

The Bucharest University of Economic Studies

FABIZ - IMBA

**Research Methods Project**

**Unemployment Analysis in Romania**

Examiners: Prof. Dr. Petre Caraiani

Prof. Dr. Daniel Traian Pele

Prof. Dr. Wolfgang Karl Härdle

Students: Amelia Daria **Nistor**

Sunita Maria **Georgescu**

Daria Elisa **Vuc**

Bucharest, 2024

**Introduction**

*Topic motivation:*

Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Romania's unemployment rate for 2022 was **5.44%**, a **0.15% decline** from 2021. The tendency of the unemployment rate to be unstable is because many job seekers are not satisfied with the salary and benefits that the employers provide, or they do not have the required skills by employers due to lack of experience. However, from a macroeconomic perspective, there are also some other factors that influence unemployment, which will be discussed in our paper. It is very important for a country to find solutions for the problem of unemployment, because it determines the level of satisfaction among people and influences the growth of the economy.

*Data:*

We downloaded our data from the World Bank website (<https://data.worldbank.org/indicator/NE.CON.GOVT.CD>, <https://data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS>, <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG>, <https://data.worldbank.org/indicator/SP.POP.GROW>, <https://data.worldbank.org/indicator/FR.INR.LEND> ).

We wanted to analyze which are the major determinants of unemployment in our country. Thus, the dependent variable is the unemployment rate while the independent or explanatory variables are the macroeconomic factors which are: interest rate, government expenditure, population growth and foreign direct investment.

For the scope of this project, we used data from 1992 to 2022.

The unemployment rate refers to the number of people unemployed in the labor force. Unemployment affects families' disposable income in a negative way, which also weakens their purchasing power, lowers employee morale, and lowers the production of an economy.

The unemployment rate (our dependent variable noted by y) is explained in our project by the following independent factors x):

- interest rate;

- government expenditure;

- population growth;

- foreign direct investment.

*Research questions:*

What are the most relevant variables that influence the rate of unemployment in Romania among government expenditure, foreign direct investment, inflation rate and population growth, from a statistical point of view? In other words, what drives unemployment in our country?

**Methodology**

The purpose of the research is to gain insights regarding the unemployment rate in Romania from 1992 until 2022. In terms of data we used the World Bank website, and for the analysis, we used Google Colab, where data has been stored, processed and displayed.

We read the unemployment rate csv file in panda. We edited the files in order to have only Romania as a country and its values. Using Plotlib, we calculated the correlation matrix and used it to generate a heatmap with Seaborn. We analyzed the heatmap and selected the factor that has the highest positive correlation with Unemployment rate. We continued with the multiple linear regression, pairplot and scatter plot.

* Import Libraries

import pandas as pd

import numpy as np

* Import Dataset

data\_df=pd.read\_csv('/content/drive/MyDrive/UnemploymentandVariablesDatasetFinal.csv')

data\_df.head()

|  | **Foreign direct investment** | **Government Expenditure** | **Interest rate** | **Population Growth** | **Unemployment rate** |
| --- | --- | --- | --- | --- | --- |
| **0** | 0.306508 | 3587916667 | 211.205588 | -0.903463 | 8.260 |
| **1** | 0.356585 | 3253996448 | 255.166862 | -0.136109 | 8.373 |
| **2** | 1.133915 | 4139835296 | 136.759393 | -0.145379 | 8.170 |
| **3** | 1.119418 | 4584710419 | 32.242485 | -0.202319 | 8.010 |
| **4** | 0.712022 | 4311709374 | 38.829301 | -0.288129 | 6.740 |

**Correlation index**

The correlation index tells us about the strength and direction of the linear relationship between variables.

