduced in Tryon [8], encompasses a number of different algorithms and methods for grouping objects and data of similar kind into respective categories. In other words, cluster analysis is an exploratory data analysis tool which aims at sorting different objects and data into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. Given the above, cluster analysis can be used to discover structures in data without providing an explanation. Thus, cluster analysis simply discovers structures in data without explaining why they exist.

Fig. 1



The above diagram divides a data set so that records with similar content are in the same group, and groups are as different as possible from each other [9]. Also, clustering may also be defined as the technique of grouping data together based on their locality and connectivity [10].

However, clustering, as a methodology is vulnerable on two fronts, i.e. the classifications delivered are not sufficiently compelling to convince the experts always, due to the lack of prior theoretical assumptions; and the second is that the techniques themselves are not based on probability models [7].

3.2 K-Means Clustering

Very good reviews of Clustering Algorithms have been provided by various researchers, see [11, 12]. There exist several methods (e.g. Nearest Neighbor, Furthest Neighbor, Centroid, Median, Group Average, Ward's, and *K*-Means) for grouping observations from a multivariate dataset into clusters of similar points. All of the methods except *K*-Means are hierarchical clustering methods. *K*-Means provides one non-hierarchical method.

In hierarchical methods, each observation begins in its own cluster. Two clusters are then merged to form a new cluster that replaces the two old clusters. This process is repeated until only one cluster is left. Then observations are combined into successively larger clusters until it reaches the number of clusters that is specified.

However, the algorithms used differ in how they compute the distance between the two clusters. Thus, one should choose the Distance Metric, which is the method that will calculate the distance between clusters. There exist various Distance Metrics such as: Euclidean distance, Squared Euclidean distance, City-block (Manhattan) distance, Chebychev distance, Power distance, Percent disagreement, etc [14].

As seen, in cluster analysis the objective is to divide a set of observations (here the collection of employment data for the thirteen regional departments of Greece) into groups or clusters in such a way that most pairs of observations which are placed in the same cluster are more similar to each other than are pairs of observations which are placed in two different clusters.

We use the Euclidean distance as a measure of similarity, which is probably the most commonly chosen type of distance [11]. The Euclidean distance is the geometric distance. It is computed as: distance (x,y) = $\{ \sum_i (x_i - y_i)^2 \}^{½}$

In the *K*-Means method [13], formation of clusters begins with an initial partition then uses a search algorithm to test other partitions to identify the one with the least error. Note that Euclidean (and squared Euclidean) distances are usually computed from raw data. This method has certain advantages (e.g. the distance between any two objects is not affected by the addition of new objects to the analysis, which may be outliers) [14].

We use K-means, a nonhierarchical clustering method, which fixes the number of clusters, K, and divides the observations into K clusters in such a way that an objective function, i.e. the total sum of squared Euclidean distances between observations and their respective cluster centroids (average value of the observations) is minimized [11, 15].

The K-means algorithm aims to minimize the squared error function. The objective function is $J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$ where $\left\| x_i^{(j)} - c_j \right\|$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centers.

The minimization for our dataset was performed with *Stat-Graphics Plus* software.

There is no consensus in the statistical community on a method to select an appropriate value for the number of clusters [15], K and so based on evidence from a previous study we have chosen the number of clusters.

The K-means clustering method was chosen because it is effective in using a heterogeneous high-dimensional multivariate data set to create a manageable set of relatively homogeneous classes which could be employed in long term regional planning [15].

In the next section, the data and the classification results are presented.

4. Empirical Results and Discussion

4.1 Data and Variables

For this paper we use data that come from the National Statistic Service of Greece [1]. The variables constructed, for every regional territory of Greece, are:

L1-91: labor in the agricultural sector in 1991
L2-91: labor in the industrial sector in 1991
L3-91: labor in the service sector in 1991
L1-01: labor in the agricultural sector in 2001
L2-01: labor in the industrial sector in 2001
L3-01: labor in the service sector in 2001
TL91: total labor in 1991
TL01: total labor in 2001
RL1: % change of labor in the agricultural sector between 1991-2001
RL2: % change of labor in the manufacturing sector between 1991-2001
RL3: % change of labor in the service sector between 1991-2001
P1-91: % of total labor employed in the agricultural sector in 1991
P1-01: % of total labor employed in the agricultural sector in 2001
P2-91:% of total labor employed in the manufacturing sector in 1991
P2-01:% of total labor employed in the manufacturing sector in 2001
P3-91:% of total labor employed in the service sector in 1991
P3-01: % of total labor employed in the service sector in 2001

4.2 The Classification

Using K-means we partition the above mentioned variables into distinct clusters. As seen, there is no consensus in the statistical community on a method to select an appropriate value for the number of clusters, K and so given some empirical evidence from a previous study [16] using a different methodological approach we decided to use four (4) clusters.

