Bugudui, Elena

Article

The use of multivariate techniques for the unemployment analysis

Academic Journal of Economic Studies
Vol. 2, No.2, June 2016, pp. 134–146
ISSN 2393-4913, ISSN On-line 2457-5836

The Use of Multivariate Techniques for the Unemployment Analysis

Elena Bugudui

Department of Finance - Accounting, Artifex University of Bucharest,
E-mail: bugudui@yahoo.com

Abstract *The present paper includes the results of an analysis of different aspects which influence the unemployment evolution in Romania; the analysis was conducted using exploratory multidimensional techniques such as Principal Component Analysis, Factor Analysis and Cluster Analysis. The present research aims to extend the comprehension of the labor market state, offering a formula which reflects the unemployment multidimensionality, the simultaneous action of more factors and establish a framework for the evaluation of the labor market conditions. There are presented different typologies of unemployment, extracted according to the level of education, the unemployment duration and gender for the period 1996-2014 for four age groups. The analysis highlights the specific behavior of the young generation and also of the generation over 50 years.*

Key words Unemployment, principal component analysis, factor analysis, cluster analysis, multivariate analysis

JEL Codes: C33, C54, J21, J70

1. Introduction

The phenomenon encountered in the labor market is a constant concern and a challenge for any company. Human resource is the strength of any economy, the use of labor is among the major sources of global economic growth. Unemployment is a generator of unfair distribution of income, a source of change in the level of conflict of economy and society. Numerous studies conducted by specialized institutions, international bodies such as the World Bank, the ILO, IMF, European Commission, but also by prestigious universities discuss labor market issues. The research of the past decade has gained momentum because of databases availability. The question, however, the limitations in the literature and how policy makers use research results.

Structural component and result of labor market functioning, unemployment is a multidimensional phenomenon, complex, covering a large number of situations: people looking for their first job, especially young graduates of education; people laid off, who have involuntarily lost their jobs for economic reasons; people, mostly women, who wish to resume work after a break; persons employed part-time, casual

or seasonal seeking a job with full work program; elderly people without any source of income seeking work, etc. International studies pays special attention to the multiple dimensions of unemployment who can be enumerated: temporal size structure - dynamic developing economies lead to structural unemployment cyclic apparition in different time periods; spatiality - globalization, technological progress differentiated demographic make changes in the structure of employment in different regions/countries; contingency - differentiated effect of unemployment on future generations; frictionally - offer jobs versus expectations of individual labor market may lead to a short or long-term unemployment, namely the various expenses associated; discrimination and exclusion social discrimination by gender (male/female), ethnic discrimination/other types of discrimination.

2. Literature review

A review of some unemployment approaches and methods used in the analysis is presented below.

The paper "*Explaining European Unemployment*" of O.J. Blanchard (Research Associate at the NBER's Programs on Monetary Economics and Economic fluctuations and Growth and a professor of economics at MIT, 2004), introduces the analysis, which can be today by topical dimensions: unemployment shocks and institutions (Blanchard, 2004).

Based on new evidence about the time evolution of institutions, Steven Nickell (Nickell, 2003) claimed that adaptation institutions played an important role in the evolution of unemployment. For example, the Dutch trade unions have accepted the argument of the companies that they need a restoration of profit margins in order to increase employment. This is stated and Thomas Philippon (Blanchard and Philippon, 2003) trust between capital and labor is a factor that can influence the evolution of unemployment, thus explaining most of the differences in the evolution of unemployment between different countries. In another work, Blanchard (2004) appreciates that reforms in financial markets will force reforms on the labor market. The method used in the study by O.J. Blanchard was the panel approach. As for, approach the dimension of social exclusion (gender, religion, ethnicity, age), it is found in the paper "*The Multidimensional Analysis of Social Exclusion*" by Ruth *et al*. (2007) from the Faculty of Sociology and School for Social Policy, Townsend Centre for International poverty study and Bristol Institute for Public Affairs University of Bristol. Another work that addresses the multidimensional phenomenon of unemployed is "*Unemployment Index: A Multidimensional Measure of Labor Market Efficiency*," by Della Lee Sue (Sue, 2008). In the study are inserted simultaneously the incidence and duration of unemployment. The incidence refers to the probability, that a person, will face the state of unemployment and the duration

