

---

# THE INFLUENCE OF THE UNEMPLOYMENT RATE, THE DEGREE OF INCLUSION IN EDUCATION AND THE AVERAGE MONTHLY NET EARNINGS ON THE GDP OF ROMANIA DURING 1995-2019

**Alexandra Diana CHIRESCU** (*chirescualexandra18@stud.ase.ro*)

*Faculty of Agri-food and Environmental Economics, Department of Agri-food  
and Environmental Economics, Bucharest University of Economic Studies*

## Abstract

*Unemployment rate, education coverage and average monthly net earnings are a set of indicators specific to the socio-economic dimension of a state. Thus, in the period 1995-2019, the indicators registered an oscillating trend in Romania, respectively a decrease of approximately 6.6 percentage points (p.p) in the case of the unemployment rate and an increase of approximately 8.3 p.p in the case of the degree of education. Moreover, the average net monthly salary decreased by approximately 98.59 p.p in the same period. The purpose of this paper is to perform an analysis of the indicators mentioned above and to identify the level of correlation between them and GDP. The research objectives are represented by establishing the degree of influence of independent variables, namely unemployment rate, degree of education and monthly average net earnings on the dependent variable, Gross Domestic Product, and identifying the importance of these indicators for the development of the country's economy. The present research aims to capture both the evolution of indicators and the impact they have on GDP, so it is considered that the study could contribute to the development of knowledge in this field and to solving practical situations. Regarding the structure of the paper, it consists of the following sections: Introduction, Review of Scientific Literature, Case Study, Conclusions and Bibliography. For the Review of the scientific literature section, the most important papers in the chosen field were analyzed and a bibliometric analysis was performed using the VOSviewer software. Next, the econometric study was performed for the chosen variables, using the Data Analysis function in Excel.*

**Keywords:** *econometric analysis, bibliometric analysis, socio-economic indicators, Romania, economic development*

---

### Introduction

Regression analysis is performed to determine the correlations between two or more variables that have a cause-effect relationship and to draw predictions about the phenomenon in question. Besides indicating the existence of relationships between the dependent variables and the independent variable, the regression analysis highlights the strength of the relationship between them, as well as the possibility of making predictions for the evolution of the dependent variable. The model for multiple linear regression can be formulated as below:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon,$$

where  $y$  = dependent variable

$x$  = independent variable

$\beta$  = parameter

$\varepsilon$  = error

The assumptions for multivariate regression are normal distribution, linearity, lack of extreme values, and lack of multiple links between independent variables (Uyanik & Güler, 2013). The sample data for this research were taken from the website of the National Institute of Statistics, more precisely from the Tempo Online platform. In order to give complexity to the study, a set of representative indicators for the socio-economic dimension were chosen, respectively the unemployment rate, the degree of inclusion in education and the average net monthly salary in Romania in the period 1995-2019. Therefore, by means of multiple linear regression, the existence of the correlation between the three variables and the Gross Domestic Product was verified. In other words, the research objectives are to establish the degree of influence of the independent variables, namely the unemployment rate, the level of inclusion in education and the average monthly net earnings, on the dependent variable, Gross Domestic Product, and identifying the importance of these indicators for economic development.

### Review of scientific literature

The review of the scientific literature was realised by using VOSviewer software. Bibliometric analysis involves conducting an inventory of journalistic activity in a particular field, as well as establishing collaborative links between states and keywords most often used by authors in their research (Hamidah, Sriyono & Hudha, 2020). Thus, approximately 2,000 documents published in the period 2010-2021 on the topic of linear regression were analyzed. The Scopus database was used to collect the data needed for the

---

analysis. The figure below (Figure 1) shows the Analysis of collaboration relationships between states.

