Aspects Regarding the Multiple Regression Used in Macro-economic Analysis

Prof. Constantin ANGHELACHE PhD.
Academy of Economic Studies, Bucharest
“Artifex” University of Bucharest

Assoc. prof. Alexandru MANOLE PhD.
“Artifex” University of Bucharest
Ligia PRODAN PhD. Student
Andreea Gabriela BALTAC PhD. Student
Zoica DINCA (NICOLA) PhD. Student
Academy of Economic Studies, Bucharest

Abstract
The regression function serves as a basis for carrying out the numerous analyzes micro or macroeconomic indicators. Information obtained by use of the model simple linear regression are not always sufficient to characterize changes in an economic phenomenon and, in particular, to identify possible future evolution of the latter. To remedy these shortcomings, in the literature had been entered into multiple regression models in which the dependent variable is defined on the basis of two or more variables factor incomes shall be described.

Key words: multiple linear regression, GDP, analysis, economics, macro-economics.

The economic situation in which correlations involves only two variables are very rare. Rather we have a situation where a dependent variable, \( Y \), can depend on a whole series of variables factorial or regressor. In practice, there are correlations of the form:

\[ Y = \beta_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + ... + \beta_kX_k + \varepsilon \]

where values \( X_j (j = 2, 3, ..., n) \) represents the variable factor or regressors, the values \( \beta_j (j = 1, 2, 3, ..., k) \) are the regression parameters, and \( \varepsilon \) is the residual factor. Residual factor reflects the random nature of human response and any other factors, others than \( X_j \), which might influence the variable \( Y \).

We adopted the usual notation, respectively assigned to the first factor notation \( X_2 \), the second notation \( X_3 \) and so on. Sometimes it is convenient that the parameter \( \beta \) to be considered that coefficient of one variable \( X_1 \) whose value is always equal to unity. Then the relationship is rewritten as:

\[ Y = \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + ... + \beta_kX_k + \varepsilon \]
In the case of regression with two variables \( E(\epsilon) = 0 \), then, substituting, for given values of the variables \( X \), we get:

\[
E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + ... + \beta_k X_k
\]

The relationship is multiple regression equation. For now, conventional, we consider that it is the linear form. Unlike the case of two-variable regression, we cannot represent this equation in a two-dimensional diagram. \( \beta_j \) are regression parameters. Sometimes, they are also called regression coefficients. \( \beta_1 \) is a constant (intercept) and \( \beta_2, \beta_3 \) and so on, are the regression slope parameters. \( \beta_4 \) measuring the effects of \( E(Y) \) produced by changing one unit of \( X_4 \), considering that all other factor variables remain constant. \( \beta_2 \) measures the effects on \( E(Y) \) produced by changing one unit of \( X_2 \), considering that all other variables remain constant factor. As the population regression equation is unknown, it has to be estimated based on data sample. Suppose that we have available a sample of \( n \) observations, each observation containing the dependent variable values for both \( Y \) and for each factorial variables \( X \). We write the values for observation \( i \) as:

\[
Y_i, \ X_{2i}, X_{3i}, X_{4i}, ..., X_{ki}
\]

For example, \( X_{37} \) is the value of \( X_3 \) in the 7th observation and \( X_{24} \) is the value \( X_2 \) taken in the 4th observation. For a similar manner, \( Y_6 \) is the variable \( Y \) in the observation of 6 and so on.

Given that it is assumed that the sample data were generated by the correlation of the population, each observation have to involve a set of values satisfy the multiple equation regression.

We can write the equation:

\[
Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + \epsilon_i
\]

for all the values, where \( \epsilon_i \) represents the residual value for the observation of the \( i \).

We can rewrite the relationship in a simple matrix form, as follows:

\[
Y = X\beta + \epsilon,
\]

where \( X \) is a matrix the form of \( n \times k \) with \( k \) column of values and then all sample values of the \( k-1 \), \( X \) variables.

Thus, the fourth column of \( X \), for example, contains the values of \( X_4 \) of the sample \( n \), the seventh column contains the values of \( X_7 \) and so on. \( \beta \) is a vector of \( k \times 1 \) column containing the parameters \( \beta_j \) and \( \epsilon \) is an vector of \( n \times 1 \) column containing the residual values.

