

Tatarczak, Anna; Boichuk, Oleksandra

Working Paper

The Use of Multivariate Techniques for Youth Unemployment Analysis in Poland

Institute of Economic Research Working Papers, No. 130/2017

Provided in Cooperation with:

Institute of Economic Research (IER), Toruń (Poland)

Suggested Citation: Tatarczak, Anna; Boichuk, Oleksandra (2017) : The Use of Multivariate Techniques for Youth Unemployment Analysis in Poland, Institute of Economic Research Working Papers, No. 130/2017, Institute of Economic Research (IER), Toruń

This Version is available at:

<https://hdl.handle.net/10419/219952>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/3.0/>



Institute of Economic Research Working Papers

No. 130/2017

The Use of Multivariate Techniques for Youth Unemployment Analysis in Poland

Anna Tatarczak, Oleksandra Boichuk

Article prepared and submitted for:

9th International Conference on Applied Economics Contemporary Issues in Economy, Institute of Economic Research, Polish Economic Society Branch in Toruń, Faculty of Economic Sciences and Management, Nicolaus Copernicus University, Toruń, Poland, 22-23 June 2017

Toruń, Poland 2017

© Copyright: Creative Commons Attribution 3.0 License

Anna Tatarczak, Oleksandra Boichuk
E-mail address: anna.tatarczak@poczta.umcs.lublin.pl olaboichuk@gmail.com
Maria Curie Skłodowska University, Lublin, Poland

The Use of Multivariate Techniques for Youth Unemployment Analysis in Poland

JEL Classification: C38; J42

Keywords: *labour market; unemployment; young people; cluster analysis.*

Abstract

Research background: The labour market situation is considered to be the most widely discussed part of economic development. However, it should be noted that the unemployment situation of young people (aged 15 -24 years) in Poland in general terms seems to be problematic. Overall, the unemployment rate among young people in Poland is significantly higher than the overall unemployment rate in the EU. Moreover, the situation varies greatly across the regions.

Purpose of the article: Using multivariate techniques as a theoretical framework, the main goal of the paper is to identify groups of Polish regions that share similar patterns regarding unemployment among young people.

Methodology/methods: The initial calculation is based on the concept of the taxonomic measure developed by Hellwig. The final method used to create clusters of objects (across 16 voivodeships of Poland) is cluster analysis. A segmentation of the voivodeships is observed for the years 2005 and 2014, based on selected indicators to determine the labour market situation.

Findings: Through the exploration of the advantages of multivariate methods, the nature of youth unemployment is revealed in more precise detail. Indeed, dendrogram analysis divided the voivodeships into five groups, which are characterized by similar features associated with the labour market. It was found that the groups which emerged in 2005 have a different composition of regions than in 2014; this difference seems to be connected to the economic crisis.

Introduction

Unemployment is an economic indicator that refers to the portion of people who are actively looking for a job and are unable to find work. The youth unemployment ratio is calculated as the ratio of youth unemployment to the adult unemployment rate. Youth unemployment is often estimated separately because the rate has historically been higher than that for older age groups.

Moreover, this phenomenon, which has recently attracted increasing attention, has negatively influenced countries' labour markets at a time of economic crisis. In fact, there is a wide range of both theoretical and empirical literature devoted to the connections between the global economic crisis and labour markets, e.g. Rose & Spiegel (2011), Madianos *et al.*, (2014), Boeri & Jimeno (2016). However, the rather narrow range of literature regarding youth unemployment has focused on the dispersion of unemployment across Poland's regions during recessions.

In this context, the main aim of this paper is to show the local diversity of the situation on the labour market among the voivodeships of Poland in 2005 (before the economic crisis) and 2014 (under the effects of the crisis). In order to contribute to the achievement of this goal, our ambition is to detect the presence of homogeneity among different regions based on a multivariate statistical method, namely cluster analysis and Hellwig's method.

The paper consists of five main sections. Section 2 focuses on literature data and methodology. In Section 3, the main facts about the specific unemployment situation in Poland are presented, paying particular attention to youth unemployment. The results of the analysis are presented and commented on in Section 4. The main conclusions are presented in Section 5.

