Classification of the EU countries labour markets

B. Kába

Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Statistics

Abstract

The objective of the paper is to classify the labour markets of the EU member states on the basis of selected employment and unemployment indicators. In order to achieve the study target, the adequate multivariate exploration procedures have been chosen. In the first part of processing original data, principal component analysis (PCA) was employed. PCA is a multivariate statistical procedure used to reduce the number of observed variables into a smaller number of uncorrelated variables with a minimum loss of information. Moreover, the PCA results can be used for effective ranking of the EU countries according to observed indicators of labour markets. This paper describes the crucial steps in PCA and procedure for ranking mentioned and it reviews how PCA-based statistics are constructed and interpreted. The results of the study have demonstrated the range of application and advantages of the multivariate statistical approaches represented in this paper.

Key words

Classification, employment, unemployment, principal component analysis, cluster

Anotace

Cílem příspěvku je klasifikace trhů práce členských zemí EU na základě vybraných ukazatelů zaměstnanosti a nezaměstnanosti. Pro dosažení uvedeného cíle byly zvoleny odpovídající vícerozměrné postupy průzkumové analýzy dat. V prvé fázi zpracování disponibilních dat byla využita analýza hlavních komponent (PCA). Jedná se o vícerozměrnou statistickou proceduru užívanou k redukci počtu studovaných proměnných na menší počet nekorelovaných proměnných s minimální ztrátou informace. Výsledky PCA mohou být dále využity pro účelnou klasifikaci studovaných objektů (členských zemí EU) podle uvažovaných ukazatelů trhů práce. Daný příspěvek popisuje klíčové etapy PCA a zmiňované klasifikační procedury a shrnuje, jak statistiky založené na PCA jsou konstruovány a interpretovány. Výsledky studie demonstrují okruh použitelnosti i přednosti vícerozměrných statistických postupů uvedených v tomto příspěvku.

Klíčová slova

Klasifikace, zaměstnanost, nezaměstnanost, analýza hlavních komponent, shluk

Introduction

An important prerequisite for sustaining social cohesion and political stability in the European Union is a well-functioning and adaptable labour market. It is therefore understandable that such a great degree of attention is paid to monitoring its development – not only in the individual EU 27 countries, but also through a comparison of them. Eurostat uses a number of indicators for monitoring employment and unemployment. As far as comparative purposes are concerned, various levels of employment and unemployment, structured according to such factors as gender, age, level of education attained by job seekers, the length of unemployment, etc., are the most appropriate ([3], [5]). An isolated analysis of individual indicators does not make it possible to unequivocally evaluate the status of the labour market in individual countries, as these indicators reflect processes that take place simultaneously and with complex levels of interaction. This means that, in order to use the information contained within all the individual indicators in a comprehensive manner, it is necessary to select the corresponding multi-dimensional statistical procedures ([1], [4], [5]).

The objective of this particular paper, which is methodological in nature, is to classify the labour market in the twenty-seven member states of the

EU using selected available employment and unemployment indicators. The realisation of this objective was founded on the use of multi-dimensional Principal Component Analysis (PCA) and Cluster Analysis (CA) statistical methods ([1], [2], [6]). The analysis included an assessment of whether it is possible to find groups of countries amongst the individual EU 27 member states that have a similar labour market situation. Attention was also paid to identifying the indicators that are decisive for monitoring employment and unemployment.

Material and methods

The following were included for all of the twenty-seven EU member states in the analysis:

P1 – employment rate – total

P2 – employment rate, by highest level of education attained – levels 0 – 2 (ISCED 1997)[1]

P3 – employment rate, by highest level of education attained – levels 3 – 4 (ISCED 1997)

P4 – employment rate, by highest level of education attained – levels 5 – 6 (ISCED 1997)

P5 – unemployment rate (ILO definition) – total

P6 – unemployment rate – females

P7 – unemployment rate, by age group – less than 25 years

P8 – unemployment rate, by age group – between 25 and 74 years

P9 – long-term unemployment rate – total.

All the data that were used pertain to 2009 and were obtained from the EU Labour Force Survey (LFS) database. The computations have been performed using the SAS programme package, version 9.1.

