

# **Cluvar: a grouping technique of statistic units and variables selection.**

Giorgio Skonieczny- Benedetto Torrisi

**University of Catania, Faculty of Economics,  
c /so Italia n.57, Catania  
e-mail: giorgios@unict.it-b.torrisi@unict.it**

## Abstract

First of all researches regarding the division – relating to a complex phenomena- of a territory in homogeneous areas present problems relating to the identification of the variables set that must individuate the hypothesis in a univocal way and define the relation among the areas.

On the other hand, the choice of a technique of classification is strongly conditioned by the set of definite indicators and by the hypothesised grouping methods. The not hierarchical classification techniques need the previous determination of the groups' number or of the elements number per every group. The goal of this research is to define a technique of classification that allows the formation of homogeneous groups without the previous definition of their number. At the same time this classification makes possible significant links among the units. This technique gives the description of the relations among the single units and the groups.

In this research this methodology is used to determine the regional diversification of the Italian financial system.

With the cluster analysis applied to the territorial units we obtain the structural elements which allow the individuation of indicators' groups of real and financial economy that synthesize both the information obtained from the index and a system of relation among the considered territorial units.

## **1. Introduction**

In the ambit of the research on the economic and social phenomena the choice of the most adaptable variables to represent the structure of the phenomena is often really difficult and can strongly conditioned the results. Particularly, for Cluster Analysis, the problem becomes even more complicate if the unit to be classified comes from heterogeneous socio-economic realities and then it is individuated by a set of variables totally, or more often just in part, different.

The goal of this research is to define a technique of classification to consent the formation of homogeneous groups without the previous definition of their number, at the same time this classification makes possible significant links among the units. This technique gives the description of the relations among the single units and the groups.

In this research this methodology is used to determine the regional diversification of the Italian financial system.

## 2. The applied methodology: cluvar

Cluvar is the acronym of cluster variables. The goal of this technique is to individuate groups and relations among units characterized by an high level of homogeneity, this level of homogeneity must exist among units and among the variables that define the units and finally the typologies of relations among single units, and between units and Cluster identifying among variables which have contributed more to couple units links.

Particularly, if it is given a data table where  $x_{ij}$  represents the determination of the variable  $X_j$  per  $j=1, \dots, m$  on units  $U_i$  per  $i=1, \dots, n$ .  $S_{n,m}$  would be the correspondent table of the standardized values

and  $d_i$  the most opportune distance;  $m$  matrices must be determined of the A distances  $A \begin{pmatrix} n \\ 2 \end{pmatrix} \begin{pmatrix} m \\ k \end{pmatrix}$

the lines of which are constituted by  $\begin{pmatrix} n \\ 2 \end{pmatrix}$  combination of  $n$  unities and the columns by  $\begin{pmatrix} m \\ k \end{pmatrix}$

combinations of  $m$  variables. To every unit couple (line) and for every matrix it must be associated that combination among the others of  $j$  variables (column) that has the smallest value of distance. Starting from  $k-1$  and for every  $m$  matrices must be considered the  $i$ -esima couple of unit and to this must be associated the couple among the others that has the smallest distance, it would be the couple that has for first term of the combination the second term of the previous couple. The procedure continues until one of the couple is selected another time. All the units that made the "slip knot" are part of the Cluster. The procedure restarts changing the Cluster units with its "centre" and imposing the condition that the units that are part of the successive "slip knot" have among them smaller distances than to their distance from one of the "centres" previously determinate. In the end we will have  $m$  variables that make the relation among every unit and all the other units. Just for an exemplificative purpose the schema and the hypothesis of a research about the classification of economic and financial indicators are realized. It explains the regional unbalance of the bank sector in Italy (Skonieczny and Torrisi, 2002).

## 3. Indicators

The subdivision of territory in homogeneous areas in relation with the studied phenomenon presupposes the individuation of specific indicators able to synthesize the needed statistic information.