### Analysis of collaboration relationships between states

Fig.1

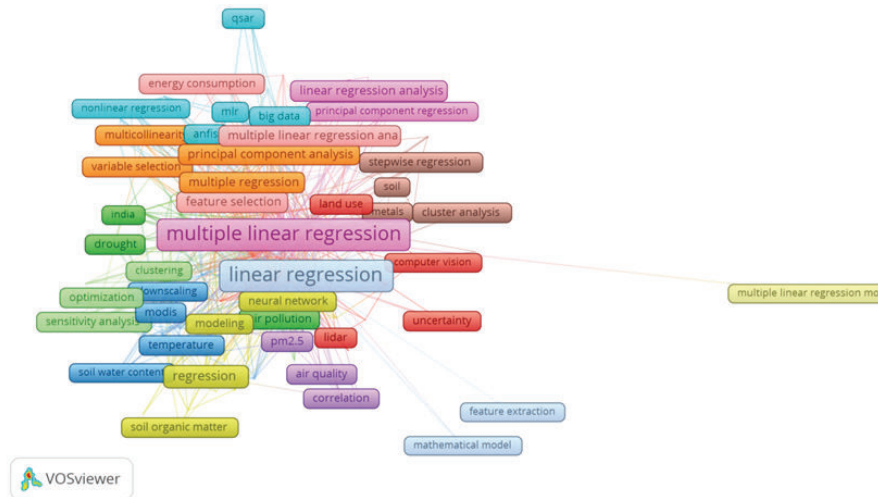


Source: own conceptualization based on Scopus database

The criteria taken into account for conducting the bibliometric analysis were that the minimum number of documents published by a state or citations to be 1. Thus, 107 states complied with this condition. On the first place in terms of the number of collaborative relationships, is the United States with 53 links, followed by China with 39 links and Spain with 23 links. In this analysis, Romania established 4 collaboration relationships with India, Serbia and the United States of America. Figure 2 shows the Analysis of the keywords used by the authors.

## Analysis of the keywords used by the authors

Fig.2



Source: own conceptualization based on Scopus database

According to the graph, the most commonly used word is „linear regression” with 249 occurrences, followed by „multiple linear regression” with 242 occurrences and „regression” with 26 occurrences. After consulting scientific articles in this field, it was found that multiple regression is considered to be an extremely flexible system for examining the relationship between a collection of independent variables (or predictors) with a single dependent variable (or criterion) (Aiken, West & Pitts, 2003). Also, multiple regression can be used to create predictions of the analyzed variables (Abuella & Chowdhury, 2015). In this context, in addition to the analysis of the correlation of the unemployment rate, the degree of inclusion in education and the average monthly net salary on GDP, predictions can be made about the evolution of these indicators.

### **Case Study. The influence of the unemployment rate, the degree of inclusion in education and the average monthly net earnings on the GDP of Romania in the period 1995-2019**

In this chapter it will be performed a multiple linear regression analysis in order to determine the level of correlation between the dependent variables and the independent variable. Table 1 presents the data used for the research.

### Data required to perform the multiple regression

Table 1

Year	GDP (mil.lei) Y - dependent variable	Unemployment rate (%) X1-independent variable	Degree of inclusion in education (%) X2- independent variable	Average monthly net salary (lei) X3- independent variable
1995	7.610,6	9,5	64	211.373
1996	11.387,7	6,6	64,7	321.169
1997	25.500,1	8,9	65,1	632.086
1998	37.007,7	10,4	66,3	1.042.274
1999	55.126,4	11,8	67,3	1.522.878
2000	80.873,1	10,5	68,9	2.139.138
2001	117.391,4	8,8	70,6	3.019.424
2002	152.271,5	8,4	72,9	3.789.202
2003	191.917,6	7,4	73,2	4.839.648
2004	244.688,3	6,3	72,7	5.986.386
2005	286.861,9	5,9	74,4	746
2006	342.762,6	5,2	76,7	866
2007	425.691,1	4	81,1	1.042
2008	539.834,6	4,4	85,9	1.309
2009	530.894,4	7,8	86,5	1.361
2010	528.514,5	7,0	88,3	1.391
2011	558.889,9	5,2	83,9	1.444
2012	591.799,1	5,4	82,9	1.507
2013	634.967,8	5,7	82,6	1.579
2014	669.703,9	5,4	73,6	1.697
2015	711.929,9	5,0	72,4	1.859
2016	763.652,5	4,8	72	2.046
2017	857.895,7	4	72,2	2.338
2018	951.728,5	3,3	72,2	2.642
2019	1.058.973,2	2,9	72,3	2.986

Source: own conceptualization based on data collected from Tempo Online database

According to the table (Table 1), the dependent variable (Y) is GDP, and the independent variables are the unemployment rate (X1), the degree of inclusion in education (X2) and the average monthly net earnings (X3). The data are analyzed over a long period of time, respectively 25 years, so that the model can be valid and relevant. Table 2 shows the first data obtained after conducting the regression.

