

## Principal Components Analysis of Employment in Eastern Europe

Savić Mirko\*

**Summary:** For the last decade, the employment structure is one of the fastest changing areas of Eastern Europe. This paper explores the best methodology to compare the employment situations in the countries of this region. Multivariate statistical analyses are very reliable in portraying the full picture of the problem. Principal components analysis is one of the simplest multivariate methods. It can produce very useful information about Eastern European employment in a very easy and understandable way.

**Key Words:** Employment, Multivariate analysis, Principal components analysis

**JEL:** J21, C82

### 1. Introduction

The former communist countries missed at least a half-century of normal economic development. The nature of their growth brought about serious structural distortions in their economies, making them highly inefficient compared to the rest of the world. This led to an unavoidable output decline after the collapse of communism. Making up for this lost time will take at least 15 years in the case of the Czech Republic, 20 years in the case of Hungary and Poland, and 30 years in the case of Romania.

According to Schmitt and Baker (2006), over the last five years, the EU has nearly closed the employment gap with the United States. The small remaining difference is almost entirely due to low employment rates among women in Italy and Spain, two large economies with weak welfare-state institutions and long traditions of low female employment. New EU members from Eastern Europe could spoil this balance of employment.

Labour force activity rates increased in most Western European countries during the last decade. In the Eastern European countries the employment situation was different, although experience from other countries shows that trade reforms likely contributed to accelerated employment growth (Goldar (2002)). The economic activity rates have decreased for both men and women, the decrease for men often being larger. In a majority of eastern countries women constitute a larger proportion of the total labour force - between 40 and 50%. Turkey provides an exception because it has considerably less women in the labour force – fewer than 30%.

Total employment as a percentage of the total population is an indicator of a country's capacity to support its population. Bosnia and Herzegovina, The Former Yugoslav Republic of Macedonia and Serbia have the lowest employment ratios in all of Europe. The economic activity rates in these countries are not greatly below average, and the reasons for the low ratio can be the high unemployment in Bosnia, Herzegovina, and Serbia.

The last decade has seen a general trend in Eastern European countries increasing service sector employment at the expense of the agricultural and industrial sectors. There are exceptions: the share of total employment in agriculture is still high in Albania (72%), Moldova (51%) and Romania (43%).

Women are less likely than men to be employers or self-employed workers in Eastern Europe, especially in Turkey and Albania. The proportion of employed women among these groups is over 50 % only in Moldova and Belarus with Ukraine, Lithuania, Estonia, Latvia and Russia slightly below 50%.

Lithuania has the highest proportion of women (47 %) among legislators, senior officials, and managers in Eastern Europe. In the rest of the region there is a clear majority of men in these occupations, with the lowest recorded proportion of women in Turkey at 8%.

Part-time employment remains a female domain, but varies considerably among countries. In Moldova the proportion of all employed women working part-time is less than 1%. In most of the countries, the trend has been increasing since 1990. In general, employed persons in Eastern Europe have longer working hours than employed persons in Western Europe (except in Greece and Iceland). In all countries men spend more time in paid work than women, but the difference between women and men is more pronounced in Western Europe than in Eastern Europe.

Employment is a very dynamic and complex economic issue. Its intricacies can be scientifically examined using quantitative methods. Multivariate statistical analysis offers a range of methods and techniques: discriminant analysis, conjoint analysis, principal components analysis, AID and CHAID methods, factor analysis, cluster analysis, correspondence analysis etc. These methods are gradually finding their place in varying segments of economics. At the moment, marketing research represents the area most intensively using multivariate statistical methods. If these procedures are valuable in marketing, it is only logical to test their applicability to other areas of economic science.

Principal components analysis is one of the well known multivariate statistical methods. The main reasons behind undertaking this research is to implement this method into analysis of employment structure to prove its usefulness and to compare performances of different Eastern European countries in creating jobs in their respective economies.