refer to the time spent by a person unemployed. This paper provides a basis for creating an *index of unemployment* in which different components of unemployment are highlighted and used together as an indicator of labor market efficiency. This indicator can be useful for setting policy on the labor market when both scope and duration are relevant in assessing economic conditions and in comparing unemployment between different groups identified, gender, ethnicity, age, marital/ residence region and so on.

Leighton and Mincer (1982), show that time spent unemployed person is wasted time for the labor market, in this context, tall tales and his "*Regional Unemployment, Gender and Time Allocation of the Unemployed*" by Gimenez-Nadal and Molina (2012) which shows the relationship between time allocation decisions by the unemployed, both women and men and the regional unemployment rate. The results are consistent with Burda and Hamermesh (2010) and Aguiar *et al.* (2012) study was conducted on a sample of 4122 people, with 1884 men and 2238 women, i.e. 2002-2003, 2009-2010, Disabled aged between 21 and 65 years inclusive. as a method of analysis and estimation method was used regression with OLS, with the series for men and women, with the variables of time between activity reference to individual and from region j at t time, the unemployment rate in region j at time t and a matrix containing the individual characteristics of the region j i at time t; a matrix of dummy variables to highlight a day of the week; a control variable for the region j.

Besides the above mentioned methods and techniques used in the analysis of unemployment size and can include those listed in "*Recent Advances Research on Support Vector Machines*" (Yngjie *et al*., 2012).The paper refers to the programming tools used in economic applications, finance and management that facilitates weather high quality Cao and Tay (2003), compares SVMs with back propagation multilayer (BP) neural networks and regulating the basic function of radial (RBF). "*Creating and Manipulating Tables with Multidimensional OLAP Cubes Using Location Data*" by Konrad Dramowicz (2005) includes tools that enable customers OLAP query, browsing and synthesizing information into a highly efficient, interactive, and dynamic.

An advanced technique for analyzing large volumes of data used is the Data mining, a type of predictive analytics by applying algorithms to extract information from data (Edelstain, 1999). Data mining finds multiple applications in research. Data mining techniques are suitable macroeconomic analysis, regional.

3. Methodology of research

The methodology used in the present research is the application of methods Principal Component Analysis, Factor Analysis and Cluster Analysis in the age groups, but also at conjugate views.

Both Principal Component Analysis and Factor Analysis (Beaumont, 2012), is exploratory multivariate techniques, aimed, on the one hand reduce the dimensionality to speed data analysis, initial variables replacing it with new, representative and uncorrelated and, elsewhere, identifying hidden patterns and relationships between variables (AF). Principal Component Analysis (Abdi and Williams, 2010), can be regarded as a technique to analyze the factors or algorithms a factorial analysis, principal component analysis implies. If we just want to reduce the dimensionality and visualize data in a smaller space, use Principal Component Analysis, if we want to realize a model to explain the correlations between variables, we continue to apply the algorithm and Factor Analysis. The two methods can be used only with the help of specialized software, such as SPSS, Statistica, Matlab, EViews.