Working to this goal and in relation to the hypothesises it was individuate an exhaustive set of variables, and, macro indicators were built; they regard the following aspects of the phenomenon:

1. *real economy*
  - a) unemployment rate
  - b) labour force/population
  - c) GDP/number of employed people
  - d) GDP/population
  - e) consumption of electricity/population
  - f) importations/ GDP
  - g) exportation/GDP
2. *bank-density:*
  - a) GDP/bank counters
  - b) bank counters/ bank places
  - c) surface/ bank counters
  - d) population/ bank counters
  - e) (regional bank counters/ national bank counters)/(regional population/ national population)
3. *bank productivity:*

- a) bank deposit/bank counters
  - b) availments/ bank counters
  - c) number of bank credit/ bank counters
  - d) current account amount/ bank counters
  - e) savings deposits amount/ bank counters
4. *financing productive activity:*
- a) availments/GDP
  - b) availments/population
  - c) availments/deposits
  - d) loans/GDP
  - e)(availments /loans)/GDP
  - f) need of facility credits/total credits
5. *credit intermediation:*
- a) number of bank credit/population
  - b) bank deposit/ population
  - c) current account amount/population
6. *credit agency dimensional:*
- a) small banks availments /tot. regional availments
  - b) medium-big banks availments/tot. regional avail.
  - c) small banks deposits/total region deposits
  - d) medium-big banks deposits/total region deposits
7. *risk and interest rate*
- a) utilized credit/granted credit
  - b) straying/granted credit
  - c) utilized factoring/ granted factoring
  - d) utilized leasing/ granted leasing
  - e) straying leasing/granted leasing
  - f) guarantee granted/population
  - g) unpaid bills/availments
  - h) rate of cash financial decay
  - i)(active rates-passive rates)
8. *interregional financials fluxes:*
- a)deposits influx/deposits flow
  - b) availments influx/ imp.flow
  - c)(dep. infl. - dep. flow)/(dep. infl.+dep. flow)
  - d)(avail. infl.-avail. flow)/(avail. infl.+avail. flow)
9. *families financial activities:*
- a)bank deposits/available deposits
  - b)postal deposits/available deposits
  - c) house financing/availments
  - d)credits - di firma- value
10. *payments system*
- a) amount savings deposits/population
  - b) indirect raising/population
  - c) interest –bearing security/population
  - d) amount money supply/population
  - e) amount ATM operations/population
  - f) amount POS operation/population
  - g) home corporate banking/population
  - h) phone banking/population

## 4. Empiric Analysis

Following the previous list, we used singularly groups of variables to understand the territorial divisions in Italy in relation with every single aspect that defines the various comport mental typologies of the phenomenon. The goal of the proposed technique is to individuate homogeneous groups and the relations among units (the relations among the singular unit and cluster, so that would be possible the identification of the variables that have mainly contributed to the links among couples of units). If we have a table of data where  $x_{ij}$  represent the determination of the variable  $X_j$  per  $j=1, \dots, n$ .  $S_{n, m}$  would be the correspondent table of standardized values and  $d_i$  the most opportune distance. We will determine  $m$  matrices of distances. For this reason the grouping technique "Cluvar" (Skonieczny G, 1995) was utilized. The software package WinCluvar+, developed by Angelo Mazza, was adopted <sup>1</sup>.

A thematic analysis of the phenomenon and of its territorial links was possible using the variables groups, even if it didn't allowed a whole vision as the consideration of the single aspects and of the events caused by their integration.

Often the accumulation of concause gives manifestations and behaviours different from common interpretative schemes.

Composed indicators were used; they were built using weight averages through factorial weights of the considered 55 variables (simple indicators). The cluster analysis was applied on the total of the 10 macro indicators, they were put together to determine a picture on the territorial diversification of the economic-financial system and at the same time the cluster analysis was applied on variables groups (simple indicators) of every macro-indicator with the goal to build the regional differences connected to the various aspects contemplated by every single macro-indicator: real sphere, bank density and productivity, the capacity of financing the productive activities, the bank intermediation and so on.<sup>2</sup>

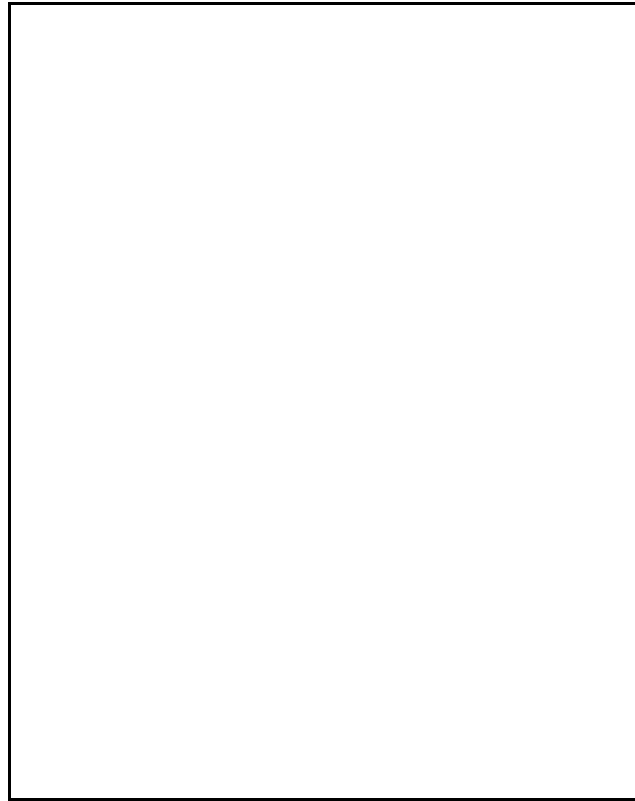
The cluvar technique applied upon the 10 complex indicators produced the results findable in the following schema:

