Regional Unemployment Disparities in Romania

Grecu Mihaela
Bucharest University of Economic Studies, Department of Statistics and Econometrics
mihaela_grecu24@yahoo.com

Abstract

The unemployment phenomenon and the factors that contribute to its evolution have always been in the attention of the analysts having the main purpose to study the size, the structure and the dynamic of the unemployment.

The aim of this study is to make a segmentation of the Romanian regions by taking into account the unemployment rate, the number of emigrants, the number of immigrants and the population for 42 regions for 2016 year. We decided to include the unemployment rate and the migration rate because these second index has strong implications on the first index.

Key words: unemployment, migration, clustering
J.E.L. classification: J16, J21, J61, J62, J64

1. Introduction

According to OECD, the unemployment rate is defined as being the ratio of the number of people without work divided by the total number of working people (https://stats.oecd.org/glossary/detail.asp?ID=3094).
From a historical point of view, in a family, men have been the ones who have power in terms of income. Compared to men, women had less or no opportunity to work. Over time, this strong gap begun to reduce gradually, moving from male breadwinner model to the dual breadwinner model (Gush *et al*, 2015, p. 705). Therefore, the role of the man in the household has not changed much in comparison with the women's role, it has become more complex, women have to combine household chores with the employee status.

Young men are more likely to enter the labor market when the unemployment rate is high. This can be seen as favoring the "added worker" effect. At the opposite end, middle-aged and elderly men, high levels of unemployment are associated with low participation, which implies that the "discouraged worker" effect predominates (Kakinka *et al*, 2011, p. 7). Even though, the data from this study suggest that in Romania, the unemployment rate is higher for men.

2. Data and methods

The unemployment rate is gaining more and more attention in the scientific field. In order to see what are the main causes that lead to a high level of unemployment it is important to analyze its evolution of the unemployment rate divided by gender, by region, by age, by attainment level or parallel with another economic phenomena.
In this study, we will focus on analyzing the unemployment rate by gender and by region.

In the next figure, we have the evolution of the unemployment rate by gender. We can see that the evolution of these two rates is quite similar. For the first years included in the analysis, the unemployment rate for males is smaller than the unemployment rate for females. This happened until 1998 when the unemployment rate for females matched the unemployment rate for males. The biggest gap was in 1993 with 4.8 percentage points between the males and females rate (8.1% for males compared to 12.9% for females). After 1998, the unemployment rate for females was smaller than the unemployment rate for males in each year, except 2008 year for which we had equal rates

(4.4%). The biggest gap for this second case was in 2004 with a difference of 1.4 percentage points (5.6% for females compared to 7.0% for males).

Figure 1. Unemployment rate by gender in Romania for 1991 – 2016 period
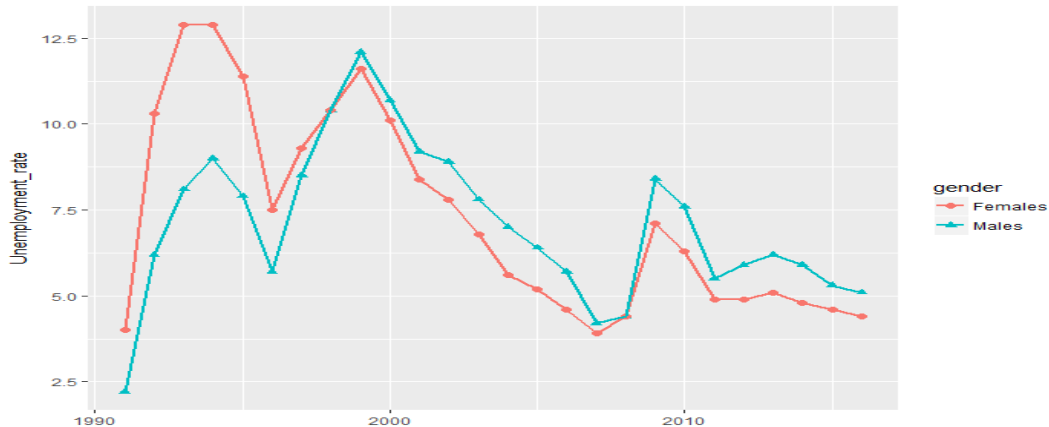
The next analysis is made for males and for females by using the K-Means algorithm. This algorithm is maybe the most used clustering method. It uses simple principles and can be very easily understood even though you do not have statistical knowledge. It is very flexible and it can be adapted to many real-world cases. The statistical tool used for this study is R with k-means library. We collected the data from the National Institute of Statistics site (www.insse.ro).

3. Results and discussion

Before applying the K-Means algorithm, we looked at the data. We noticed that the unemployment rate for women has the smallest value compared to the unemployment rate for males and total level, even though the female population is bigger than the male population. This means that females find a job easier than males.