The representation of space dimensionality original items are considered as points in a space whose axes are the characteristics of those objects (Ex: Q32004: RS_MASC, RS_FEM, RS_SUP, RS_SEC, RS_PRIM, RS_LD; RS_SC,) in this situation, characteristic values are projections point to the axes of space (coordinates). This representation leads to great losses of information, so is reduced dimensionality space (other axles) and thus create new features. In this case, the coordinates will be different and will lead the new features. New features (variables) are called Principal Components and coordinates these new axes are called scores. To be relevant and appropriate analysis, the new coordinates, must contain the information content of the original characteristics. Transform envisages a maximum conservation information and minimal information loss. The new coordinates are orthogonal and maximizing variance. The Need to reduce dimensionality occurs for Several Reasons: the many more variables, diminishing the significance of each variable; could result in multicolinearity or if colinearity. Some have strongly interrelated and not least, the existence of several variable leads to a very complex calculation, in addition, viewing positioning objects, it is very difficult, if more than three variables due to limitations plotting. Even three variables can be too much, because plotting is no longer relevant.

Principal Component Analysis allows graphical representation of objects that are characterized by a large number of variables. The number of principal components obtained as linear combinations of the original characteristics, can be equal to the initial features, only that charging information is different (variance). The first principal component is a linear combination of normalized maximum variance (Ruxanda, 2009), the second CP is unrelated to the first, has ranged as high as possible, but not higher than the first and a, m, d. Principals Components has orthonormal vectors. Main components will each retain many original version maximum explain (those components for which the cumulative variance is around

80% of the variance initial space). In the space of initial features, subsets can be formed, made up of very highly correlated features. Usually, the number of subsets Principal Component gives us the number detained (in our case, the age groups 25-34 years, 35-49 years and 50 years and above is very strong correlations observed in two subsets, that which is anticipated that the first two Principal Components, significantly explain the initial version).

Analysis Cluster used in the present study aims to identify clusters, classes (periods) within the age groups studied, the unemployment rate according to various criteria such as level of education (RS_SUP, RS_SEC, RS_PRIM) and duration (RS_SD, RS_MD, RS_LD) saw-like characteristics and values. It checks the similarities or differences between different periods analyzed belonging to the same age groups, to obtain homogeneous classes, where unemployment short- or long term was mostly, or the different social categories, by level of education, they was more or less affected by unemployment. In summary, the algorithm for the realization of classes (clusters) is as follows (Ruxanda, 2009):

- choosing the variables based on which the groups/clusters will be made (in our case, unemployment by educational level and duration);
-choosing the distance for the determination of proximity (in our case the Euclidean distance that measures the distance between two points in a straight line);
- establishing the Principles for forming the groups (within each age group);
- the clusters will be built;
- checking the significance and consistency of the clusters;
- choosing a suitable number of groups;
- interpreting the significance of groups,

The Discriminant/Cluster Analysis allows the identification of the affiliation of unemployed people to certain classes/groups/clusters with well-defined individuality, allows a better correlation of the unemployment phenomenon with other socio-economic phenomena such as: the realization of a certain type of retraining, the optimal allocation of resources on economic activities at a territorial level, the evolution of living standards.

This process can be used to search for the answer to the evolution of unemployment and labor migration, with the phenomenon of globalization which includes Romania too.

The data used come from statistical research "*Household Labour Force Survey*", the National Institute of Statistics. The source variables used in the research are: the unemployment rate by level of education and unemployment by duration for four age groups.

4. The results of the multivariate analysis

The results obtained by applying the principal component analysis(PCA), the age group 15-24 years, reveals information about the variables that define the new space, the number of 7 (equal to the initial variables) loading each principal component and information on their dispersal.