The effective values of \( Y \) will not coincide with the expected values of \( Y \) and, in the case of two-variable regression, the differences between them are known as residual values.

\[
\hat{Y}_i = Y_i + \epsilon_i
\]

for all values of \( i \)

where \( \epsilon_i \) is the residual corresponding to the observations of \( i \).

The relationship can be written as:

\[
\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + ... + \hat{\beta}_k X_{ki} + \epsilon_i
\]

for all values of \( i \) or on matrix form:
\[ Y = X^\hat{\beta} + e \]

where \( X \) and \( Y \) are already defined.

There are two issues to be retained on the residual values. First, regardless of the method used to estimate the regression equation, we get such residual values - one for each of the sample observations. Second, as expected \( \hat{\beta} \) when it becomes known and can be used to calculate them. Now, we need to calculate the differential with the vector \( \hat{\beta} \) and equalizer to zero the result. Such of this matrix lead to the following relation:

\[
\frac{\partial S}{\partial \beta} = -2X'Y + 2X'X\hat{\beta} = 0
\]

The above equation is a set of \( k \) equations that can be written as:

\[ X'X\hat{\beta} = X'Y \]

### The average number of employees, on the activities of the national economy

- thousand persons -

<table>
<thead>
<tr>
<th>Year</th>
<th>Agriculture, forestry and fishing</th>
<th>Industry, including energy</th>
<th>Construction</th>
<th>Trade, repair of motor vehicles and motorcycles; transport and storage; hotels and restaurants; information and communications</th>
<th>Financial intermediation and insurance; Real estate transactions</th>
<th>Other service activities</th>
<th>Total labour force</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>762</td>
<td>3846</td>
<td>784</td>
<td>1427</td>
<td>429</td>
<td>988</td>
<td>8156</td>
</tr>
<tr>
<td>1991</td>
<td>768</td>
<td>3643</td>
<td>488</td>
<td>1344</td>
<td>338</td>
<td>1055</td>
<td>7574</td>
</tr>
<tr>
<td>1992</td>
<td>634</td>
<td>2246</td>
<td>680</td>
<td>1269</td>
<td>206</td>
<td>1022</td>
<td>1849</td>
</tr>
<tr>
<td>1993</td>
<td>648</td>
<td>3017</td>
<td>536</td>
<td>1153</td>
<td>294</td>
<td>1025</td>
<td>6672</td>
</tr>
<tr>
<td>1994</td>
<td>575</td>
<td>2856</td>
<td>515</td>
<td>1140</td>
<td>289</td>
<td>1043</td>
<td>6428</td>
</tr>
<tr>
<td>1995</td>
<td>693</td>
<td>2641</td>
<td>450</td>
<td>1201</td>
<td>262</td>
<td>1046</td>
<td>6409</td>
</tr>
<tr>
<td>1996</td>
<td>442</td>
<td>2586</td>
<td>431</td>
<td>1191</td>
<td>254</td>
<td>1035</td>
<td>5939</td>
</tr>
<tr>
<td>1997</td>
<td>357</td>
<td>2483</td>
<td>387</td>
<td>1193</td>
<td>218</td>
<td>999</td>
<td>5597</td>
</tr>
<tr>
<td>1998</td>
<td>316</td>
<td>2979</td>
<td>370</td>
<td>1188</td>
<td>210</td>
<td>979</td>
<td>5269</td>
</tr>
<tr>
<td>1999</td>
<td>244</td>
<td>1991</td>
<td>309</td>
<td>1044</td>
<td>236</td>
<td>972</td>
<td>4761</td>
</tr>
<tr>
<td>2000</td>
<td>199</td>
<td>1873</td>
<td>316</td>
<td>1022</td>
<td>248</td>
<td>965</td>
<td>4633</td>
</tr>
<tr>
<td>2001</td>
<td>167</td>
<td>1901</td>
<td>309</td>
<td>1008</td>
<td>249</td>
<td>969</td>
<td>4619</td>
</tr>
<tr>
<td>2002</td>
<td>162</td>
<td>1891</td>
<td>309</td>
<td>996</td>
<td>277</td>
<td>952</td>
<td>4508</td>
</tr>
<tr>
<td>2003</td>
<td>155</td>
<td>1884</td>
<td>325</td>
<td>1013</td>
<td>258</td>
<td>962</td>
<td>4571</td>
</tr>
<tr>
<td>2004</td>
<td>155</td>
<td>1741</td>
<td>323</td>
<td>1006</td>
<td>257</td>
<td>957</td>
<td>4469</td>
</tr>
<tr>
<td>2005</td>
<td>147</td>
<td>1672</td>
<td>308</td>
<td>1086</td>
<td>317</td>
<td>989</td>
<td>4659</td>
</tr>
<tr>
<td>2006</td>
<td>136</td>
<td>1632</td>
<td>352</td>
<td>1159</td>
<td>371</td>
<td>1017</td>
<td>4667</td>
</tr>
<tr>
<td>2007</td>
<td>127</td>
<td>1643</td>
<td>466</td>
<td>1241</td>
<td>452</td>
<td>1074</td>
<td>4885</td>
</tr>
<tr>
<td>2008</td>
<td>105</td>
<td>1606</td>
<td>478</td>
<td>1371</td>
<td>136</td>
<td>1365</td>
<td>5646</td>
</tr>
<tr>
<td>2009</td>
<td>110</td>
<td>1371</td>
<td>464</td>
<td>1339</td>
<td>136</td>
<td>1423</td>
<td>4774</td>
</tr>
<tr>
<td>2010</td>
<td>99</td>
<td>1257</td>
<td>537</td>
<td>1218</td>
<td>136</td>
<td>1365</td>
<td>4311</td>
</tr>
<tr>
<td>2011</td>
<td>90</td>
<td>1259</td>
<td>334</td>
<td>1227</td>
<td>126</td>
<td>1303</td>
<td>4249</td>
</tr>
<tr>
<td>2012</td>
<td>104</td>
<td>1296</td>
<td>356</td>
<td>1246</td>
<td>121</td>
<td>1320</td>
<td>4443</td>
</tr>
</tbody>
</table>

