# Unemployment Evolution in Romania

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

## Abstract

*The labor market occupies the main stage in the specialty literature research. One of the most important phenomenon to understand is the unemployment, having implication in many areas as sociology, demography, economics and others.*

*The purpose of this study is to find a relationship between the unemployment rate and other economic indices: number of emigrants, number of immigrants, import, life expectancy, GDP and available population for Romania for the 1991 – 2016 period.*

**Key words:** unemployment, multiple regression, regression with interaction
**J.E.L. classification:** J11, J60, J61

## 1. Introduction

Strong and widely accepted evidence shows that the natural rate of unemployment varies over time with substantial amplitude. The frictions in the labor market that account for positive normal levels of unemployment are not simple and mechanical (Hall, 2003).
Within a dynamic labor market, overlapping individuals and jobs is a complex process, precisely for this reason there will always be jobseekers as well as available jobs waiting for individuals to be hired. Taking into account economic efficiency, we can say that there is an optimal level of unemployment and that the unemployment rate will always be positive.

Unemployment occurs when a worker departs from a job and spends time finding a new job. In addition, unemployment arises when a person looks for a new job after a period out of the labor force (Hall, 2005, p. 3).

In order to assess the evolution of this economic phenomenon, it is important to take into account all the repercussions and effects that unemployment may have (both in the short and long term). A high level of unemployment can be the result of unemployment and unemployment flows and / or a high level of unemployment (Blanchard, 2005, p. 6).

## 2. Data and methods

The data included in this analysis contains indices as unemployment rate, the number of immigrants, the number of emigrants, the available population, the GDP and life expectancy. We collected the data from the National Institute of Statistics site for 1991 - 2016 period and the statistical tool used is R (www.insse.ro ).

Before performing the statistical analysis, we included three more indices in our database. They have actually the unemployment rate value as: for the current year we will have the unemployment rate for the previous year, for two years ago and for three years ago – we called these indices "unemployment rate t-1", "unemployment rate t-2" and "unemployment rate t-3". Doing this we will bring more accuracy to do prevision of the unemployment rate. We had to remove the first three years in order to have complete information and the years included in the analysis are from 1998 until 2015 (for 2016 we did not have data for GDP).

In the below picture we have the correlation matrix. If we look at the unemployment rate we can see that it is highly correlated with the import (-0.82) and with GDP (0.78). We also have medium correlation between the unemployment rate and the available population (0.65), life expectancy (-0.67) and GDP (0.45). We also have low correlation between unemployment rate and the number of emigrants (0.17).

*Figure 1. Correlation matrix*

| | unempl_rate | emigrants | immigrants | import | available_population | GDP | life_expectancy | unempl_rate_t1 | unempl_rate_t2 | unempl_rate_t3 |
|---|---|---|---|---|---|---|---|---|---|---|
| unempl_rate | 1.0000000 | 0.1795983 | -0.3067518 | -0.8287797 | 0.6459943 | -0.7808544 | -0.6716665 | 0.7873320 | 0.5377599 | 0.4347001 |
| emigrants | 0.1795983 | 1.0000000 | 0.0407417 | -0.2758063 | 0.6226018 | -0.2710394 | -0.3160269 | 0.2418358 | 0.3094522 | 0.1836177 |
| immigrants | -0.3067518 | 0.0407417 | 1.0000000 | 0.6385186 | -0.3370474 | 0.6541519 | 0.7089370 | -0.3781827 | -0.4326616 | -0.4665037 |
| import | -0.8287797 | -0.2758063 | 0.6385186 | 1.0000000 | -0.7384423 | 0.9914482 | 0.9573293 | -0.8034333 | -0.7721717 | -0.7490915 |
| available_population | 0.6459943 | 0.6226018 | -0.3370474 | -0.7384423 | 1.0000000 | -0.7018409 | -0.7662491 | 0.6117616 | 0.5473647 | 0.4355353 |
| GDP | -0.7808544 | -0.2710394 | 0.6541519 | 0.9914482 | -0.7018409 | 1.0000000 | 0.9652603 | -0.8081659 | -0.8194794 | -0.7889695 |
| life_expectancy | -0.6716665 | -0.3160269 | 0.7089370 | 0.9573293 | -0.7662491 | 0.9652603 | 1.0000000 | -0.7020221 | -0.7582627 | -0.7673134 |
| unempl_rate_t1 | 0.7873320 | 0.2418358 | -0.3781827 | -0.8034333 | 0.6117616 | -0.8081659 | -0.7020221 | 1.0000000 | 0.7960144 | 0.4999910 |
| unempl_rate_t2 | 0.5377599 | 0.3094522 | -0.4326616 | -0.7721717 | 0.5473647 | -0.8194794 | -0.7582627 | 0.7960144 | 1.0000000 | 0.7697083 |
| unempl_rate_t3 | 0.4347001 | 0.1836177 | -0.4665037 | -0.7490915 | 0.4355353 | -0.7889695 | -0.7673134 | 0.4999910 | 0.7697083 | 1.0000000 |