Figure 2. Male summary

Region	unempl_rate_m	immigrants_m	emigrants_m	population_m
Length:42	Min. : 1.000	Min. : 20.00	Min. : 48.0	Min. :101710
Class :character	1st Qu.: 3.650	1st Qu.: 47.25	1st Qu.: 124.2	1st Qu.:154655
Mode :character	Median : 5.850	Median : 67.50	Median : 183.0	Median :199542
	Mean : 6.117	Mean : 383.71	Mean : 238.3	Mean :229757
	3rd Qu.: 7.875	3rd Qu.: 205.00	3rd Qu.: 271.5	3rd Qu.:275292
	Max. :14.800	Max. :3535.00	Max. :1623.0	Max. :856053

Figure 2. Female summary

Region	unempl_rate_f	immigrants_f	emigrants_f	population_f
Length:42	Min. :1.100	Min. : 20.00	Min. : 58.0	Min. :101487
Class :character	1st Qu.:3.600	1st Qu.: 47.25	1st Qu.: 152.0	1st Qu.:157874
Mode :character	Median :4.700	Median : 67.50	Median : 224.0	Median :207347
	Mean :5.014	Mean : 279.69	Mean : 304.8	Mean :240726
	3rd Qu.:6.500	3rd Qu.: 195.50	3rd Qu.: 333.2	3rd Qu.:289098
	Max. :8.900	Max. :2378.00	Max. :2011.0	Max. :988259

This happens at the aggregate level and not per region. We have regions where the unemployment rate for females is bigger than the unemployment rate for males, as Cluj where the unemployment rate for females is 2.5% while for males is 2.0%, as Constanta with an unemployment rate for females of 4.4% compared to 2.9% for males, as Bucuresti with a rate of 1.7% for females and 1.4% for males.

We have regions where the gap between the two-unemployment rates is quite big. In Vaslui the difference between the unemployment rate for males and for females is 5.9 perceptual points (14.8% the unemployment rate for males and 8.9% the unemployment rate for females). Teleorman is another region with the same pattern, with an unemployment rate for males of 13.2% and 8.2% for females.

In order to establish the correct number of clusters we made some simulations and we decided to use six clusters. A smaller number of clusters would lead to a higher number of regions per cluster, which were not likely to be very useful.

The k-means algorithm assigns each of the n examples to one of the k clusters, where k is a number that has been determined ahead of time. The goal is to minimize the differences within each cluster and maximize the differences between the clusters (Lantz, 2013, p. 289).

Firstly, we perform the clustering for the male database. After running the K-Means algorithm, we obtained 6 clusters of different size. The first cluster contains 7 regions, the second cluster contains 7 regions, the third cluster contains 12 regions and the fourth cluster 5 regions, the fifth cluster contains 10 regions and the last cluster contains 1 region. The smallest cluster has 2.4% of the regions, while the largest cluster has 40% of the regions.

Figure 3. Male clusters

```
> male_clusters$size
[1]   7   7  12   5  10   1
```

Source: *www.insse.ro*

Although the large gap between the smallest cluster and the largest cluster is concerning, we cannot draw a conclusion without examining the data more carefully. We will find more details after looking at the cluster's homogeneity.
For a more in-depth look, we will examine the centroids (coordinates of the clusters).

Figure 4. Male centroids

	unempl_rate_m	immigrants_m	emigrants_m	population_m
1	4.700000	708.4286	367.4286	340700.1
2	6.042857	285.5714	264.0000	272271.4
3	6.450000	48.5000	129.6667	155355.4
4	7.520000	27.8000	84.8000	113083.0
5	6.530000	544.0000	198.4000	207327.7
6	1.400000	2997.0000	1623.0000	856053.0

Source: *www.insse.ro*

The rows of the output refer to the number of clusters and the numbers across each row indicate the mean of each cluster for each variable of interest. For the unemployment rate, we can see that the last cluster has the smallest value. This last cluster has also the highest number of population even though it contains only one region. The number of emigrants and immigrants is the highest compared to the rest of the clusters. This leads us to conclude that the region included in this cluster is the most developed region compared to rest of them. The region included in this cluster is Bucuresti, the capital of Romania, which is by far the most developed area in Romania.

At the opposite side, we have the fourth cluster with a mean for the unemployment rate equal to 7.52%. The regions included in this cluster have the smallest mean value for the number of emigrants, immigrants and for population. If we look at the difference between the mean number of emigrants and the mean number of immigrants (570000), we can say that this cluster maybe contains undeveloped regions. The regions included in this cluster are Salaj, Covasna, Tulcea, Ialomita and Mehedinti.

We also have a cluster (cluster number 3) for which we have a big unemployment rate (6.45%), a small number of immigrants and a big number of emigrants (compared to the number of immigrants). This means that the regions included in this cluster contains undeveloped regions as well. The regions included in this cluster are Bistrita-Nasaud, Satu Mare, Alba, Harghita, Braila, Vrancea, Calarasi, Giurgiu, Teleorman, Gorj, Valcea si Caras-Severin.