Source: Statistical Yearbook, the AVERAGE NUMBER of employees, on the ACTIVITIES of the NATIONAL ECONOMY

Agriculture, forestry and fisheries in the period 1990-1999, a reduction in the number of employees. So if in 1990 were in this branch employees, 762 in
1996 were 442 employees, and in 1999 it was at 244 employees. The period between the years 2000-2012 is characterized by a significant drop in the number of employees, especially in 2010 when there were 95 employees in 2011 and their number was 98. In 2012 we notice a slight increase in the number of employees, 104.

In the field of industry and energy, we are also to a reduction in the number of employees, which in 1990 were 3846, and in 2012 the number fell to 1296. We find, as in the previous fall in veriginoasă branch of the number of employees since 2000.

Constructions, constitutes another branch of national economy which has been confronted also with the reduction from one year to another in the number of employees, with major discrepancies between 1990 and 2002, whereas the number of employees was reduced to more than half, from 704 to a total of 300. After an increase in the number of employees in 2007-2009, starting in 2010 their number starts again to fall, and in 2012 begins to increase timidly.

Trade, repair of motor vehicles and motorcycles; transport and storage; hotels and restaurants; information and communication constitutes a series of activities of the national economy which have recorded a large number of employees in the period 1990-1999, and after 2000 we are witnessing a decline in them. In 2002 shall be registrata the smallest number of employees of the analysis period, concerned 986 employees.

Financial Intermediation and insurance; real estate constitutes a branch of national economy who knows fluctuations increasing and decreasing, and at the level of 2007 total number of employees was 422, very close to the total number of employees at the level of 1990, concerned 429.

Other service activities (professional activities, Scientific and technical knowledge; service activities and administrative support services activities; public administration and defense; social security from the public system; education; health and social work; activities of performances, Cultural and recreational; repair of household products and other services) represents a branch with dramatic changes to the number of employees. Thus, in the period 1991 to 1995 shall number of employees is increasing, and during the period 1996-2005, we are witnessing a reduction in their number. Starting with 2006, the number of employees begins to increase, the year with the highest number of employees in 2009.