Research Methodology

Cluster analysis is one of multidimensional methods that allows observations to be classified into groups. Cluster analysis techniques include several different algorithms, which can be broadly divided into two methods: hierarchical and non-hierarchical. Dendrograms are often used to aid visualization in the form of a tree showing the linkages between observations. In order to group the voivodeships into clusters, a hierarchical Ward's algorithm based on a squared Euclidean distance, has been chosen. This method is the most highly recommended one due to the efficiency criterion of presenting the actual data structure.

Next, in the current research, the concept of taxonomic measure proposed by Zdzisław Hellwig (Hellwig, 1968) is used. This is a commonly applied method in spatial economic research which allows the researcher to produce a synthetic measure d_i , which takes into account the impact of many indicators on socio-economic development. The main advantages of Hellwig's method are its methodological simplicity and the flexibility of its application.

The aim of this paper is to use cluster analysis and Hellwig’s synthetic measure to create a new perspective for discussing differences and similarities of youth unemployment in the regions of Poland. In addition, a similar method can be applied in the context of other countries to visualize which of the regions are more similar.

Analysed data are obtained from the databases of the Central Statistical Office of Poland and Eurostat. Indicators which determined the situation of young people in Poland for two different years – 2005 and 2014 were selected. Statistical calculations were performed using SPSS statistical software, version 22.

The starting point of the analysis were indicator data for the labour market and wages. The initial set of potential diagnostic variables includes six indicators of the structure of unemployment by age, education, work experience and other measures of unemployment. The variables are presented in Table 1. All of the variables were standardized with classic standardization formulae.

Table 1. Diagnostics variables for the purpose of describing youth unemployment.

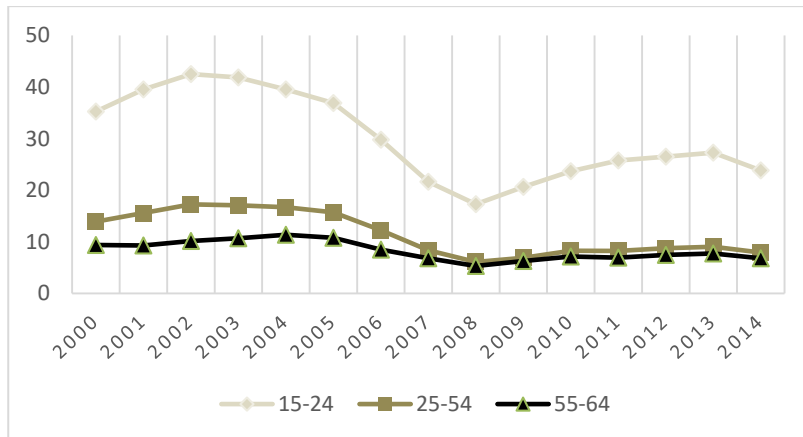
Variable	Description of diagnostic variables
X ₁	participation of unemployed persons in age group of 15-24 years in the total number of unemployed
X ₂	participation of unemployed persons without internship in the total number of unemployed
X ₃	participation of unemployed persons with higher education in the total number of unemployed
X ₄	job vacancy rate
X ₅	number of unemployed persons attributable to one offer
X ₆	the monthly average gross salary in relation to the national average

Source: own calculations based on Eurostat (2017).

Statement of the Problem and Research Tasks

A main driver of Polish unemployment divergence is youth unemployment. Figure 1 plots the annual change (horizontal axis) in the unemployment rate (vertical axis) over the 2000 – 2014 period for three age groups. As Figure 3 illustrates, the highest level of unemployment is experienced by young people. Youth unemployment rates at a national level have experienced considerable turbulence. The unemployment rates within other age groups are much more stable. Since 2006, the unemployment rate among young people has declined by about 10 %, and it is still twice as high compared to those of other groups. As shown in Figure 3, the recession affected youth unemployment rates earlier than it affected other age groups; this is indeed the group most affected by the economic crisis.

Figure 1. Unemployment rate by age group in Poland between 2000 and 2014.



Source: own calculations based on OECD Statistics.