---

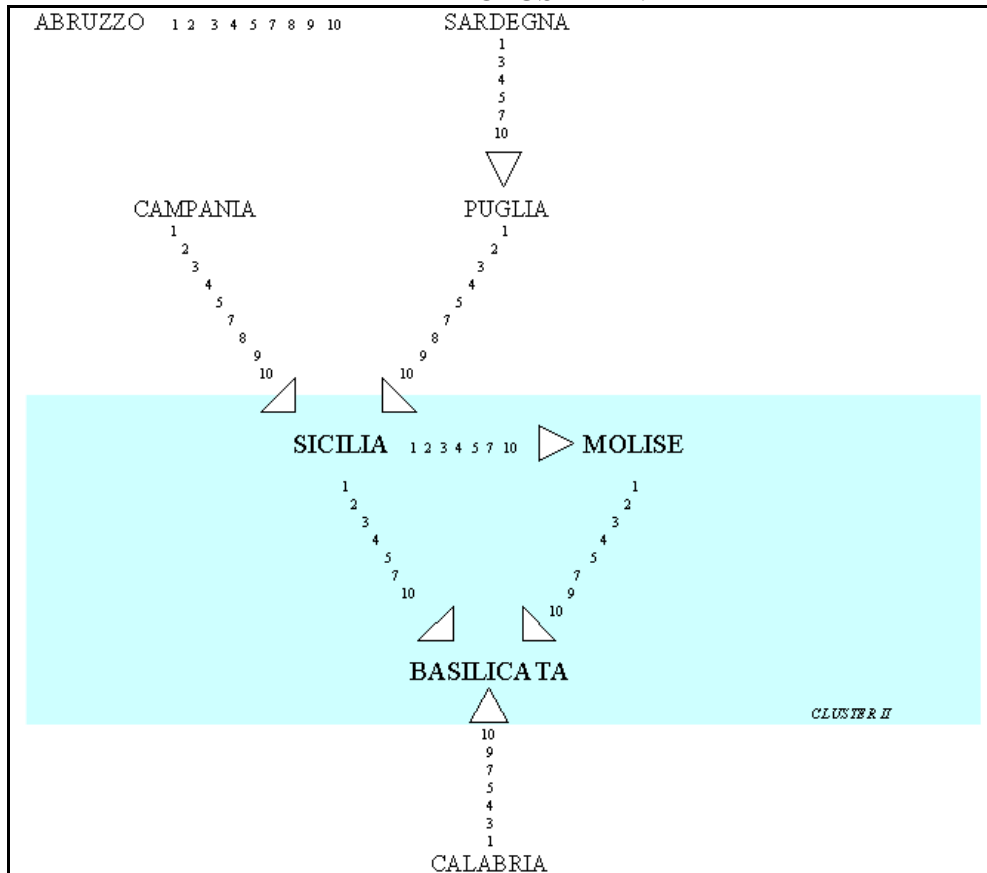
<sup>1</sup> A. Mazza, "Un'applicazione informatica per l'analisi dei cluster in ambito territoriale", *Annali della Facoltà di Economia dell'Università di Catania*, 2000, XLVI.

<sup>2</sup> For the application to the single simple indicators see Skonieczny-Torrisi 2002 "The financial unbalances in Italy" research presented to SIS- Milano.

CLUSTER N. 1



CLUSTER n. 2



## 5. Conclusion

The cluster analysis (applied to the main component obtained through the factorial analysis upon 10 groups of regional indicators) allowed obtaining a picture of the financial national system to underline the divergence between North and South.

The structural elements (that were classified in the previous phase) synthesise the various aspects of the financial system of Italian regions. From a first exam of the groups the country's gap is even confirmed in the financial sector. We can easily see a North characterized by a competitive credit system strongly rising. Centre and South are quite underdeveloped in the financial field.