The classification and comparison of the labour markets in the member states of the EU 27 was based on the use of principal component analysis (PCA) and cluster analysis (CA) techniques. PCA is a statistical method that makes it possible to compress multi-dimensional statistical data and reduce the number of original variables (which are often highly co-related) through the use of a lower number of uncorrelated variables, or principal components. Each component is constructed as linear combination of the original variables and the weights for each principal component are given by the eigenvectors of the correlation matrix of the initial variables. The principal components are sorted in descending sequence according to the decrease in their variability as measured by eigenvalues that have been analysed using a correlation matrix. As it is generally only the first few principal components (two or three) that contain a significant portion of the variability for the set of objects being analysed, it is possible to limit the analysis to using only these components.

For the purposes of this particular study, the PCA method was supplemented with CA procedures. In order to identify and create clusters, i.e. groups of objects where the objects within one group are mutually similar whilst objects that are not mutually similar are in different groups, procedure K – means clustering (a non-hierarchical classification algorithm) was applied ([6]).

Results

The PCA method is mathematically founded on a certain decomposition of a correlation matrix of available variables, which should contain several correlation coefficients that are more important. In order to assess this characteristic, i.e. factorability, in an exact manner, the KMO (Kaiser-Meyer-Olkin) value is used ([6]). PCA is considered to be an appropriate method to use for a particular set of data if the KMO values for the data are greater than 0.5 (see ([6]). Table 1 provides the average KMO value as well as the individual KMO values for the individual variables.

The presented data show that the prerequisites for the correct application of the PCA method have been met.

[1] Levels 0 – 2: pre-primary, primary and lower secondary education. Levels 3 – 4: upper secondary and post-secondary non-tertiary education. Levels 5 – 6: tertiary education (according to the International Standard Classification of Education, ISCED 1997)

When converting the original variables into principal components, the first step was to calculate the eigenvalues for the correlation matrix. Eigenvalues and the proportion of the total variation explained by each principal components are listed in table 2.

Application of Kaiser–Guttman criterion (see [6]) of retaining only those components whose eigenvalues are greater than 1 for subsequent analysis yielded the first two principal components PC1 and PC2, which accounted for 81,06 % of the total variance. The remaining components were considered less significant. From the eigenvectors obtained in the PCA, the first component can be given as:

PC1= – 0,316P1 – 0,230P2 – 0,330P3 – 0,302P4 + 0,359 P5 + 0,363 P6 + 0,371P7 + 0,359 P8 + 0,345P9. (1)

Similarly, the second principal component can be expressed as:

PC2 = 0,497P1 + 0,435P2 + 0,304P3 + 0,252P4 + 0,377P5 + 0,308P6 + 0,183P7 + 0,369P8 – 0,040P9. (2)

An important output from the analysis of the principal components is the component loadings, which represent the correlation of a component with the individual variables that are being analysed. These aforementioned component loadings – determined on the basis of a varimax rotation (the rotation procedure enhances interpretation of the components without changing their statistical explanatory power – see [6]) – are summarised in Table 3.

From Table 3, it is obvious that the first principal component correlates most strongly with the unemployment rates P5 – P9 in the EU 27 member states. Taking into account that the first principal component explains the greatest proportion of overall variability, these unemployment indicators P5 – P9 can be designated as being the most important for describing the variability of the analysed data. The second component correlates strongly or mid-level strongly with the employment levels (indicators P1 – P4) in the EU 27 member states. These particular indicators are therefore less important from the perspective of describing the variability in the database that is available.

During the next phase of the analysis, the first two principal components, which helped to summarise the multi-dimensional data contained in the indicators for the labour markets in the EU 27 member states, were used to sort the individual countries according to the monitored employment and unemployment indicators. For the purpose of organising the data in this way, the study constructed an indicator PC that aggregated the information provided by all of the indicators considered. This indicator was defined as a linear combination,

$$PC = w_1 \cdot PC1 + w_2 \cdot PC2 , \quad (3)$$

where PC1 and PC2 represented the values from the first and second principal components respectively and wi (i = 1, 2) were the weights assigned on the basis of the PCA results. The specified weights represent the proportion of overall variance, which was explained by the applicable component. It is necessary to note that the equation (3) can easily be generalised even for a greater number of principal components, which could be identified by the aforementioned Kaiser criterion. In this case, the PCA-based indicator would have shape (4):

$$PC = \sum_{i=1}^{k} w_1 \cdot PC(i) . \quad (4)$$

The absolute values of the weights are defined as “explanation ratios of total variance” and their signs (plus or minus) are determined according to the predominant number of pluses or minuses for the component loadings. If more than half of the component loadings of the PC(i) is negative then wi is negative, otherwise it becomes positive. The values for the PC indicator calculated using the above-specified method for the individual EU 27 member states and the applicable sequence for these countries are specified in Table 4 (see columns 1 and 2).