corrolation\_list = ['Foreign direct investment', 'Government Expenditure','Interest rate','Population Growth','Unemployment rate']

up\_data = data\_df[corrolation\_list].corr()

display(up\_data)

| **Foreign direct investment** | **Government Expenditure** | **Interest rate** | **Population Growth** | **Unemployment rate** |  |
| --- | --- | --- | --- | --- | --- |
| **Foreign direct investment** | 1.000000 | 0.128906 | -0.363938 | -0.279338 | -0.145553 |
| **Government Expenditure** | 0.128906 | 1.000000 | -0.564051 | -0.136761 | -0.673612 |
| **Interest rate** | -0.363938 | -0.564051 | 1.000000 | 0.235475 | 0.408903 |
| **Population Growth** | -0.279338 | -0.136761 | 0.235475 | 1.000000 | -0.007826 |
| **Unemployment rate** | -0.145553 | -0.673612 | 0.408903 | -0.007826 | 1.000000 |

**Correlation Heatmap Display:**

In order to sustain the correlation index, we decided to do a graphical representation through correlation heatmap display, presenting the relationship between the variables as a color-code matrix: The lighter the color in the chart, the stronger the positive correlation and the darker the color in the chart, the stronger negative correlation.

For this we imported seaborn, which is a data visualization library based on matplotlib.

import seaborn as sns

import matplotlib.pyplot as plt

# to help you visualize these correlations, see the plots below

# the lighter the colour the higher the correlation

plt.rcParams.update({'font.size': 20})

plt.figure(figsize=(20,20))

sns.heatmap(up\_data, annot=True);



These are the Pearson correlation values: 1 = a strong positive correlation 0 = no correlation -1 = strong negative correlation.

**Multiple linear regression:**

We used multiple linear regression to see how multiple independent variables are used to predict the dependent variable.

First we created x and y, and for the purpose of the analysis, we separated the independent and dependent variables into separate tables.

* Define x and y

x=data\_df.drop(['Unemployment rate'], axis=1).values

y=data\_df ['Unemployment rate'].values

print(x)

[[ 3.06508326e-01 3.58791667e+09 2.11205588e+02 -9.03462928e-01]

[ 3.56585212e-01 3.25399645e+09 2.55166862e+02 -1.36109142e-01]

[ 1.13391484e+00 4.13983530e+09 1.36759393e+02 -1.45379065e-01]

[ 1.11941808e+00 4.58471042e+09 3.22424847e+01 -2.02318796e-01]

[ 7.12021743e-01 4.31170937e+09 3.88293013e+01 -2.88129486e-01]

[ 3.41529919e+00 4.51510476e+09 1.54763480e+02 -2.87897952e-01]

[ 4.87096009e+00 4.95877452e+09 5.90965828e+01 -2.06980242e-01]

[ 2.89543421e+00 6.01082383e+09 4.58037811e+01 -1.56978616e-01]

[ 2.78361317e+00 6.28060982e+09 4.56665940e+01 -1.29440040e-01]

[ 2.86420760e+00 6.38110664e+09 3.44770124e+01 -1.39542999e+00]

[ 2.48342020e+00 6.82889316e+09 2.25398863e+01 -1.83065499e+00]

[ 3.18995908e+00 9.27492020e+09 1.52734890e+01 -7.21262297e-01]

[ 8.59368513e+00 1.10418160e+10 1.18743636e+01 -5.69786273e-01]

[ 6.60067176e+00 1.59889664e+10 9.01491283e+00 -6.17530957e-01]

[ 9.02006052e+00 1.94301639e+10 6.55851412e+00 -5.92402560e-01]

[ 5.78679016e+00 2.68701733e+10 4.83732928e+00 -1.47722298e+00]

[ 6.37738086e+00 3.40359401e+10 7.85080256e+00 -1.66638264e+00]

[ 2.66377051e+00 2.80384019e+10 5.58742004e+00 -8.33088755e-01]

[ 1.89010749e+00 2.55199345e+10 6.09141669e+00 -5.93959184e-01]

[ 1.23049331e+00 2.59357385e+10 5.78925329e+00 -4.91866213e-01]

[ 1.70128984e+00 2.46786229e+10 3.33492268e+00 -4.45177936e-01]

[ 2.03110515e+00 2.70617927e+10 3.98471236e+00 -3.71323063e-01]