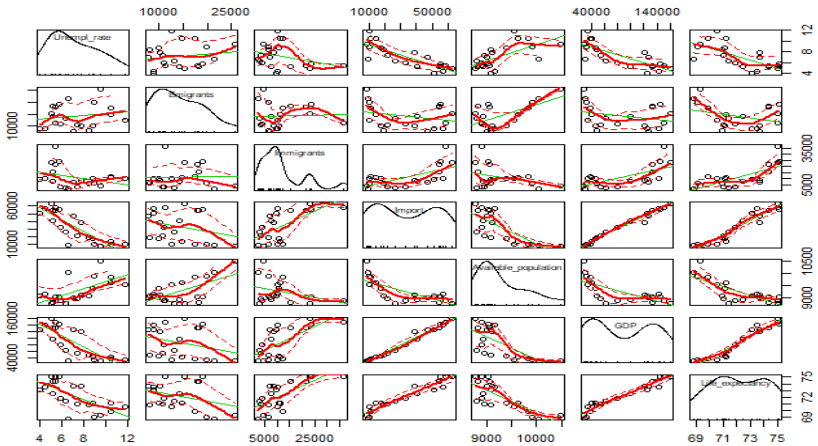
*Source*: [www.insse.ro](http://www.insse.ro)

The next analysis made is 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. [4]

If we look at the correlation between the other indices, we have high correlation between the number of immigrants and life expectancy (0.70), between import and life expectancy (0.95). The number of immigrants is highly correlated with GDP (0.86) and with the import (0.85). The import has the biggest correlation value with the GDP (0.99). We also have low correlation between the indices included in the analysis. The number of emigrants is very low correlated with the number of immigrants (0.04), the number of emigrants import is low correlated with the import (-0.27) and with GDP (-0.27).

Next, we made a scatterplot between each index and the rest of them, in order to see the graphic representation of the correlation. We can see that we have both positive and negative correlations between the indices included in the analysis. The very strong correlation between import and GDP and import is represented in the below figure showing us the linear distribution.

*Figure 2. Scatterplot matrix*

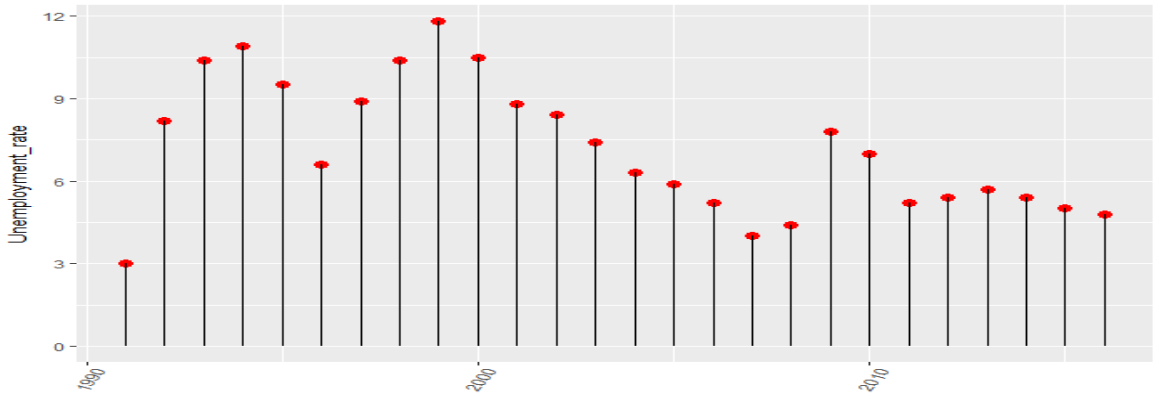



*Source*: [www.insse.ro](http://www.insse.ro)

## 3. Results and discussion

Next, we analyzed the evolution of the unemployment rate starting with 1991 until 2016. According with the below figure, we can observe that we have years for which we have the unemployment rate above 10% (1993, 1994, 1998, 1999 and 2000) with the highest rate in 1999 – 11.8%. At the opposite side we have years with a level of unemployment rate below 5% (1991, 2007, 2008, 2015 and 2016) with the lowest level in 1991 – 3%.
Starting with 2011 we can see that the unemployment rate is quite stable with values vary from 5.7% in 2013 to 4.8% in 2016.

*Figure 3. Unemployment rate in Romania for 1991 – 2016 years*




*Source*: [www.insse.ro](http://www.insse.ro)

After doing some tests, we reduced the number of indices included in the multiple regression. Therefore, our database for the multiple regression contains only the following indices: unemployment rate (dependent variable), import (independent variable), immigrants (independent variable). In the next part, we conducted a multiple regression with interaction (between the independent variables) in R with the LM model.
The equation of our model is:

$$\textbf{unempl\_rate} = \beta 0 + \beta 1 * \textbf{import} + \beta 2 * \textbf{immigrants} + \beta 3 * \textbf{import} * \textbf{immigrants}$$

Below you can find the output of the regression model.