Table 2 shows the changes among the investigated factors in 2005 and 2014, these are the main unemployment facts that influenced our analysis. We observe a significant decrease in almost all chosen variables, namely X_1, X_2, X_4, X_5 .

Table 2. Specification the investigated factors

Characteristics	2005	2014	Growth Rate (in %)
age: 15-24	774 575	347 325	-55.16
education: higher education	152 400	225 441	47.93
unemployed persons without internship	656 600	315 400	-51.96
job vacancy rate (%)	66	50	-24.24
number of unemployed persons attributable to one offer	158	46	-70.89
the monthly average gross salary (PLN)	4439.76	6516.26	46.77

Source: own calculations based on Eurostat (2017).

The rise in youth unemployment in Poland and increasing levels of European unemployment dispersion across countries and regions are two facts that give us the motivation required to analyse the situation. In order to pursue these problems, the paper proceeds to develop an analytic framework for identifying groups of Polish regions that share similar patterns regarding unemployment among young people.

The Result of Research

In order to compare the labour market situation of young people between the Polish regions in 2005 and in 2014, we implemented a two-step procedure. Firstly, we use the concept of the taxonomic measure to arrange the items being studied in a linear manner from the best to the worst. In addition, we classified the Polish regions according to their labour market performance, using cluster analysis methods. Similar results were obtained in both classifications for the years 2005 and 2014.

Table 3. Polish regions arranged according to Hellwig's measure in the years 2005 and 2014.

Voivodeship	d_i in 2005	Voivodeship	d_i in 2014
Masovian	0.632	Lower Silesian	0.656
Lower Silesian	0.629	Masovian	0.62
Pomeranian	0.554	Pomeranian	0.51
West Pomeranian	0.485	West Pomeranian	0.503
Opole	0.479	Silesian	0.465
Lubusz	0.471	Lubusz	0.461
Łódź	0.444	Łódź	0.443
Silesian	0.415	Opole	0.436
Greater Poland	0.347	Kuyavian-Pomeranian	0.427
Kuyavian-Pomeranian	0.326	Greater Poland	0.414
Warmian-Masurian	0.282	Warmian-Masurian	0.329
Lesser Poland	0.258	Lesser Poland	0.216
Podlaskie	0.155	Podlaskie	0.173
Podkarpackie	0.149	Podkarpackie	0.09
Lublin	0.106	Świętokrzyskie	0.076
Świętokrzyskie	0.025	Lublin	0.058
Arithmetic average	0.360	Arithmetic average	0.367
Standard deviation	0.186	Standard deviation	0.190

Source: own calculations based on Eurostat (2017).

According to the results presented in Table 3, there is a high diversity for the different provinces with respect to the synthetic measure. For a group of provinces characterized relative to each other, the best situation on the labour market in 2014 occurred in the Lower Silesian and Masovian regions. The lowest synthetic measure value was obtained for Lublin and Świętokrzyskie voivodeships. Both of these voivodeships stand out from the others in terms of the values of the measure. This means that in these voivodeships we have to deal with the worst situation with regard to youth participation in the labour market from the viewpoint of the adopted diagnostic features. Positive changes in the labour market were observed at the Silesian voivodeship where the dynamic of the changes reached the highest level (from 8th position to 5th). It may be stated with certainty that the situation in the labour market has worsened in the Opole voivodeship (falling from 5th to 8th position). Also, a decrease by one position was observed for the Masovian, Greater Poland and Lublin voivodeships. The process of arranging the regions showed that seven of them (Pomeranian, West Pomeranian, Lubusz, Łódź, Warmian-Masurian, Podlaskie and Lesser Poland) occupied the same position with regard to the value of d_i .

In order to create groups of voivodeships, which are similar in terms of the structure of the labour market, a hierarchical cluster analysis has been used. The composition of clusters in the year 2005, is shown in Table 4, while in

the year 2014 in Table 5. A detailed description of the clusters in 2005 and 2014 is presented in Table 6 and Table 7. The territorial distribution of the clusters produced by the analysis is presented in Figure 2.