Clusters 6 and 1 contain regions with a small mean for unemployment rate of 1.4% and 4.7%. The different thing for these clusters is the number of immigrants, which is very big (almost double) compared to the number of emigrants. This lead us to conclude that the regions included in these two clusters are very popular for people who are looking for a better job, for better living conditions. The regions included in the cluster 1 are Cluj, Iasi, Suceava, Constanta, Prahova, Dolj and Timis.

The rest of the clusters (5 and 2) contains the following regions. Cluster 5 contains Maramures, Sibiu, Botosani, Neamt, Vaslui, Buzau, Ilfov, Olt, Arad and Hunedoara regions, while cluster 2 contains Bihor, Brasov, Mures, Bacau, Galati, Arges and Dambovita regions.

In the next part, we will perform the clustering for females. As per male database, for females we obtained 6 clusters of different size as well.

The first female cluster contains 8 regions, the second cluster contains 9 regions, the third cluster contains 6 regions, the fourth cluster contains 1 region, the fifth contains 10 regions and the last cluster contains 8 regions. Excluding the cluster number 4, the rest of the clusters have almost the same number of regions – between 14% and 24%. As per male database, we have here as well one cluster with only one region. The next step in our study is the centroids analysis.

Figure 5. Female clusters


```
> female_clusters$size
[1]   8   6   1   9   8  10
```

Source: *www.insse.ro*

The third cluster, which contains only one region, is the most developed one, as per male clustering. It has a very small unemployment rate (1.7%), the biggest cluster population and the biggest flow in terms of immigrants and emigrants. The immigrants are coming mostly from the rest of the Romania regions, while the emigrants leave Bucuresti to migrate outside the country, to more developed countries. This cluster and the last cluster are the only ones that have the number of immigrants bigger than the number of emigrants. The difference between these two clusters is that the sixth cluster has a big unemployment rate (5.14%). The regions included in the sixth cluster are Maramures, Sibiu, Botosani, Neamt, Vaslui, Buzau, Ilfov, Olt, Arad and Hunedoara.

Figure 6. Female centroids

	unempl_rate_f	immigrants_f	emigrants_f	population_f
1	4.95	34.75000	136.7500	125541.1
2	4.10	498.66667	503.0000	362701.8
3	1.70	2116.00000	2011.0000	988259.0
4	5.70	51.33333	159.5556	166859.1
5	5.25	266.75000	329.1250	287043.5
6	5.14	376.50000	260.8000	214362.4

Source: *www.insse.ro*

At the opposite side, we have the fourth cluster, which has the highest unemployment rate (5.7%), a small population and immigrants and a big number of emigrants. The regions included in this cluster are Satu Mare, Alba, Harghita, Braila, Vrancea, Calarasi, Teleorman, Gorj and Valcea. Another cluster with big gap between the number of immigrants and the number of emigrants is the first cluster. The unemployment rate for this cluster is 4.95%, the population is the smallest compared to the rest of the clusters and the number of emigrants is four times higher than the number of immigrants. The regions included in this cluster are Bistrita-Nasaud, Salaj, Covasna, Tulcea, Giurgiu, Ialomita, Mehedinti and Caras-Severin.

The fifth cluster has a big unemployment rate, while the difference between emigrants and immigrants is not as big as the above-mentioned clusters. The regions included in this cluster are Bihor, Brasov, Mures, Bacau, Suceava, Galati, Arges and Dambovita. The second cluster has almost the same path as the fifth cluster. The regions included in the second cluster are Cluj, Iasi, Constanta, Prahova, Dolj and Timis.

4. Conclusions

After making the clustering for females and for males, we obtained 6 clusters with different number of regions. Bucuresti is by far the most developed region, being the only region that makes up a cluster. This cluster is the only cluster that has the unemployment rate for males smaller than the unemployment rate for females. We also have clusters where the difference is slightly bigger. The cluster 5 for females has the unemployment rate 5.14% while the cluster 3 for males has the unemployment rate 6.53%. Here we have same structure in both clusters.

The rest of the clusters do not have the same structure in terms of regions included. The cluster 3 for females and 5 for males include almost the same regions except Suceava region, which is included in addition in the male cluster. The unemployment rate for males (4.7%) is higher than the unemployment rate for females (4.1%).

5. References

- Gush K., Scott J., Laurie H., 2015, Households' responses to spousal job loss: 'all change' or 'carry on as usual'?, *Work, employment and society*, Vol. 29(5) 703-719
- Kakinka M., Miyamoto H., 2011, Unemployment and Labor Force Participation in Japan*, IUJ Research Institute, International University of Japan, Economic and Management Series*
- Lantz B., 2013, *Machine Learning with R*, Second Edition, Published by Packt Publishing Ltd.
- www.insse.ro
- https://stats.oecd.org/glossary/detail.asp?ID=3094