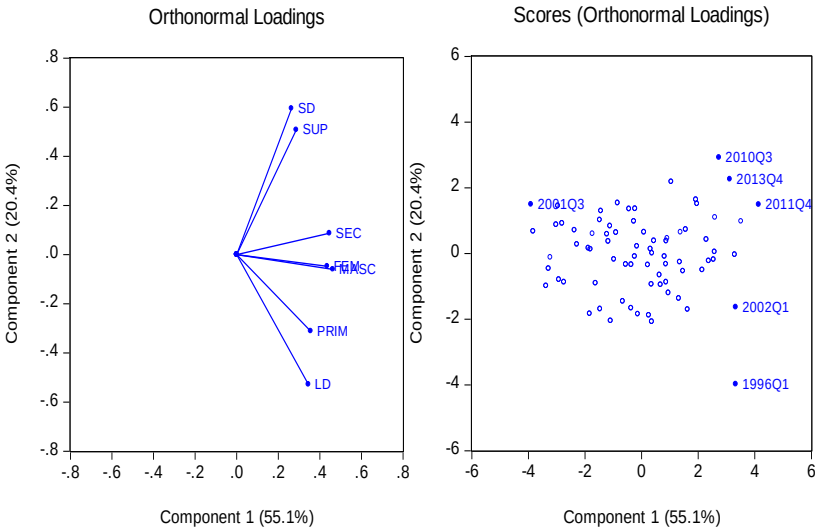


Figure 1. Representing the Principal Components for the age group 15-24 years

Subsequently transformation algorithm and obtaining CP can continue to identify the principal factors.

Factor analysis aims to reduce the dimensionality of research, but also to identify hidden relationships between baseline characteristics and obtaining patterns in the evolution of a phenomenon.

Similar to PCA, factor analysis from original reconfigures space, reducing the dimensionality by replacing the original variables with abstract factors, which preserves a rate or other variability (variance) original space. The new space is defined by common factors (latent) and a single factor.

Data show that our model allows for one principal factor. The principal factor is a linear combination of variables with coefficients presented in Table 1. From estimating variable dispersion is observed that the unemployment rate for the male

population has the highest dispersion, the dispersion is lower for short-term unemployment.

The principal factor assigns weights close enough for all 7 baseline variables. It notes, however, a small difference in that which concerns variables and SEC MASC an interpretation would be that unemployment in this age group is sensitive to all the features considered, special attention being paid to young graduates of secondary education.

Table 1. The principal factors for age group 15-24 years

Factor Method: Principal Factors					
Date: 06/06/16 Time: 13:12					
Covariance Analysis: Ordinary Correlation					
Sample: 1996Q1 2014Q3					
Included observations: 75					
Number of factors: Minimum average partial					
Prior communalities: Squared multiple correlation					
	Loadings				
	F1	Communality	Uniqueness		
SUP	0,497815	0,247819	0,752181		
SEC	0,867856	0,753174	0,246826		
SD	0,488558	0,238689	0,761311		
PRIM	0,689835	0,475872	0,524128		
MASC	0,903593	0,816481	0,183519		
FEM	0,837076	0,700696	0,299304		
LD	0,665508	0,442901	0,557099		
Factor	Variance	Cumulative	Difference	Proportion	Cumulative
F1	3,675633	3,675633	---	1,000000	1,000000
Total	3,675633	3,675633		1,000000	
	Model	Independence	Saturated		
Discrepancy	0,594876	5,454526	0,000000		
Parameters	14	7	28		
Degrees-of-freedom	14	21	---		

Looking at all age groups and levels of education, it is easy to see that the highest level of unemployment is within young people (particularly age group 15-24) for the average level of education. The observations in the area of factors reveal that the joint action age - training will continue to have even a greater effect on this typology, most observations being placed above the modeling surface of the unemployment rate. Subsequently transformation algorithm and obtaining hp, can continue to identify the principal factors.

Principal Component Analysis results for the age group 50 years and over show that it can retain two main components, explaining them in a proportion of 81.2% of the original space evolution characteristics. The first principal component explains 69.7% contains the most information, the second component explains the phenomenon evolution proportion of 11.5%. From original contribution to the first principal component and this time are minor differences. The second major component is dominated by unemployment among those with higher education.