The first group of regions is characterized by a strong nucleus made by Friuli Venezia Giulia and Valle d'Aosta. The relations between the two regions are distributed upon all the range of indicators. The only two exceptions are: the dimensional variables of the credit bank (6) and the financial interregional fluxes (8), this one for what concern the relation Valle d'Aosta – Friuli Venezia Giulia.

All the other regions interact with the cluster with the modalities and the variables groups underlined in the Graph.n.1.

For what regard South the cluster algorithm allowed the individuation of a cluster including Sicilia, Molise and Basilicata, the relation among these regions are generally characterized by all the variables, with the exception of the variables regarding the dimension of credit agencies (6), regional financial fluxes (8), and for Sicilia - Molise, and, Basilicata – Sicilia even the families financial activities.

The classification technique applied on the total of the regional financial indicators showed an Italy with two velocities in its financial system. Partially the process of group's formation changes but more or less the whole results are the same. This contributes to partially confirm the validity of the determinant factors, synthetic indexes representative of the regional financial differences.

The groups number was chosen to analyse in the best way the various areas of the country, in every group was always maintained a sufficient level of homogeneity.

The cartographic representation of the regional financial structure shows a clear disparity between the two big areas in which Italy is usually divided.

The first group puts together centre-north regions; this group reveal the presence of a large number of really productive banks.

The NEC regions have a decisive importance; they are characterized by high values of “credit bank dimension” factor.

The regions grouped in the second cluster have an opposite situation; this cluster puts together the southern area. The financial system is quite underdeveloped. The credits banks are largely disperse on the regional area. Every bank counter serves averagely a high number of people and a really vast area. In this group the activities of the big banks are predominant, probably this kind of banks has its legal seat in Northern financial centres.

All we read until now can help to explain the reason why economic activities are scarce financed; financing activities are mainly managed by the big northern financial institutes, South dependency from external credit is high, even because this kind of activity is really risk full.

The investment in financial activities is scarce (for ex, state bonds); the utilization of new income and payment instruments is still insufficient.

The separation of the financial southern system is evident under the various considered aspects. Inside this group the behaviour seem to be more homogeneous.

The comparative analysis of the financial sector confirmed the deep difference between North and South. This analysis underlines a kind of classification quite analogue to the one obtained for the economic development. This gap doesn't show evidence of changes.

## References

- Alessandrini P., *Squilibri regionali e dualismo finanziario in Italia*, in “Moneta e Credito” n.177, 1992.
- Arbia G., *Spatial data configuration in statistical analysis of regional economic and related problems*, Kluwer Academic, 1989.
- Arrow K. J., Sala – I – Martin X., *Economic growth*, in McGraw Hill, 1994
- Banca D’Italia, *L’innovazione finanziaria in Italia*, in “Temi Di Discussione” n.69, 1986.
- Brancati. R.,(a cura di), *Settore finanziario e struttura reale nell’economia regionale*, in quaderno del centro ricerche Economiche Angelo Curella, n°4, ed. Giuffrè, 1989.
- Busetta P., *Statistiche finanziarie*, ed. Cedam, Milano 1998.
- Conigliani C., *La fabbrica delle statistiche monetarie e finanziarie*, in *Bancaria* n.2, 1992.
- Cusimano G., *Il sistema bancario nelle due Italie. Alcune evidenze empiriche*, in *Rivista bancaria - Minerva Bancaria*, nr.3, Roma, 1997.
- Cusimano G., *Le statistiche bancarie in Italia*, in Cusimano G. (a cura di), *Le statistiche creditizie*, Quaderni del Centro Ricerche economiche Angelo Curella, n°8, ed. Giuffrè, 1992.
- D’amico N., Parigi G., Trifilidis M., *I tassi d’interesse e la rischiosità degli impieghi bancari*, in Banca d’Italia, *Il sistema finanziario nel Mezzogiorno*, 1990.
- Mazza A., *Un’applicazione informatica per l’analisi dei cluster in ambito territoriale*, in *Annali della Facoltà di Economia dell’Università di Catania*, 2000, XLVI
- Skonieczny G. *Cluvar plus: un metodo di classificazione con la selezione delle variabili*, in *Giornate di analisi dei dati multidimensionale*, Anno1995.
- Silipo D., *Credito e sviluppo territoriale: un’analisi empirica del comportamento bancario*, in “Mezzogiorno Senza Credito”, 1998.
- Torrisi B. , *Economie di scala e di scopo. Un’applicazione della Hybrid traslog cost function al sistema bancario siciliano”* in *Bancaria* Editrice N.7/8 luglio/agosto 2000.