*Figure 4. Regression output*

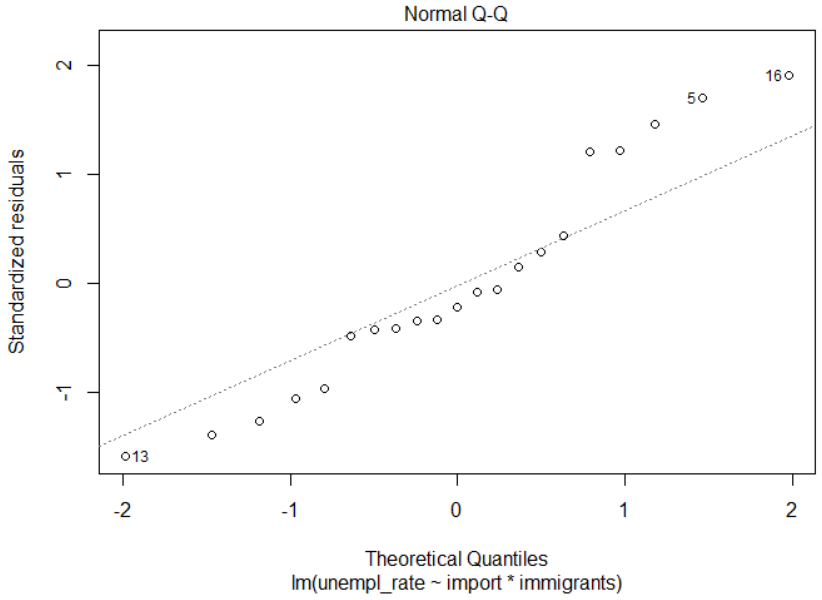
```
Residuals:
    Min      1Q  Median      3Q     Max
-1.2635 -0.3456 -0.1685  0.3490  1.5103

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       7.296e+00  8.313e-01   8.776 1.01e-07 ***
import           -5.744e-05  2.096e-05  -2.740 0.013958 *
immigrants        4.403e-04  9.940e-05   4.429 0.000367 ***
import:immigrants -6.715e-09  1.856e-09  -3.619 0.002120 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8615 on 17 degrees of freedom
Multiple R-squared:  0.8703,    Adjusted R-squared:  0.8474
F-statistic: 38.03 on 3 and 17 DF,  p-value: 9.294e-08
```

*Source*: [www.insse.ro](http://www.insse.ro)

Since our p-values are smaller than 0.05, we can affirm that our independent variables are statistically significant for this analysis. The overall p-vale has a value smaller than the critical value (9.294e-08 < 0.05) as well. The adjusted R-squared tells us the independent variables explain around 85% of the dependent variable variation.

In order to validate the model, we have to perform some more tests. The Q-Q plot shows that our data residuals is normally distributed. The points are around the line showing the normal distribution.

*Figure 5. Normal Q-Q plot*




*Source*: [www.insse.ro](www.insse.ro)

Since the correlation between our independent variables is medium (0.63), we do not have the multicollinearity assumption met.

Another test made in order to test the normality is the Shapiro Test. According to this test, we have the null hypothesis, which tells us that the data is normally distributed. This test returns two values – one of them is W (0.9697) which tells us that the data is almost perfectly fitted (W = 1 represent the "perfect fit"), the second value is the p-value (0.792). Since the p-value is higher than the critical value, we cannot reject the null hypothesis, which tells us that the data is normally distributed.

*Figure 6. Shapiro – Wilk normality test*

```
        Shapiro-Wilk normality test

data:  rstandard(model.lm.new)
W = 0.94494, p-value = 0.2724
```

*Source*: [www.insse.ro](www.insse.ro)

Below you can find the model equation:

**unempl_rate = (7.296e+00) + (-5.744e-05*import) + (4.403e-04*immigrants) + (-6.715e-09*import*immigrants)**

## 4. Conclusions

After performing the multiple regression with interaction, we obtained the coefficients of the independent variables. For the "import" variable, we obtained the coefficient = -5.744e-05, at one unit of "import" variable increasing the "unemployment rate" is decreasing with -5.744e-05 units.

For the "immigrants" variable we obtained the coefficient = 4.403e-04, at one unit of "immigrants" variable increasing, the "unemployment rate" is increasing with 4.403e-04 units. For the interaction of the independent variables we obtained a coefficient = -6.715e-09, at one unit of "import*immigrants" variables increasing, the "unemployment rate" is decreasing with -6.715e-09 units.

## 5. References


- Blanchard O., European Unemployment: The Evolution of Facts and Ideas, *National Bureau of Economic Research*, Working Paper 11750, November 2005, p. 6
- Hall R. E., 2003, Modern Theory of Unemployment Fluctuations: Empirics and Policy Applications, *IMPACT OF MODERN ECONOMIC THEORY*, Vol. 93, No. 2
- Hall R. E., JOB LOSS, JOB FINDING, AND UNEMPLOYMENT IN THE U.S. ECONOMY OVER THE PAST FIFTY YEARS, *NATIONAL BUREAU OF ECONOMIC RESEARCH*, Working Paper 11678, September 2005, p.3 (http://www.nber.org/papers/w11678 )
- www.insse.ro