Table 4. Distribution of Polish regions by clusters in 2005.

2005				
1	2	3	4	5
Łódź	Lower Silesian	Opole	Podkarpackie	Świętokrzyskie
Greater Poland	Silesian	West Pomeranian	Warmian-Masurian	
Kuyavian-Pomeranian	Pomeranian	Lublin	Podlaskie	
Lubusz	Lesser Poland			
	Masovian			

Source: own calculations based on Eurostat (2017).

Table 5. Distribution of Polish regions by clusters in 2014.

2014				
1	2	3	4	5
Opole	Silesian	Podlaskie	Warmian-Masurian	Łódź
Kuyavian-Pomeranian	Greater Poland	Podkarpackie	Lublin	Lesser Poland
Pomeranian	Lower Silesian	Świętokrzyskie		Masovian
West Pomeranian	Lubusz			

Source: own calculations based on Eurostat (2017).

Table 6. Cluster description in 2005.

Year	2005				
Cluster	1	2	3	4	5
participation of unemployed persons in age of 15-24 years, %	22.5	22.2	19.67	24.33	24
participation of unemployed persons without internship, %	20.25	23.2	21	26.67	29
participation of unemployed persons with higher education, %	4.5	17	8.67	6.33	8
indicator of the new working places using, %	66	73.2	62.33	51.33	57
number of unemployed persons attributable to one offer	159.25	116.2	251.33	408.67	1070
the monthly average gross salary (PLN)	3397.16	4424.18	3325.02	3162.91	3208.39
Number of voivodships in cluster	4	5	3	3	1

Source: own calculations based on Eurostat (2017).

Table 7. Cluster description in 2014.

Year	2014				
Cluster	1	2	3	4	5
participation of unemployed persons in age of 15-24 years, %	16	15	18.53	18	15
participation of unemployed persons without internship, %	15.5	15	21.67	21	17
participation of unemployed persons with higher education, %	9	11.25	14.33	12	12.67
indicator of the new working places using, %	55	56	29.33	35	52
number of unemployed persons attributable to one offer	43	27	103.67	78	61
the monthly average gross salary (PLN)	5186.82	5513.26	4949.3	4740.23	6002.67
Number of voivodships in cluster	4	4	3	2	3

Source: own calculations based on Eurostat (2017).

Cluster 1 (2005)

This cluster is characterized by the lowest level of unemployed persons without internship and the lowest percentage of unemployed persons with higher education.

Cluster 2 (2005)

This cluster is characterized by high levels of unemployment among the highly educated, although there are a lot of vacancies which are being regularly taken and the average salary in this cluster is the highest. Since the cluster includes five of the 16 voivodeships, the situation regarding unemployment among people with higher education and a lack of jobs for them was important for Poland.

Cluster 3 (2005)

The lowest indicator of unemployment among young people is a characteristic of this group. The rate of unemployment among people with higher education is almost twice as low as it is for cluster 2, but it stands at the highest level compared to clusters 1, 4, 5.

Cluster 4 (2005)

Being the cluster which is characterized by the highest rate of unemployment in the age group of 15-24, it has the lowest level of job vacancy rate. The voivodeships in this cluster include the lowest levels of monthly average gross salary.

Cluster 5 (2005)

The single voivodeship in this cluster is characterized by the greatest number of unemployed persons attributable to one offer as well as the highest level of unemployment among people without experience.

Cluster 1 (2014)

It includes four voivodeships which are characterized by the lowest percentage of unemployed people with higher education, as well as by the lowest percentage of unemployed persons without internship.

Cluster 2 (2014)

This group consists of four voivodeships. Compared with the first cluster it has lower indicators of participation of unemployed people in the age group of 15-24 years. This cluster is characterized by the lowest number of unemployed persons attributed to one offer and by participation of unemployed persons without internship.

Cluster 3 (2014)

Among all of the clusters this one is next to first in term of the participation of unemployed persons in age of 15-24 years, without internship and with high education.