**The results of the regression – Regression statistics**

*Table 2*

<i>Regression Statistics</i>	
Multiple R	0,84927678
R Square	0,72127105
Adjusted R Square	0,68145263
Standard Error	179905,255
Observations	25

*Source: own conceptualization*

The most important indicators in the analysis above are R Square and Adjusted R Square. Given that R Square has a high value of about 72%, it can be said that the variables analyzed have a high degree of correlation. This means that the evolution of the three independent variables influences the evolution of the dependent variable. A value of the Adjusted R Square indicator of 0.68 means that this regression model explains about 68% of the variation of the dependent variable (GDP) around the average value of the observations (sample average) (Dikov, 2020). Regarding the Standard Error, it provides an estimate of the standard deviation of the error (residues).

**The ANOVA test**

*Table 3*

ANOVA	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	1.758.828.164.475,48	586.276.054.825,16	18,114	4,88288E-06
Residual	21	679.683.914.608,32	32.365.900.695,63		
Total	24	2.438.512.079.083,80			

*Source: own conceptualization*

This table provides a general significance test on the regression parameters. Variance analysis, written briefly ANOVA, is a statistical technique that compares sample populations based on their mean and data distribution. The model helps us answer the question of whether the averages of two or more groups are significantly different. In this context, we can rely on the Analysis of Variance (ANOVA) to assess the impact of different factors by comparing the averages of different samples given by the mentioned factors. Thus, ANOVA is a form of hypothesis testing, where there are the following two hypotheses:

• Null hypothesis - all sample averages are equal or not significantly different in statistical terms.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_n$$

• At least one sample average is significantly different from the others.

$$H_1: \mu_l \neq \mu_k$$

Where l and k represent numbers from 1 to n.

The biggest disadvantage of the ANOVA test is that it provides indications whenever two groups have different sampling media, but does not help to identify the two groups. Taking into account the values of p in Table 4, compared to the threshold of 0.05 or 5% selected for alpha (significance level), it can be concluded that the null hypothesis should not be eliminated (Dikov, 2020). The following table provides information about the coefficients in the multiple regression model and is the most interesting part of the analysis.

### Regression analysis

Table 4

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	742050,4943	486579,1404	1,525035565	0,14217	-269846,2227	1753947,211	-269846,223	1753947,211
Unemployment rate (%) X1- independent variable	-96165,17079	17969,4782	-5,351583932	2,63E-05	-133534,7464	-58795,59515	-133534,746	-58795,59515
Degree of inclusion in education(%) X2- independent variable	4491,475508	5735,006291	0,783168366	0,442272	-7435,122975	16418,07399	-7435,12297	16418,07399
Average monthly net salary (lei) X3- independent variable	-0,030213399	0,023673171	-1,276271749	0,215794	-0,079444454	0,019017655	-0,07944445	0,019017655

Source: own conceptualization

Thus:

- Coefficients are estimates derived by the least squares method;
- Standard error - standard deviation of the least squares estimates;
- T-State — the null hypothesis according to which the coefficient is equal to zero, compared to the alternative hypothesis which means that it is different from zero;
- The p value for the t test
- Lower and upper coefficients of 95% define the confidence interval for coefficients

Table 5 shows the regression residues.

**Residual output**

*Table 5*

<i>Observation</i>	<i>Predicted GDP (mil.lei) Y-dependent variable</i>	<i>Residuals</i>	<i>Standard Residuals</i>
1	109549,507	-101938,9074	-0,605748154
2	388255,225	-376867,5252	-2,239447268
3	159478,063	-133977,9631	-0,796132761
4	8226,90363	28780,79637	0,171023162
5	-136433,541	191559,9405	1,138300216
6	-22851,7672	103724,8672	0,616360803
7	121668,099	-4276,699058	-0,025413286
8	147206,951	5064,549082	0,0300949
9	212982,02	-21064,41987	-0,125170396
10	281871,117	-37182,81684	-0,220950207
11	508819,225	-221957,3252	-1,318929577
12	586461,613	-243699,0128	-1,448124479
13	721616,992	-295925,8924	-1,758470516
14	704701,94	-164867,3396	-0,979685668
15	380433,673	150460,7269	0,894077736
16	465449,559	63064,94075	0,374748684
17	618782,773	-59892,87313	-0,355899413
18	595056,36	-3257,26002	-0,019355507
19	564857,191	70110,60923	0,416615924
20	553279,897	116424,0027	0,691822451
21	586351,3	125578,5996	0,746221505
22	603782,094	159870,4056	0,949992554
23	681603,704	176291,9961	1,047574022
24	748910,139	202818,3615	1,205200754
25	787814,961	271158,239	1,611294518