## 2. Principal Components Analysis

The objective of principal components analysis is to take  $p$  original variables ( $X_1, X_2, \dots, X_p$ ) and find combinations of these to produce indices or new variables ( $Z_1, Z_2, \dots, Z_p$ ). These new variables are uncorrelated in order of their importance, and describe the variation in the data. The lack of correlation means that the indices are measuring different “dimensions” of the data, and the ordering is such that  $\text{Var}(Z_1) \geq \text{Var}(Z_2) \geq \dots \geq \text{Var}(Z_p)$ . The new variables are then the principal components. In principal components analysis, there is always the hope that the variances of most of the new variables will be so low as to be negligible. In this case, most of the variation in the full data set can be adequately described by the few  $Z$  variables featuring variances that are not negligible.

The best results of principal components analysis are obtained when the original variables are very highly correlated, either positively or negatively. If that is the case, then it is quite conceivable that 20 or more original variables can be adequately represented by two or three principal components. If this desirable state of affairs does occur, then the important principal components will be of some interest as measures of the underlying dimensions in the data. It will also be of value to know that there is a good deal of redundancy in the original variables, with most of them measuring similar phenomenon.

A principal component analysis will help us make the plot of Eastern European countries against their values for two principal components. The picture is rather meaningful in terms of what is known about employment in this region. The countries with similar employment situations will be grouped together, and it is possible to see the position of each country in comparison with other countries.

The procedure for principal component analysis starts with data on  $p$  variables for  $n$  individuals. The first principal component is the linear combination of the original variables ( $X_1, X_2, \dots, X_p$ ):

$$Z_1 = a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p \quad (1)$$

that varies as much as possible for the individuals, subject to the condition that

$$a_{11}^2 + a_{12}^2 + \dots + a_{1p}^2 = 1. \quad (2)$$

Thus the variance of  $Z_1$ ,  $\text{Var}(Z_1)$ , is as large as possible given this constraint on the constants  $a_{1j}$ . If the constraint were not introduced, then  $\text{Var}(Z_1)$  could be raised by simply increasing any one of the  $a_{1j}$  values. The second principal component is

$$Z_2 = a_{21} X_1 + a_{22} X_2 + \dots + a_{2p} X_p \quad (3)$$

This is chosen so that  $\text{Var}(Z_2)$  is as large as possible subject to the constraint that

$$a_{21}^2 + a_{22}^2 + \dots + a_{2p}^2 = 1 \quad (4)$$

and also to the condition that  $Z_1$  and  $Z_2$  have zero correlation for the data. Further principal components are defined by continuing the same way. If there are  $p$  variables, then there will be as many as  $p$  principal components.

The steps in a principal components analysis:

- 1) Code the variables to have zero means and unit variances. Sometimes this is omitted when it is thought that the importance of variables is reflected in their variances. In the case of employment, the variables will have equal importance.
- 2) Calculation of the covariance matrix. This is a correlation matrix if step 1 has been done.
- 3) Find the eigenvalues and the corresponding eigenvectors. The elements of the eigenvectors matrix are the coefficients of the principal components, while the eigenvalues are their variance.
- 4) Discard any components that account for only a small proportion of the variation in the data. For example, for the 20 variables there will be 20 principal components but only the first four components account over 90% of the total variance. On this basis, the other 16 components may reasonably be ignored.

The principal components analysis gives us very useful information about differences in employment structures, but it could be just the start for more serious multidimensional analyses. One example is factor analysis. This method based on the principal components analysis helps to reduce a vast number of variables (for example, all the questions tapping several variables of interest in a questionnaire) to a meaningful, interpretable, and manageable set of factors.

### 3. Data

The data were collected from UNECE's (United Nations Economic Commission for Europe) statistical database. This information is maintained by the Statistical Division of the UNECE Secretariat. It provides detailed statistical information on countries in Europe, North America and Central Asia. These are probably not the latest data on employment in the Eastern European region, but it is fresh enough to learn the insights.

The statistics appearing in the UNECE publications have been compiled from a wide range of national and international sources. Data for the tables have three main sources: the replies of the National Statistical Offices to annual ECE questionnaires, the published or unpublished data collections of other international organizations, and the official national sources.