[ 1.93737111e+00 2.88090052e+10 1.06830988e+00 -3.74575498e-01]

[ 2.42727525e+00 2.47530561e+10 -5.94156437e-01 -4.70052228e-01]

[ 3.37423267e+00 2.82190037e+10 -1.54479667e+00 -5.73660845e-01]

[ 2.83273374e+00 3.30964009e+10 1.33902121e+00 -5.78007015e-01]

[ 3.01811604e+00 4.05378808e+10 4.62548443e+00 -5.87493307e-01]

[ 2.93423090e+00 4.40300619e+10 3.82785433e+00 -5.26814822e-01]

[ 1.43315648e+00 4.69217425e+10 2.63107311e+00 -5.50759893e-01]

[ 4.10699674e+00 5.11053339e+10 5.05232876e+00 -7.46036479e-01]

[ 3.95173511e+00 5.04005460e+10 1.37954887e+01 -3.93250874e-01]]

print(y)

[8.26 8.373 8.17 8.01 6.74 5.52 5.63 6.24 6.97 6.56 8.11 6.95

7.72 7.17 7.27 6.41 5.79 6.86 6.96 7.18 6.79 7.1 6.8 6.81

5.9 4.93 4.19 3.91 5.03 5.59 5.436]

* **Split the dataset in training set and test set**

**from sklearn.model\_selection import train\_test\_split**

**x\_train,x\_test,y\_train,y\_test= train\_test\_split(x,y,test\_size=0.3,random\_state=0)**

* **Train the model on the training set**

**from sklearn.linear\_model import LinearRegression**

**ml=LinearRegression()**

**ml.fit(x\_train,y\_train)**

**LinearRegression**

**LinearRegression()**

* **Predict the test set results**

Here, these are all predicted results, which are 10.

**y\_pred=ml.predict(x\_test)**

**print(y\_pred)**

**[7.28107814 5.27836863 6.55628869 7.26529439 5.55475737 5.99108793**

**6.15337027 6.95343103 6.24221816 6.31009248]**

In order to see more clearly, we are using the values of our x variable from the first data set and value it to see how close it is from our actual outcome which is y:

| **0** | 0.306508 | 3587916667 | 211.205588 | -0.903463 |  |
| --- | --- | --- | --- | --- | --- |

**ml.predict([[ 0.306508 ,3587916667 ,211.205588, -0.903463]])**

**array([7.55093803])**

So, according to the model that we have prepared, the value of unemployment rate should be 7.55093803, compared to the actual value which is 8.260. It is not completely accurate, but close to the actual value.