The resulting clusters are presented in Table 2, where the regional departments' numbering is based on Table 1. Table 2 also presents corresponding clusters of regional departments by Lagos et al. [16] using the Boudeville classification.

Note that in [16] the regional departments were categorized according to the traditional eight (8) Boudeville types. However, it was obvious that the regional departments tended to form four (4) distinct groups.

Table 2: Regional Classification

Cluster	**Regional Departments (K-means)**	**Regional Departments (Boudeville)**
1	12, 9	12, 9
2	11, 13, 2, 6	11, 13
3	1, 5, 3, 8	1, 5, 2
4	4, 7, 10	4, 7, 10, 3, 6, 8

In the next section, the results of the paper are compared with those obtained by a previous study and our concluding remarks are stated.

5. Conclusion

In the present paper we attempted to group, the thirteen regional departments of Greece into several clusters of regions which share similar employment characteristics, using the Clustering Analysis methodology. The results do show that the thirteen regions can, indeed, be grouped in four (4) clusters and our findings are, generally, consistent with the findings of a previous study using an alternative methodology.

It is evident that the thirteen regional departments of Greece formed four (4) clusters and the estimated clusters are, in general terms, consistent with the findings in [16] where it was obvious that the regional departments also tended to form four (4) very similar groups.

Any differences between the two approaches' results (e.g. in the members of each cluster) should not be surprising and are due to the fact that the Clustering Analysis in general, and the *K*-means method in particular, does not make any particular theoretical assumptions about the characteristics of each regional department.

Conclusively, we believe that future and more extended research on the subject would obviously be of great interest, including the use of alternative clustering algorithms, as well as the incorporation of additional regional variables (e.g. output, population, etc) in the model.

References

[1] National Statistical Service of Greece (1991, 2001), *Results of Census* (http://www.statistics.gr).

[2] Kazakos, P. (1984), *Lessons of regional development*, Athens: Papazisis Publications, p. 66.

[3] Konsolas, I. N. (1985), *Regional economic policy: I. General speculation*, Second Edition, Athens: Papazisis Publications, pp. 347-50.

[4] Konsolas, I. N. (1997), *Contemporary Regional Economic policy*, Athens: Papazisis Publications, p. 239.

[5] Papadaskalopoulos, A. D. (1990) *Principal Methods of Regional Analysis*, Athens: Papazisis Publications.

[6] Milios, J. (2000), *Greek Social Formation: from expansionism to capitalist development*, Athens: Kritiki Publications.

[7] Indian Statistical Institute, *Introduction to Clustering*, Documentation Research and Training Centre (http://drtc.isibang.ac.in/~aparajita/project/project%2020appu/Project/Chapter2.doc).

[8] Tryon, R.C., *Cluster Analysis*, Ann Arbor, MI, 1939, Edward Brothers.

[9] Aerts, M., Molenberghs, G. and L. Ryan, *Topics in Modelling of Clustered Data*, 2002, CRC Press.

[10] Wu, W., Xiong, H. and S. Shekhar (ed), *Clustering and Information Retrieval*, 2003, Kluwer Academic Publishers.

[11] K. Jain and R. C. Dubes, *Algorithms for Data clustering*, Prentice Hall, 1988, ch. 3.3.

[12] Kaufmann, L. and P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, 1990, John Wiley & Sons.

[13] MacQueen, J. B. Some Methods for Classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, Berkeley, University of California Press, 1:2, 81-297

[14] Statsoft, *Clustering Analysis*, Electronic Textbook, 2004, StatSoft Inc

[15] Glascoe, L., Glaser, R., Hung-Neng S. Chin, and G. Loosmore, Regional-Scale Wind Field Classification Employing Cluster Analysis, *13th Conference on the Applications of Air Pollution Meteorolgy with the Air and Waste Management Association, Session 7: Variability and Uncertainty*, 24 August 2004 (ams.confex.com/ams/pdfpapers/78711.pdf)

[16] Lagos, D., Kalligianakis, I., Economakis, G. and P. Michaelides, Regional Structure of Employment in Greece (1991-2001): An Empirical Investigation, *Proceedings of the 2nd International Conference on Enlarged Europe and Regional Disparities*, Technological Educational Institute of Epirus, School of Management and Economics, 2004, Preveza, 3-4 June.