Only a few variables were taken for analysis, but this is just a start for further and more detailed scientific research. Variables included in the model are: percentage of employment in agriculture, percentage of employment in

industry, percentage of employment in services, percentage of women in the labour force, unemployment rate, and youth unemployment rate.

For employment the National Accounts definition is used: employment comprises employees and self-employed - engaged in some productive activity that falls within the production boundary. It includes both the residents and the non-residents who work for resident producer units. Employment by major economic sectors is based on International Standard Classification of Occupations 1988 (ISCO-88). Unemployment rate is calculated by relating the number of workers who are unemployed during the reference period to the total of employed and unemployed persons at the same date. Registered unemployment comprises unemployed population registered at Employment or Labour Offices. This administrative approach to unemployment reflects national rules and conditions and usually yields different results from those of surveys using the ILO concept of unemployment, which includes persons often not covered in registered unemployment statistics, such as persons seeking work for the first time. Youth unemployment refers to 15-24 years of age.

#### **4. Results**

It is appropriate to begin with step 1 of the aforementioned analysis process. Standardization of the measurements ensures that they all have equal weight in the analysis. Procedure of standardization was conducted on the basis of the following formula:

$$z_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_i} \quad (5)$$

where:

$z_{ij}$  = standardized value of original value  $j$  from variable  $X_i$

$x_{ij}$  = original value  $j$  from variable  $X_i$

$\sigma_i$  = standard deviation for variable  $X_i$

$\bar{x}_i$  = sample mean for variable  $X_i$

The covariance matrix for the standardized variables is the correlation matrix. The correlation matrix for all variables shows that correlation variables are not particularly high, which indicates that several principal components will be required to account for the variation in the data. The eigenvalues of this matrix are shown in table 1.

**Table 1:** Eigenvalues of correlation matrix, and related statistics (Employment - EEC + Developed Countries in Employment in Eastern Europe) Active variables only

-	Eigenvalue	% Total	Cumulative	Cumulative
1	2.633371	43.88952	2.633371	43.8895
2	2.065042	34.41736	4.698413	78.3069
3	0.821423	13.69038	5.519836	91.9973
4	0.371016	6.18360	5.890851	98.1809
5	0.109108	1.81846	5.999959	99.9993
6	0.000041	0.00069	6.000000	100.0000

The corresponding eigenvectors are shown in table 2, standardized so that the sum of the squares of the coefficients is one for each of them. These eigenvectors are providing the coefficients of the principal components.

**Table 2:** Eigenvectors of correlation matrix (Employment - EEC + Developed Countries in Employment in Eastern Europe) Active variables only

-	Principal Components					
	1	2	3	4	5	6
<b>Agriculture</b>	0.278519	-0.608002	-0.189799	-0.085429	0.015234	0.713587
<b>Industry</b>	-0.534732	0.142312	-0.100197	0.728522	-0.015719	0.390867
<b>Services</b>	0.015761	0.648560	0.303039	-0.385817	-0.022126	0.581330
<b>% of Women</b>	0.137201	0.365365	-0.902360	-0.093115	0.157320	0.003238
<b>% Unemployment</b>	-0.559808	-0.183099	0.004075	-0.361583	0.722708	0.004858
<b>% Youth Unemployment</b>	-0.551389	-0.149781	-0.218709	-0.416761	-0.672291	-0.006121

The Scree Method is one approach to deciding how many principal components to select as the most important. The analyst looks for natural break points in the percentage of total variance dependent from each principal component.

It is a matter of judgment as to how many components are important. To some extent, the choice of the number of components that are important will depend on the use that is going to be made of them. For the present case, it will be assumed that a small number of indices are required in order to present the main aspects of difference between the countries, and for simplicity only the first two components will be examined further. Between them, they account for about 78% of the variation in the original data.