Based on analysis of the relationships (1) – (3), it becomes apparent that the labour market in the majority of the countries which have a negative PC score (there are fourteen in total) is, when compared to the overall EU 27 level, as a rule characterised by an above-average level of employment (overall rate and rate for persons with an education level of 3 – 4 or 5 – 6) and, in particular, a below-average unemployment rate (broken down according to the overall unemployment rate, the unemployment rate for women and the unemployment rate for persons between 25 and 74 years of age) and generally also a lower than average unemployment rate for persons under the age of 25 and a below-average long-term unemployment rate. On the other hand, the labour market in those EU 27 member states with a positive score – as compared to the overall EU 27 level – generally has a below-average level of employment (this in particular applies to persons with an education level of 3 – 4) and an above-average unemployment rate (in particular in the case of the unemployment rate for persons under the age of 25 and the long-term unemployment rate). Table 4 shows that the first ten countries, sequenced according to their PC ranking, include only two of the "new" member states that were accepted into the EU in 2004: Cyprus (in fourth place) and Slovenia (in fifth place). The average value of the PC scores for the "old" EU member states, i.e. the EU 15, is –0.346 as compared to the average PC value for the twelve new EU member states, which is 0.432. For this reason, the next phase of the analysis tested the hypothesis that the value of the PC scores for the new EU member states is based on the same distribution as the PC scores for the EU 15 as opposed to a one-sided alternative. In order to verify this hypothesis, a non-parametric Wilcoxon Rank Sum test was performed, from which the resulting p – value was 0.098. The tested hypothesis was therefore not rejected, i.e. from the perspective of the considered employment and unemployment indicators there was no difference proven as regards the labour markets in the old and new EU member states.

KMO Measure: Overall				0,700				
P1	P2	P3	P4	P5	P6	P7	P8	P9
0,720	0,656	0,642	0,867	0,619	0,774	0,643	0,617	0,916

Source: own calculation according to data of the EU Labour Force Survey

Table 1. Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy.

Components	Eigenvalue	Individual Percent	Cumulative Percent
PC1	6,164	68,49	68,49
PC2	1,131	12,57	81,06
PC3	0,710	7,89	88,95
PC4	0,387	4,30	93,25
PC5	0,310	3,44	96,69
PC6	0,131	1,46	98,15
PC7	0,099	1,10	99,25
PC8	0,066	0,74	99,99
PC9	0,002	0,01	100,00

Source: own calculation according to data of the EU Labour Force Survey

Table 2. Eigenvalues of correlation matrix of indicators P1 – P9.

Variables	PC1	PC2
Total employment rate	– 0,267	0,907
Employment rate, by highest level of education attained – levels 0 – 2	– 0,146	0,721
Employment rate, by highest level of education attained – levels 3 – 4	– 0,426	0,772
Employment rate, by highest level of education attained – levels 5 – 6	– 0,407	0,685
Total unemployment rate	0,943	– 0,260
Unemployment rate – females	0,904	– 0,323
Unemployment rate, by age group – less than 25 years	0,833	– 0,438
Unemployment rate, by age group – between 25 and 74 years	0,937	– 0,266
Long-term unemployment rate – total	0,633	– 0,579

Source: own calculation according to data of the EU Labour Force Survey

Table 3. Component loadings for rotated components.

The results provided in Table 4 also indicate a tendency of certain countries to cluster. These tendencies were identified using a non-hierarchical K – means clustering method, which categorised the EU 27 member states into four clusters on the basis of the employment and unemployment indicators that were analysed. For a more illustrative description of the identified clusters, the calculated cluster averages for the indicators P1 – P9 were compared with the values for these same indicators at the overall EU 27 level. The results of this comparison are provided in Table 5.