Cluster 4 (2014)

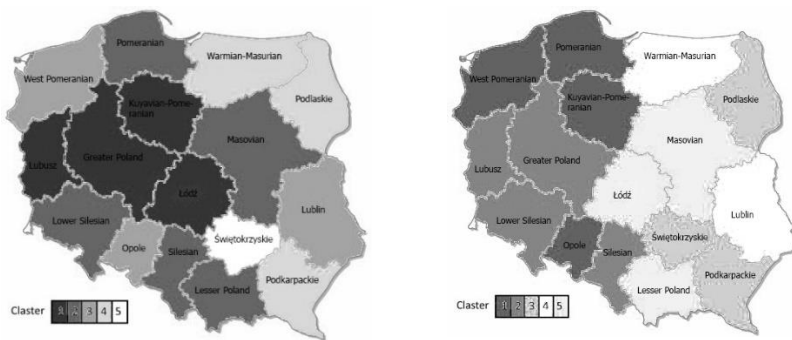
A defining characteristic of the cluster is a lower monthly average gross salary than any other cluster. This cluster shares similar problems with cluster 3 when it comes to youth unemployment. It is in second place with a higher level of unemployed persons without internship. Even though these indicators in the cluster are slightly better than they are in the third cluster, this does not affect the size of the average salary.

Cluster 5 (2014)

This cluster shares a similar situation with cluster 2 when it comes to unemployed persons in the age bracket of 15-24 years. It has the highest monthly average gross salary despite the average values of other indicators.

To sum up, cluster analysis and the use of Hellwig's method have allowed us to examine the Polish unemployment situation with regard to groups of regions according to their youth labour market performances. According to the selected indicators of the market, it may be observed that advanced regions proved to be more homogenous. The clusters with highest numbers of voivodships are those with a good or at least a moderately good situation, in the sense of the synthetic variable, although their number changed over time.

Figure 2. Groups of the voivodships distinguished on the basis of cluster analysis in 2005 (on the left) and in 2014 (on the right).



Source: own calculations based on Eurostat (2017).

Conclusions

The results of the analysis show that the methods used are suitable for inter-regional comparisons on the basis of the labour market. This paper presents data mining methodology, in particular, cluster analysis and Hellwig's methods which allowed us to divide the 16 voivodeships into five groups, which are characterized by similar features associated with the labour market. The measures provide a way to grasp the changes between the years 2005 and 2014 in the labour market situation in the Polish regions. Based on this analysis we may conclude that the labour market in Poland is considerably regionally structured. In the case of Poland, cluster analysis shows the division of the country into five groups which are homogenous in terms of unemployment, which is described by multiple characteristics at the same time.

This paper also has some important policy implications. Namely, the results could help decision makers to identify regional similarities/dissimilarities in the Polish labour market. The lack of stability in the labour market in the form of visible differences occurring in clusters requires special attention from the ruling elites who ought to take joint steps aimed at reducing the number of the young unemployed people. The state should actively contribute to reducing disparities between the regions. The main justification for the need for such a policy is to achieve equality of opportunity of development between the regions. The high level of regional diversity discriminates against people living in regions with high unemployment rates among young people and low per capita incomes. Furthermore, the findings of this paper may help to generate new ideas concerning which factors should be prioritized regarding equal opportunities for all young people in education and in the labour markets of the different regions of Poland.

References

- Boeri, T., & Jimeno, J. F. (2016). Learning from the Great Divergence in unemployment in Europe during the crisis. *Labour Economics*, 41, 32-46.
DOI: <http://dx.doi.org/10.1016/j.labeco.2016.05.022>
- Hellwig, Z. (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd statystyczny*, 4(1968), 307-326.
- Madianos, M. G., Alexiou, T., Patelakis, A., & Economou, M. (2014). Suicide, unemployment and other socioeconomic factors: evidence from the economic crisis in Greece. *The European Journal of Psychiatry*, 28(1), 39-49.
DOI: <http://dx.doi.org/10.4321/S0213-61632014000100004>

Rose, A. K., & Spiegel, M. M. (2011). Cross-country causes and consequences of the crisis: An update. *European Economic Review*, 55(3), 309-324.
DOI: <http://dx.doi.org/10.1016/j.eurocorev.2010.12.006>.