The first component is

$$Z_1 = 0.278519 X_1 - 0.534732 X_2 + 0.015761 X_3 + 0.137201 X_4 - 0.559808 X_5 - 0.551389 X_6$$

As the analysis has been done on the correlation matrix, the variables in this equation are the original values after they have each been standardized to have a mean of zero and standard deviation of one.

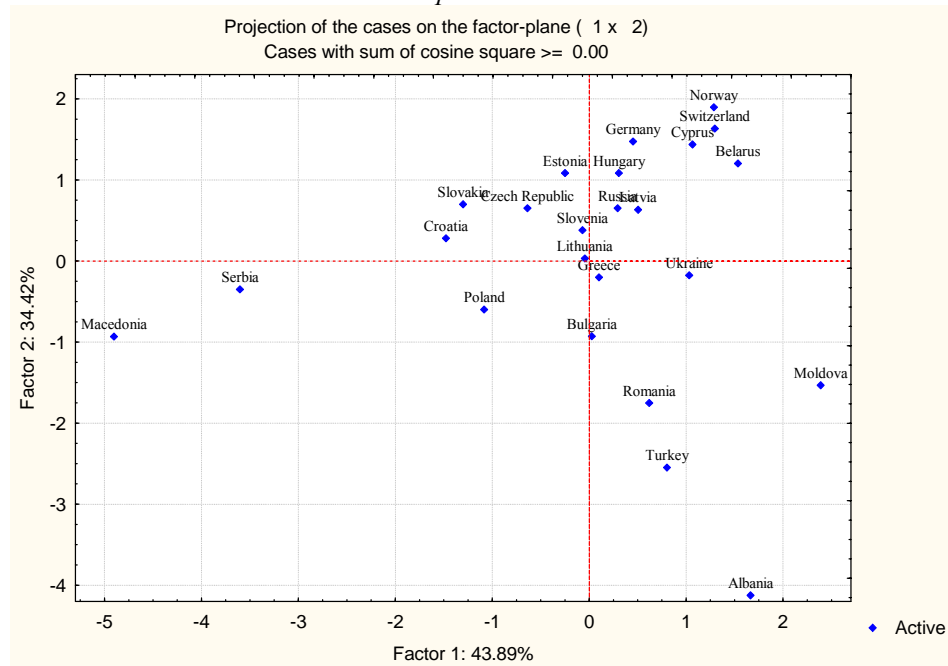
The second principal component is

$$Z_2 = -0.608002 X_1 + 0.142312 X_2 + 0.64856 X_3 + 0.365365 X_4 - 0.183099 X_5 - 0.149781 X_6$$

Figure 1 shows a plot of 21 Eastern European countries with three Western European countries (Norway, Switzerland and Germany) for comparative estimates. The plot shows the positions of the countries against their values for the first two principal components. The picture is rather meaningful in terms of what is known about the employment in the countries.

The plot on figure 1 is the result of principal components analysis. The use of this result is obvious. With one look at the plot one can get the full picture about the employment in the region and the position of each country.

**Figure 1:** *The plot diagram according to scores for the first two principal components*



The Western European countries are grouped in the upper left corner with very similar employment situations in all three major economic sectors (agriculture, industry and services) and with similar values for almost every other variable.

It is interesting that the most similar employment structure to Western European countries is found in Cyprus, Belarus, and Hungary. All countries from the former USSR are very close to one another, with the exception of Moldova. The reason for this is that Moldova has a larger percentage of

employees in agriculture, the percentage of employed women is above 50%, and the percentage of total unemployment is very low.

The most dramatic difference in employment structure is clearly in Serbia, Macedonia, and Albania. Serbia and Macedonia have the same problems: very low percentage of employment in the agricultural sector and the largest percentage of total unemployment in Europe. Albania has the opposite problem: the greatest percentage of employment in the agricultural sector and very low percentage of employed women.

OECD countries mostly display unemployment with too little variance over time according to Bassanini and Duval (2006). In Eastern Europe the unemployment rate is more varied than in the rest of the continent, but the majority of countries have higher unemployment rates in recent years. Most countries feature small differences between women and men in unemployment rates; however, in Greece, Turkey, and Albania considerably more women are unemployed than men. In Lithuania the situation is exactly opposite – the unemployment rate for men is higher than for women.