From these results it is apparent that the least favourable values for the monitored indicators characterising employment and unemployment in the EU 27 member states in 2009 were recorded in the Baltic States and Spain, which were included in the first cluster. In the case of these countries, high unemployment rates were typical in particular for persons between 25 and 74 years of age. The countries comprising the second cluster (including the Czech Republic) attained relatively favourable results for the individual labour market indicators, which had an employment rate lower than the EU27 average only in the case of persons with a "pre-primary, primary and lower secondary education". The best results for all of the analysed indicators were attained by the countries in the fourth cluster. They differ from the other EU 27 member states on the basis of low values for the individual considered categories of unemployment rates, in particular a very low value for long-term unemployment rate.

All presented results were based on the analysis of overall employment and unemployment indicators of 27 EU member states. It must be noted, however, that 91% of the EU territory is made up of rural areas, i.e. areas where the population density is below 150 inhabitants per square kilometre and 56% of the EU population live in predominantly rural (PR) and significantly rural (SR) areas1 In addition, it must be mentioned that rural areas provide 55% of employment. Therefore it would be useful to complete the set of the analyzed labour market indicators with indicators related only to rural areas. However, national statistic offices of EU countries publish such specialized indicators in a limited extent. With regard to this fact, the following variables were added to the original ones - P1 – P9:

P10 – employment rate in PR or SR respectively in rural areas

P11 – unemployment rate in PR or SR respectively

P12 – long-term unemployment rate in PR or SR respectively

P13 – employment rate in the primary sector in PR regions

In order to classify labour markets of EU member states and their ranking using the indicator (4), all disposable variables P1 – P13 were applied. The achieved results (see columns 3 and 4 of Table 4) were very similar to the above-commented results concentrated in Table 4 (columns 1 and 2). Eleven EU member states were ranked – according to the PC ranking – to the same positions as when applying variables P1 – P9. Seven countries changed their ranking by one position, 6 countries by two positions and 3 countries changed their ranking by 3 positions. In addition, it was possible to state that even on the basis of an extended set of indicators P1 – P13, no difference between the labour markets of 15 old and 12 new EU member states was proven (p – value was 0.116).

The relationship between the ranking values acquired both from variables P1 – P9 and from the extended set of variables P1 – P13 was quantified using the Spearman's correlation coefficient. This coefficient reached the value of 0.982 which signals a very close relation between both sets of ranking values. Hence we can state that both described procedures of EU labour market classification (making use of 9 or 13 employment and

[1] Predominantly Rural region (PR) – more than 50 % of the population of the region is living in rural local units

Significantly Rural region (SR) – 15 % to 50 % of the population of the region is living in rural local units unemployment indicators respectively) are in effect mutually interchangeable and both of the allow – at

Countries	PC scores (1)	PC ranking (2)	PC scores (3)	PC ranking (4)
Belgium (BE)	0,26	15	0,47	17
Bulgaria (BG)	– 0,65	11	– 0,23	14
Czech Republic (CZ)	– 0,55	12	– 0,47	11
Denmark (DK)	– 2,45	2	– 2,36	2
Germany (DE)	– 1,17	6	– 1,04	8
Estonia (EE)	1,70	23	1,21	22
Ireland (IE)	1,06	20	0,75	19
Greece (EL)	1,29	21	1,23	23
Spain (ES)	3,77	27	3,17	27
France (FR)	0,56	18	0,56	18
Italy (IT)	0,78	19	0,97	20
Cyprus (CY)	– 1,86	4	– 1,79	4
Latvia (LV)	2,97	26	2,31	25
Lithuania (LT)	1,74	24	1,20	21
Luxembourg (LU)	– 1,12	8	– 1,02	9
Hungary (HU)	1,65	22	1,79	24
Malta (MT)	– 0,54	13	– 0,30	13
Netherlands (NL)	– 3,28	1	– 2,94	1
Austria (AT)	– 2,34	3	– 2,17	3
Poland (PL)	0,30	16	0,41	16
Portugal (PT)	0,36	17	0,30	15
Romania (RO)	– 0,35	14	– 0,33	12
Slovenia (SI)	– 1,51	5	– 1,33	5
Slovakia (SK)	2,29	25	2,58	26
Finland (FI)	– 0,70	10	– 0,62	10
Sweden (SE)	– 1,14	7	– 1,14	7
United Kingdom (UK)	– 1,07	9	– 1,19	6

Source: own calculation according to data of the EU Labour Force Survey

Table 4. Ranking of EU 27 countries based on PC scores.