*Source: own conceptualization*

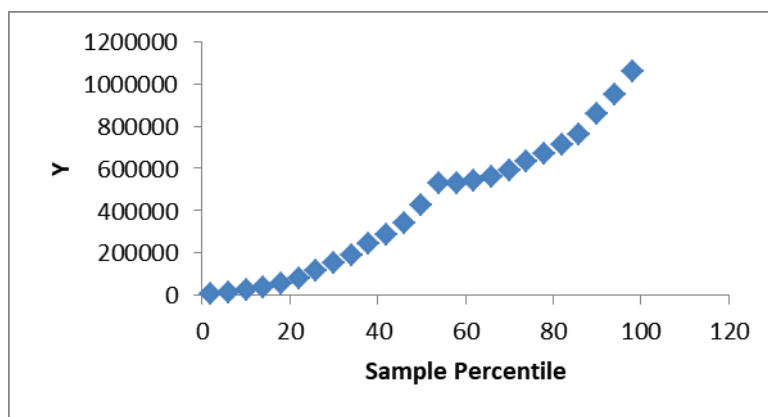
The residues provide information on how much the values of the real y deviate from the values of the predicted y. The figure below (Figure 3) shows the normal probability graph.



---

**Normal probability graph**

*Fig. 3*



*Source: own conceptualization*

The normal probability graph indicate whether the data fit a normal distribution, which is conclusive in this case.

### **Conclusion**

The purpose of this paper is to perform an analysis of the indicators mentioned above and to identify the level of correlation between them and GDP. Given the fact that the R Square indicator registered a value of 72%, it can be stated that the analyzed variables correlate. The research objective was achieved: the degree of influence of independent variables was established, respectively the unemployment rate, the degree of inclusion in education and the average monthly net earnings on the dependent variable, Gross Domestic Product. The regression model fits a homogeneous and normal distribution, and the null hypothesis cannot be rejected, as p recorded values higher than the significance level of 5%. The main barrier encountered during the study was the inexistence of a database containing a more complex inventory of indicators collected over a long period of time. However, this challenge has been overcome and the data chosen for the research is valid and relevant. It is considered that this paper is useful for those interested in analyzing the evolution of indicators in the socio-economic sector and not only, for those who want to determine the influence of independent variables on dependent variables. Moreover, the research can be used to solve practical problems, and the results can be extended and analyzed in detail. In this context, the variables chosen for the study can be included in a predictive analysis, which may prove useful in some cases.

---

### References

1. Uyanik, G. K., Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, 106, 234-240.
2. Hamidah, I., Sriyono, S., Hudha, M. N. (2020). A Bibliometric Analysis of Covid-19 Research using VOSviewer. *Indonesian Journal of Science and Technology*, 5, 209-216.
3. Aiken, L.S., West, S.G. and Pitts, S.C. (2003). Multiple Linear Regression. *In Handbook of Psychology*, I.B. Weiner (Ed.). doi: 10.1002/0471264385.wei0219
4. Abuella, M., Chowdhury, B. (2015). Solar power probabilistic forecasting by using multiple linear regression analysis. *SoutheastCon 2015*, 1-5 doi: 10.1109/SECON.2015.7132869
5. Dikov, D. (2020). Multiple Linear Regression Analysis in Excel. *Magnimetrics* consultat la: <https://medium.com/magnimetrics/multiple-linear-regression-analysis-in-excel-6f1436acd896>
6. Dikov, D. (2020). Perform a One-Way Analysis of Variance (ANOVA) in Excel. *Magnimetrics*. Consultat la: <https://magnimetrics.com/perform-a-one-way-analysis-of-variance-anova-in-excel/>