* **Evaluate the model**

**from sklearn.metrics import r2\_score**

**r2\_score(y\_test,y\_pred)**

**0.580280325753812**

* **Plot the results**

**import matplotlib.pyplot as plt**

**plt.figure(figsize=(10,5))**

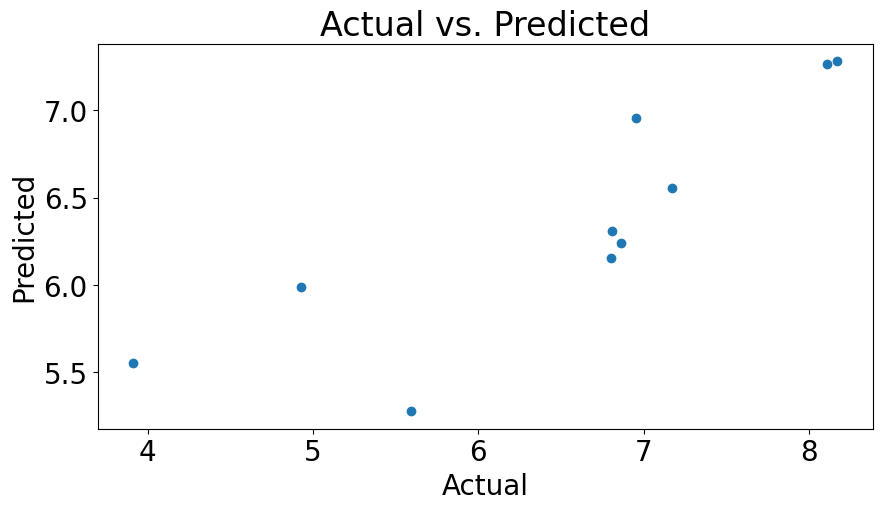
**plt.scatter(y\_test,y\_pred)**

**plt.xlabel('Actual')**

**plt.ylabel('Predicted')**

**plt.title('Actual vs. Predicted')**

**Text(0.5, 1.0, 'Actual vs. Predicted')**



* **Predicted values**

pred\_y\_df=pd.DataFrame({'Actual Value':y\_test,'Predicted value':y\_pred,'Difference': y\_test-y\_pred})

pred\_y\_df[0:30]

| **Actual Value** | **Predicted value** | **Difference** |  |
| --- | --- | --- | --- |
| **0** | 8.17 | 7.281078 | 0.888922 |
| **1** | 5.59 | 5.278369 | 0.311631 |
| **2** | 7.17 | 6.556289 | 0.613711 |
| **3** | 8.11 | 7.265294 | 0.844706 |
| **4** | 3.91 | 5.554757 | -1.644757 |
| **5** | 4.93 | 5.991088 | -1.061088 |
| **6** | 6.80 | 6.153370 | 0.646630 |
| **7** | 6.95 | 6.953431 | -0.003431 |
| **8** | 6.86 | 6.242218 | 0.617782 |
| **9** | 6.81 | 6.310092 | 0.499908 |

**Results**

1. **Correlation index:**

|  | **Foreign direct investment** | **Government Expenditure** | **Interest rate** | **Population Growth** | **Unemployment rate** |
| --- | --- | --- | --- | --- | --- |
| **Foreign direct investment** | **1.000000** | **0.128906** | **-0.363938** | **-0.279338** | **-0.145553** |
| **Government Expenditure** | **0.128906** | **1.000000** | **-0.564051** | **-0.136761** | **-0.673612** |
| **Interest rate** | **-0.363938** | **-0.564051** | **1.000000** | **0.235475** | **0.408903** |
| **Population Growth** | **-0.279338** | **-0.136761** | **0.235475** | **1.000000** | **-0.007826** |
| **Unemployment rate** | **-0.145553** | **-0.673612** | **0.408903** | **-0.007826** | **1.000000** |

The table shows the correlation coefficients between five economic indicators in Romania: unemployment rate, foreign direct investment, government expenditure, interest rate and population growth. The correlation index shows that there are weak to moderate correlations between the 5 economic indicators.

A correlation coefficient of 1 present perfect positive relationship, 0 represents no linear relationship and -1 present perfect negative relationship.

Taking into account the dependent variable y -”unemployment rate” to each x- independent variable:

* Unemployment rate and foreign direct investment have a negative correlation: -0.145553, meaning that the variables move in opposite directions, if foreign direct investment decreases, unemployment rate increases and vice versa.
* Unemployment rate and government expenditure have a moderate negative correlation: -0.673612, meaning that if government expenditure decreases, unemployment rate increases and vice versa.
* Unemployment rate and interest rate have a positive correlation: 0.408903, meaning that the variables move in the same direction, so if interest rate increases, unemployment rate will increase as well.
* Unemployment rate and population growth have a negative correlation: -0.007826, meaning that if population growth decreases, unemployment rate increases and vice versa.

**2. Correlation Matrix Heatmap:**



The figure above shows the heatmap generated from the correlation matrix. It is a diverging color palette representation of the correlation index, which is more aesthetic and appealing to our eyes. In order to interpret the heatmap, the Pearson correlation values are:

1 = strong positive correlation

0 = no correlation

-1 = strong negative correlation

Our heatmap presents:

* A positive correlation between unemployment rate and interest rate (0.41);
* A low negative correlation between population growth and unemployment rate

( -0.0078);

* A moderate negative correlation between unemployment rate and and government expenditure (-0.67) ;
* A low negative correlation between unemployment rate and foreign direct investment (-0.15).