Cluster	Countries	P1	P2	P3	P4	P5	P6	P7	P8	P9
1	EE, LT, SK, LV, ES	0,94	0,60	0,93	0,99	1,68	1,50	1,59	1,72	1,49
2	SI, DE, LU, UK, FI, BG, CZ, MT, PT	1,02	0,94	1,02	1,02	0,81	0,82	0,85	0,79	0,81
3	RO, BE, PL, FR, IT, IE, EL, HU	0,93	0,85	0,93	0,98	1,01	1,03	1,20	0,99	1,10
4	NL, DK, AT, CY, SE	1,13	1,19	1,12	1,04	0,62	0,61	0,68	0,58	0,27

Source: own calculation according to data of the EU Labour Force Survey

Table 5. Ratio of cluster averages for the indicators P1 – P9 and values at the overall EU 27 level.

least in an implicit form – consideration of certain specifics of the rural development in EU countries.

Conclusion

The majority of data for the labour market are multi-dimensional in nature. As a result, standard statistical methods are not appropriate for analysing them, as it would not be possible to describe and synthesise the relations between the individual factors and parameters that characterise the labour market. The objective of this particular study was to describe and demonstrate the usability of certain multi-dimensional methods, primarily principal component analysis, as an appropriate analytical tool to use for summarising the information contained in the larger number of indicators used for the labour market. With the help of this technique, it was possible to identify the most important indicators from the given set of EU 27 employment and unemployment indicators and, by subsequently applying a CA method, assessing which of the EU 27 member states are similar from the perspective of the considered labour market indicators. However, it has to be noted, the multivariate statistical methods employed in this study are data exploratory tools, i.e., their fundamental purpose is to describe a structure of relationships within a large data set without explaining why it exists.

The benefits brought by this study consist of the proposal and verification of the usability of a specific measuring method that makes it possible to perform a synthetic evaluation of the information provided by the analysed employment and unemployment indicators. The proposed composite indicator PC – defined by the relations (3) or (4) – integrates large amount of information into easily understood formats and it could be implemented as a fast method in routine analysis. The advantage of this measuring method, which is based on the linear combination of extracted principal components, is the fact that it does not place any demands on the distribution assumptions for the analysed data. The results of this study have proven that the proposed measuring method can be used to attain the effective and unambiguous classification of the labour markets across countries or over time.

Acknowledgements

Pieces of knowledge introduced in this paper resulted from solution of an institutional research intention MSM 6046070906 „Economics of resources of Czech agriculture and their efficient use in frame of multifunctional agri-food systems“.

Corresponding author:
Doc. RNDr.Bohumil Kába, CSc.
Czech University of Life Sciences Prague, Department of Statistics
Kamýcká 129, 165 21 Prague 6 – Suchdol, Czech Republic
Phone: +420 224 382 236
e-mail: kaba@pef.czu.cz

References

[1] Behar, J. Tax Wedges, Unemployment Benefits and Labour Market Outcomes in the New EU Members. *AUCO Czech Economic Review 3*, 2009. ISSN: 1802-4696

[2] Chamberlin, G. Data Reduction and Model Selection Techniques. *Economic & Labour Market Review, Vol. 1, No. 9,* 2007. ISSN: 1751-8334

[3] Heckman, J. J., Singer, B. *Longitudinal Analysis of Labor Market Data*. Cambridge University Press, 2008, Cambridge, U.K. ISBN: 9780521088183

[4] Odehnal, J., Michálek, J. Hodnocení konkurenceschopnosti vybraných regionů Evropské unie. *Ekonomický časopis*, 57, 2009/2. ISSN: 0013-3035

[5] Sedenkova, M., Horák, J. Multivariate and Multicriteria Evaluation of Labour Market Situation. FEEM Working Paper No. 67, 2006. Available at SSRN: http://ssrn.com/abstract=897902

[6] Tabachnick, B. G., Fidell, L. S. *Using Multivariate Statistics.* Pearson Education, Inc., 2007, Boston. ISBN: 0-205-45938-2