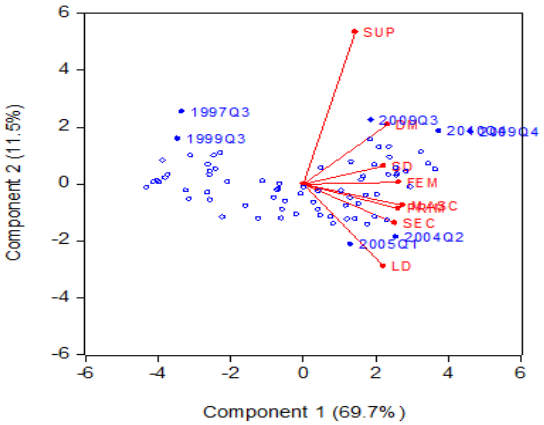


Figure 2. Representing the Principal Components for the age group 50 years and over

Analysis on all age groups and levels of training, we notice that the highest level that it touches youth unemployment (particularly age group 15-24) to the average level of training. Observations from space reveal factors that joint action age - training will continue to exert a greater effect on this typology, most observations placing it above the surface modeling unemployment rates.

The Principal Component Analysis carried out for all the observations on unemployment rate by level of education and age group in the 1996-2014 period reveals a strong simultaneous separation of the three variables taken into consideration, which is not the case for the analyzes within each age group.

Secondly, their connection with the real variables is very obvious; while the first composite factor almost merges with the age (direct effect) and unemployment (in reverse), the second one is strongly correlated with level of education. In other

words, the unemployment typology can be defined with differential rates by age groups, on one hand and by level of education, on the other hand.
Therefore it is sufficient to look at this phenomenon from the perspective of rates (which capture the effect of age) and educational level; all other factors that can be taken into account have a custom action only at the level of detail (groups only during certain periods, gender or other types of structuring of the phenomenon).
It can be observed the different behavior in relation with the non-occupation of workforce; all groups have a peak corresponding to the crisis period (2008-2010), with significantly different attenuations after 2011. Accentuated decreases in unemployment rate were recorded in the group 25-34, while in the older group (over 50 years) economic recovery does not bring significant changes in terms of unemployment.

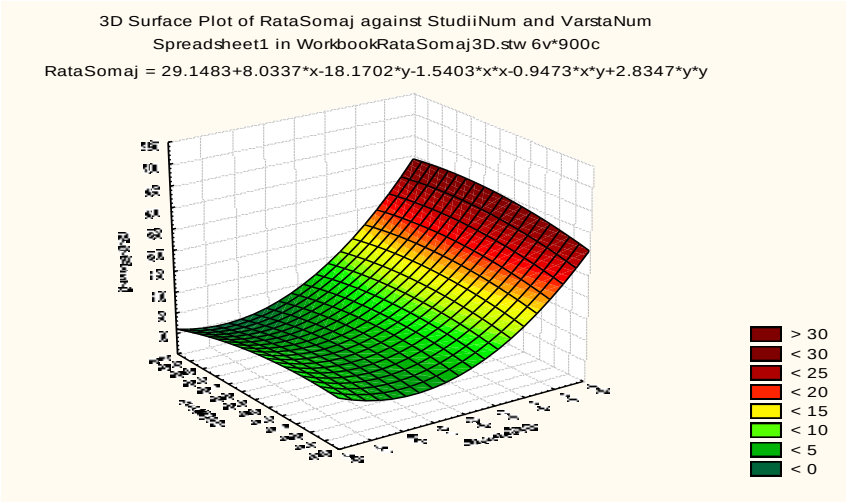
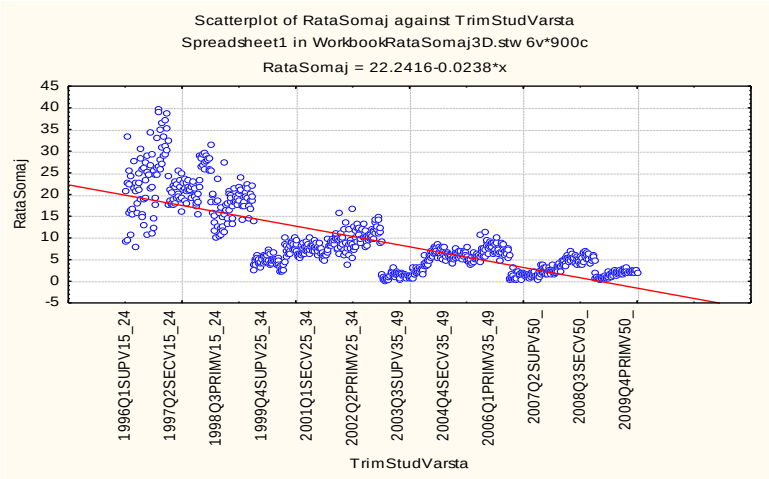