So, interest rate and unemployment rate present the strongest positive correlation, and when interest rate increases, unemployment rate increases as well.

This means that population growth, government expenditure and foreign direct investment increase while unemployment decreases and vice versa.

1. **Multiple linear regression:**

For the purpose of the analysis, we separated the independent and dependent variables into separate tables.

* Predict test results:

ml.predict([[ 0.306508 ,3587916667 ,211.205588, -0.903463]])

**array([7.55093803])**

So according to the model that we have prepared, the value of unemployment rate should be 7.55093803, compared to the actual value which is 8.260. It is not completely accurate, but close to the actual value.

* Evaluate the model:

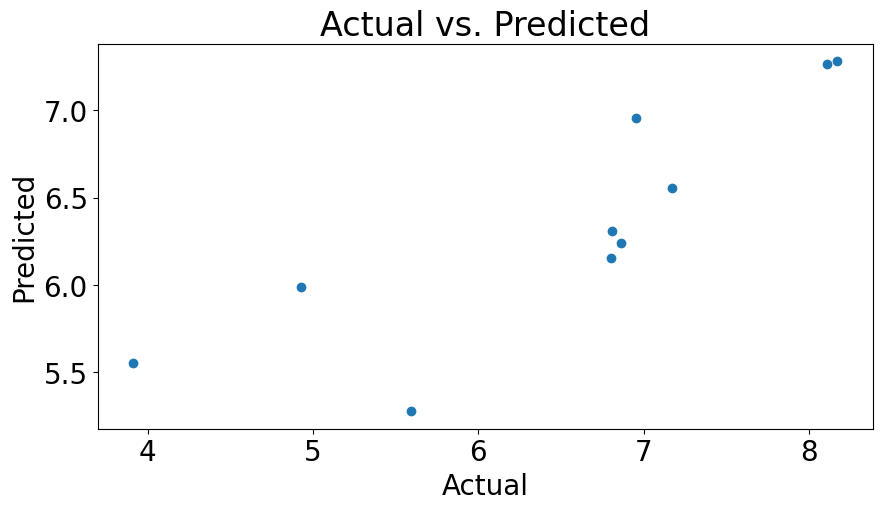
from sklearn.metrics import r2\_score

r2\_score(y\_test,y\_pred)

**0.580280325753812**

Our data set is fairly relevant, with a r2 score of 58%, but not completely. The r squared measurement says that 58% of the variance in the dependent variable (unemployment rate) is predictable from the independent variables.

* Plot the results:

****

The plot shows the relationship between the actual and the predicted unemployment rate in Romania. The actual unemployment rate is the percentage of the labor force that is actively unemployed , seeking for work. The predicted unemployment rate is the unemployment rate that economists forecast based on economic models.

In our graph, the actual unemployment rate has been higher than the predicted unemployment rate in recent years. This suggests that the Romanian economy has been growing more slowly than economists had expected.

The points on the graph fall roughly along a straight diagonal line, meaning that the actual and predicted unemployment rate have been relatively close to each other, and normally distributed.

Reasons why actual unemployment rate might be higher than the predicted unemployment rate:

-economic shock of Covid-19, the war in Ukraine;

-structural changes: the decline of the manufacturing sector and the rise of the service sector;

-the data is not always always perfect, leading to errors in forecast ;

**Conclusion**

While unemployment is affected by many external factors, we can conclude, based on our study, that unemployment decreases when foreign direct investments and government expenditures increase and it increases when interest rates increase and vice versa. If the population of Romania grows, the unemployment rate decreases, based on the correlation test results. In order to fight against a low unemployment rate, the government should spend more on creating facilities and jobs for graduates and try to attract foreign investments through EU funds, for instance. Further research can be focused on other independent variables that affect the unemployment rate in Romania by looking at more subjective factors such as minimum wages, benefits, work-life balance, working conditions and stress involved, but this would be harder because these variables are not necessarily quantifiable.