In the case of total employment in the period 1990-2012 we are witnessing permanent fluctuations from one year to another marked by up and down movements in the number of employees. The largest dicrepanta we can observe between 1990, when the total number of employees in 8156 and 2011, the total number of employees has reached 4349, thus reducing by about half their number. An extremely important factor of drastically reducing the number of employees highlighted in the year 2011 is the impact economic-financial crisis on employment.
The equation model for multiple linear regression will show in the following way: $Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 + \varepsilon$

where:
- $Y$ – Total labour force;
- $a_0, a_1, a_2, a_3, a_4, a_5, a_6$ – the regression model parameters;
- $\varepsilon$ - variable, interpreted as error (disturbance, measurement error).

Thus, the regression model can be cited under this equation maths:

**Total labour force** $= a_0 + a_1BRANCH\_1 + a_2BRANCH\_2 + a_3BRANCH\_3 + a_4BRANCH\_4 + a_5BRANCH\_5 + a_6BRANCH\_6 + \varepsilon$

The regression model features

For an pertinent analyze of the correlation between the three macroeconomic indicators presented in the table above, it is necessary in a first step of this research to identify a number of features aiming the evolution of each indicator considered in the period under review. To prove this, using the software Eviews 7.2, we studied in the first stage, the evolution of the three indicators. As can be seen from analyzing the data series under investigation, especially as in the figure shown above, in the period considered, the three of our country's macroeconomic indicators have registered a steady growth from year to year,
except to this rule making 2000 and 2009 when there was a decrease of the three indicators.

The purpose of multiple regression (term used by Pearson, 1908) is to highlight the relationship between a dependent variable (explained endogenous effect) and a lot of independent variables (explanatory factors, exogenous predictors).

Multiple linear regression model equation will look like this:

\[ Y = b_0 + b_1 X_1 + b_2 X_2 + \varepsilon \]

in which:
- \( Y \) - Gross Domestic Product- GDP;
- \( X_1 \) - Final Consumption- CF;
- \( X_2 \) - Net investments- INV;
- \( b_0, b_1, b_2 \) - parameters of the regression model;
- \( \varepsilon \) is a variable, interpreted as error (disturbance, measurement error).

The regression model may be rewrite under the following mathematical equation:

\[ \text{PIB} = b_0 + b_1 \text{CF} + b_2 \text{INV} + \varepsilon \]

**Evolution of GDP, final consumption and net investment in Romania in the period 1998 - 2011**

To estimate the regression model parameters we used the software Eiews 7.2 in which we defined an equation that has as outcome variables GDP, and factor variables the final consumption and net investments. We also thought that this
A regression model will also include a free term, which is expected to influence the dimming terms that were not taken into account when building this model. Estimation method defined in the program is the method of least square.

Based on the above, using Eviews 7.2 we obtained the following results:

![Regression Model Output]

**Characteristics of the regression model**

From the above, multiple regression model describing the relationship between the three macroeconomic indicators that are the subject of previously determined may be given in the form of equation as follows:

\[ \text{PIB} = -8.927569 + 1.165488 \, \text{CF} + 0.284958 \, \text{INV} \]

Thus, we can say that an increase with a monetary unit of final consumption (with its two components - private consumption and public consumption) will lead to an increase of 1.165488 monetary units of gross domestic product value. In case of the net investment, the difference is more significant, we can see that every leu invested brings an increase of only 0.284958 lei of the level of gross domestic product. This situation corresponds with the reality economics of Romania because in the last twenty years the Romanian economy was based almost exclusively on stimulating consumption and less on promotion of an investment policy correctly.

The influence of the free term as a picture of the factors that were not included in the analysis model is one significant. In fact, it can be said, that the factors that were not included in the econometric model of analysis, they have a significant decrease in the value of gross domestic product.

The probability for this model to be correct is very high - about 98.89%, this conclusion can be formulated on the basis of statistical tests R-squared (0.988909) and Adjusted R-squared (0.986892).

Also the validity of the regression model is confirmed by the F test value - statistically superior value table level that is considered to be the benchmark in the
analysis of the validity of econometric models and by the value of the test Prob (F - statistic) that it is zero.

Based on observations made on the analysis of Romania's GDP, using multiple regression model, we conclude that the value of this indicator is significantly influenced by the variation of final consumption and net investment less variation.

Using a multifactorial regression model allows to obtain more edifying results in macroeconomic analysis and conducting relevant research on the evolution of the national economy.

References