In a majority of regional countries there is a higher youth unemployment rate for women than for men, but in many countries the differences are not large. These disparities may be subject to annual variations based on which sectors of the economy are hardest hit by unemployment. The youth unemployment rate is generally higher than the overall unemployment rate throughout the Eastern European region.

Some countries have more than half their unemployed in that position for much longer than 12 months. The recorded long-term unemployment rate is highest in Albania, Slovenia (65 %), and Bulgaria (62 %).

The unemployment rates are high for all Eastern European countries except Belarus, Cyprus, Moldova, and Ukraine. Since by present standards any rate below 3% can be regarded as exceptionally good, Moldova would stand out in Eastern Europe. Ten countries in the region (Albania, Bulgaria, Croatia, Estonia, Latvia, Lithuania, Macedonia, Poland, Serbia, and Slovakia) have double-digit rates and hence can be classified as countries with extremely worrisome unemployment rates. The same can be said about the other four nations since their own rates exceed 7% (Czech Republic, Greece, Russia, and Turkey). In this arbitrary classification, only Moldova can claim to have performed extremely well. The reader should note that absolute levels of unemployment for these countries are not strictly comparable owing to differences in measurement techniques.

When it comes to employment by sectors, the relative size of the service sector is of particular interest. This is mainly the tertiary sector, comprised of such divergent items as banking, distribution, insurance, transportation, catering, hotels, laundries, hairdressers, etc... both publicly and privately provided. This lately has comprised the largest sector in most Eastern European economies, exceeding 50% in the majority of them.



As one would expect, since it is a natural characteristic of development, most of these countries demonstrate a decline in the percentage of the labour force engaged in agriculture. The numbers vary, with many Eastern European countries seeing double-digit declines.

## **5. Conclusion**

Employment is a political and socio-economic issue which needs to be examined in all its manifestations. Eastern European countries are undergoing a transitional process. The achievement of acceptable levels of manpower utilization will inevitably be slow and in some countries of the region may take many years. Additionally, there is the longer-term problem: the effect of evolving structures of the labour force, attitudes to work and changing social objectives which may affect employment in a fundamental sense.

EU Social Policy Agenda and European Employment Strategy emphasise the importance of ensuring a positive, mutually reinforcing interaction between economic, employment and social policies among the countries. Good employment and social policies are needed to underpin productivity and to improve the capability to change. They will also play an essential role in the transition to the knowledge-based economy. The policy implication of the presented results is that each country can find countries with similar employment structure. Those countries could exchange experience and collectively and much more effectively search the solutions in the process of creation and implementation of employment strategy.

It is necessary to monitor not only the employment in each country but to compare the employment characteristics among individual nations. The more Eastern Europe countries that are approaching EU membership, the more complicated is the analysis of the situation. Multivariate analysis is a very useful tool for monitoring employment, in this case in the region of Eastern Europe.

The principal components analysis represented in this paper is conducted on the basis of just a few employment variables. Many key indicators of the labour market are missing (labour market transitions, occupational and geographical mobility, discouraged workers, multiple jobholders, temporary employment, employment contractors, shift work, work time and leisure time, work at home, time related underemployment, inadequate employment, international movement of labour, etc.). This work could be the ground for further analysis, for example factor analysis, to get even better analysis of the relationship between variables.