Figure 3. The typology of unemployment by educational level and age

Concerning extracting some typologies, the unemployment rate determines by clustering compact groups, which do not necessarily take into account age groups, leading to mixed classes. For example, the most compact group is the 15-24 years (the top in the chart) due to the homogeneity of high rate throughout the years. Interestingly, this group is associated marginally, with the group 25- 34, but only those with low level of education (primary).
The broader category which breaks down into sub-clusters is the 35-49 age group, with all the levels of education (primary, secondary and tertiary);at this group there

are added those from the age group over 50 years, but only those with average and high educational level that were maintained in the work field, largely for transferring knowledge between generations.



Source: Personal processing in Statistics program

Figure 4. Unemployment rate simultaneously by education and age

The figure shows the types of unemployment by age group and level of education. The Group of 15-24 years has very scattered values; the highest value is achieved in 1996-1997 and decreases by 1999.

The Group of 25-34 years is much more stable, representative are the early 2000 s. The Group of 35-49 years is centered on a 5% rate with representation in the 2003-2006. The Group over 50 years has a minimum value before the crisis and is the first one affected by the crisis.

To extract typologies of unemployed according to the level of studies using cluster analysis. Clustering results show a compact group, not necessarily take into account the age groups, leading to mixed classes. The compact group occurs in the age group 15-24 years (top in the chart) due to the high rate homogeneity throughout the years. This group is associated group 25-34, only those with low level of education (primary).

Largest category, which in turn decomposes in most sub-clusters, is made up of the 35-49 age groups, with all levels of education (primary, secondary and higher); she

joins the group aged over 50, those with medium and high level training, kept in employment due to the transfer of knowledge between generations.
End clustering indicates a typology where the group returns to 25-34, but the high level of preparation, or at least medium level.

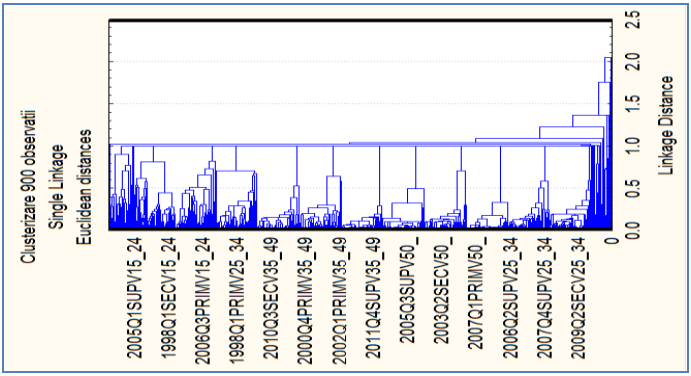


Figure 5. Typology unemployment by education level and age

5. Conclusions

As expected, the analysis showed that young people have a particular behavior towards other participants in the labor market. The analyzed period is characterized by short-term unemployment among youth. This type of unemployment shows frictions in the labor market, inadequacy, dissatisfaction or a situation of combining studies with work (exam periods, holidays). The desire to have an income alternates with periods in which the needs are not expressed. For the 50 years and more age group, we observe specific behaviors; there are high levels of long-term unemployment. The unemployment for more than a year and the very long-term unemployment, 24 months and more, can be caused by weak professional qualification (mismatch between skills, abilities and labor market requirements), personal problems to be solved preferential (problems health, children), physical characteristics and health, but also the psychological profile of those who are looking for a job. A long-term unemployment can cause discouragement when searching for a job, leading to very long term unemployment or even to not integrating in the labor market, without an influence economically speaking. Among young people, long-term unemployment appears due to lack of skills, or inadequate qualifications, possible mismatch in the labor market or the existence of alternative sources of income (parents).

References

Abdi H., Wilims, L.J. (2010), *Principal Components Analysis*, Wiley Interdisciplinary Reviews: Computational Statistics, Volume 2, Issue 4, John Wiley & Sons, Inc., pp. 433-459.

Anghelache, C., Bugudui, E., Gresoi, S., Niculescu, E. (2006), *Applied Statistics. Indexes, summaries and case studies*, Economic Publishing House, Bucharest, pp. 79-91).

Aguiar M., Hurst E. and Karabarbounis L. (2012), *Recent Developments in the Economics of Time Use*, Annual Review of Economics, Volume 4.

Beaumont, R., (2012), *An introduction to Principal Component Analysis & Factor Analysis Using SPSS 19 and R*, http://www.floppybunny.org/robin/web/virtualclassroom/stats/statistics2/pca1.pdf

Blanchard. O.J. (2004), *Explaining European Unemployment*, NBER's Programs on Monetary Economics and Economic Fluctuations and Growth.

Blanchard, O.J. (2004), *The Economic Future of Europe*, NBER Working Paper No. 10310, March 2004, and forthcoming in the Journal of Economic Perspectives.

Blanchard O.J. and Philippon, T. (2003), *The Decline of Rents, and the Rise and Fall of Unemployment in Europe*, forthcoming as an NBER Working Paper.

Bugudui, E., (2009), *Quantitative methods used in market research*, Artifex Publishing House, Bucharest.

Burda, M.C. and Hamermesh, D.S., (2010), *Unemployment, market work and household production*, Economics Letters, Elsevier, vol. 107(2), pp. 131-133, May.

Cao L.J. and Tay, F.H. (2003), *Support vector machine with adaptive parameters in financial time series forecasting,* IEEE Trans Neural Netw. 2003;14(6):1506-18.

Dramowicz K. (2005). *Creating and Manipulating Tables with Multidimensional OLAP Cubes Using Location Data*, http://www.directionsmag.com/entry/creating-and-manipulating-multidimensional-tables-with-locational-data-usin/123565

Edelstein H.A. (1999), *Introduction to data mining and knowledge discovery* (3rd ed), Potomac MD: Two Crows Corp.

Gimenez-Nadal J.I. and Molina J.A. (2012), *Regional Unemployment, Gender and Time Allocation of the Unemployed*, Review of Economics of the Household, March 2014, Volume 12, Issue 1, pp. 105-127.

Leighton L. and Mincer J. (1982), *Labor turnover and youth unemployment* in R.B. Freeman and D.A. Wise (eds), The youth labor market problem, Chicago: University of Chicago Press.

National Institute of Statistics, (2016), Tempo online, https://statistici.insse.ro/shop/

Nickell, S. (2003), *Labour Market Institutions and Unemployment in OECD Countries,* CESIFO DICE Report 1, No. 2 (2003), pp. 13-26.

Ruth, L., Pantăzis, C., Fahmy, E., Gordon, D., Lloyd, E., and Patsios, D., (2007), *The Multidimensional Analysis of Social Exclusion*, Bristol Institute for Public Affairs University of Bristol.
Ruxanda, Gh., (2009), *Multivariate data analysis*, Economic University in Bucharest.
Sue Della Lee (2008), *Unemployment Index: A Multidimensional Measure Of Labor Market Efficiency*, New York Economic Review, vol. 39, Fall 2008, p. 404.
Yingjie T., Yong S., Xiaohui L. (2012), *Recent Advances on Support Vector Machines Research,* Technological and Economic Development of Economy, issue 18, 2012, no. 1, pp. 5-33.