## **References**

- Algieri B. (2004) *Trade specialisation patterns: The case of Russia*, BOFIT - Institute for Economies in Transition, Bank of Finland, retrieved October 3, 2006, from: [www.bof.fi/bofit](http://www.bof.fi/bofit)

- Allard G. J., Lindert P. H. (2006) Euro-Productivity and Euro-Job Since the 1960s: Which Institutions Really Mattered?, Working Paper 12460, National Bureau of Economic Research, Cambridge, retrieved September 20, 2006, from: <http://www.nber.org/papers/w12460>
- Bassanini A., Duval R. (2004) Employment Patterns in OECD Countries: Reassessing the Role of Policies and Institutions, Social, Employment and Migration Working Papers No. 35, Directorate for Employment, Labour and Social Affairs Employment, Labour and Social Affairs Committee, OECD Social, Employment and Migration Working Papers, retrieved May 5, 2006, from: <http://www.oecd.org/els>
- European Commission (2001) National Reports on the Demographic Situation in 12 Central European Countries, Cyprus and Malta in 1999, Eurostat Working Papers, 3/2001/E/No 10
- European Commission (2002) National Reports on the Demographic Situation in 12 Central European Countries, Cyprus and Malta in 2000, Eurostat Working Papers
- Forcesse D. (1970) Stages of Social Research, Prentice-Hall, New York, USA
- Goldar B., (2002), Trade liberalization and manufacturing employment: The case of India, Employment Paper 2002/34, Statistical Commission and Statistical Office of the UN Economic Commission for European Communities (UNECE), retrieved September 10, 2006, from: <http://www.unece.org/stats/documents>
- Group of authors (2002), European Union and Eastward Enlargement – Reading, The International Summer School in European Studies, Belgrade, Serbia
- Gueron J. (2005), Women and Young People in the Bulgarian Labour Market, National Statistical Institute of Bulgaria, Statistical Commission and Statistical Office of the UN Economic Commission for European Communities (UNECE), retrieved September 15, 2006, from: <http://www.unece.org/stats/documents>
- Hardarson O. S. (2005) The EU Labour Force Survey and Indicators of Quality in Work, Eurostat, Statistical Commission and Statistical Office of the UN Economic Commission for European Communities (UNECE), retrieved September 15, 2006, from: <http://www.unece.org/stats/documents>
- Judd C.M., Smith E.R., Kidder L.H. (1991), Research Methods in Social Relations, sixth edition, Harcourt Brace Jovanovich College Publishers, The Dryden Press, Orlando, Florida, USA
- Lee E. (2005) Trade Liberalization and Employment, DESA Working Paper No. 5, retrieved September 20, 2006, from: [http://www.un.org/esa/desa/papers/2005/wp5\\_2005.pdf](http://www.un.org/esa/desa/papers/2005/wp5_2005.pdf)
- Manly B.F.J. (2005) Multivariate Statistical Methods, A primer, third edition, Chapman & Hall/CRC
- Morgenstern O. (1965) On the Accuracy of Economic Observations, second edition, Princeton University Press, Princeton, New Jersey, USA
- Schafer J. L. (1997) Analysis of Incomplete Multivariate Data, Chapman & Hall, UK
- Sshmitt J., Baker D. (2006) Old Europe Goes To Work: Rising Employment Rates in the European Union, Center for Economic and Policy Research, Washington DC, retrieved September 10, 2006, from: [www.cepr.net](http://www.cepr.net)
- Trochim W. (1999) The Research Methods Knowledge Base, second edition, Cornell Custom Publishing, Cornell University, Ithaca, New York, USA

UNECE, United Nations Economic Commission for Europe, Trends in Europe and North America, Retrieved October 5, 2006 from: <http://www.unece.org/stats/trends2005>

### **Analiza glavnih komponenti zaposlenosti u Istočnoj Evropi**

**Rezime:** U kontekstu tranzicionih promena u poslednjoj dekadi, u istočnoj Evropi, struktura zaposlenosti je jedno od područja koje se najbrže menja. Koja je najbolja metodologija za upoređivanje različitih zemalja u regionu kada je u pitanju zaposlenost? Multivarijantni statistički metodi za analizu predstavljaju vrlo pouzdan način da se sagleda celina problema. Analiza glavnih komponenti je jedan od najjednostavnijih metoda multivarijantne analize ali može da pruži vrlo korisne informacije o zaposlenosti u istočnoj Evropi na vrlo jednostavan i razumljiv način.

**Ključne reči:** Zaposlenost, Multivarijantna analiza, Analiza glavnih komponenti

**JEL:** J21, C82