THE PRACTICAL USE OF THE LINEAR MULTIPLE REGRESSION IN THE COMPLEX GDP ANALYSIS

Ligia DUDU[1], Raluca Georgiana MOSCU[2]

[1] *"Dimitrie Cantemir" Christian University – Bucharest, e-mail: prodanligia@yahoo.com*
[2] *"Dimitrie Cantemir" Christian University – Bucharest, e-mail: moscu.raluca@yahoo.com*

Abstract   *The economic and social phenomena meet the influence of several factors, and their contribution is more difficult to argue only through the dependence analysis of two variables. Using the simple regression model explains in a small contribution the factors of an economical result that evolving differently from year to year, provided that the factors are found together, in various proportions, in each economic outcome. Most often we use the multifactorial approach to analyze the evolution of a phenomenon.*
This article highlights the effective possibilities to use the multiple linear regression model in the analysis of gross domestic product. To emphasize the practical aspects related to the use of multiple linear regression in the macroeconomic analyzes, we have developed a practical study in which we defined as dependent variable the GDP, and the fluctuations will be explained by more causal variables. We used data with a yearly frequency, the economy of Romania and to compare we deflated, taking as a basis 1990. The objectives of this analysis are to determine the function that describes the best relationship between the indicators undergoing analysis, to observe the link that Is established between them and to estimate an econometric statistically significant valid model.

1. Introduction

The approach initiated in this article aims to identify the extent to which GDP correlates to households consumption, public consumption, gross investments, changes in inventories, net exports variation, gross disposable income as reflected on the basis of statistical data. In this regard, we consider the description of how the indicators are designed useful.

Final consumption consists of the value of purchases made by private households and public sector activity results that do not represent changes in the size of the material heritage sector.

Households consumption includes all purchases made by population and private organizations without lucrative character as private households don't have any accumulation, so no real heritage. Purchases of durable goods (capital) are as treated households consumption except housing construction, construction materials, processed goods in households products used for insemination, etc. witch represents intermediate consumption.

Public consumption is determined by decreasing production value of public sector revenues from the sale of public services and investment in their own activities.

Production value is represented by the public sector current expenditure on its employees, depreciation of capital goods and intermediate consumption. In the public sector amortization is calculated only for construction, equipment and machinery.

Investments represents the value of new durable goods witch are acquired from the market or are produced by resident producer units for use on longer than one year in their production processes.

Change in inventories is the difference between the value of stocks bought by institutional unit and the amount of the redundant calculation period; minus current and possible losses due to physical deterioration of stocks, accidental damage or theft during the period considered. The stocks comprise all goods that are not part of fixed capital and who are, at one time, in possession of the economical entities.

Net exports is the difference between the value of goods and services exported and the amount of the imported. Export value is determined based on FOB prices and imports on CIF prices. For the calculation of the balance of foreign trade operations turns CIF value of imports in FOB imports on the basis of a coefficient of CIF / FOB conversion.

Income is an important determinant of consumption, from size and dynamics that constitute the main source of economic goods procurement.

Account 4 "account of income redistribution", highlights the shift from national income (account balance 3) to national disposable income that corrects the balance of receipts and payments in relation to foreign countries. Account 4 consideres all the secondary distribution or redistribution of incomes. It shows in the right side net value added (net national product at the macro level) at the prices of factors and current transfers from abroad. The left side shows current transfers to abroad. The account balance is the national income available.

From the account 4 the national disposable income is deducted, according to:

VND = VN + (TCIS – TCPS) = VN + STCS;

in which: VN = National income;
TCIS = Current transfers received from other countries;
TCPS = Current transfers paid to other countries;
STCS expresses the balance as current transfers with foreign countries.

In this regard, we chose some data on Romania covering the spectrum of economic and political and regional developments over the last two decades. The analysis aims, on the one hand, highlighting correlations between economic development indicators over the analysis period and, on the other hand, the links between them at a time.

2. Literature review

The synthetic indicator that characterizes economic growth is the gross domestic product. Under these conditions, economic growth is defined as a complex process of increasing the size of the results of the national economy, based on the combination and the use of direct inputs: labor, fixed capital and working capital consumption materials.

A series of works - Anghelache C., Isaic-Maniu AL., Mitruţ C., Voineagu V. (2007), Anghelache, C., Capanu, I. (2003), Capanu, I., Wagner, P., Mitruţ, C. (2004) addressed issues concerning the evolution of GDP, consumption role in the development of living standards population, macroeconomic analysis, System of National Accounts, macroeconomic modeling, statistical and econometric models etc. The results of the measurement principles in System of National Accounts were treated by Anghelache C., Isaic-Maniu AL., Mitruţ C., Voineagu V. (2007), Anghelache C. and contributors (2007), Anghelache C., Capanu I. (2004), Biji, M., Lilea, E., Roşca, E., Vătui, M. (2010).

3. The methodology used

There are few cases where the interrelation between economic processes can be explained by the dependence of two variables, more frequent situations are where a complex phenomenon, Y, depends on a number of variables factor. In practice, economic processes are manifested in the form of linear models:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_k X_k + \varepsilon;$$

in which Y is the result of the variable values, X_i $(i =1, 2,..., k)$ is the variable factor values, a_i $(i = 0, 1, 2, ..., k)$ is the unknown regression parameters and ε is the variable residual, deviation (error) of the regression equation.

In these types of multiple correlation, the error reflects the image of the factors that have not been introduced into the model, different from X_i, which may influence the variable Y.

In the above relation, we adopted the usual notation, attributing factor symbol X_0 first, second X_1 and so on. Considering that X_0 is constant, equal to unity, we obtained a linear model that includes a constant term (the term free model). a_0 parameter models the autonomous behavior of the outcome variable and influence a_i quantify intensity of the X_i factor on the variable Y.

In multiple linear regression the significance of regression parameters is the same as those of simple linear regression, with the remark that for each factorial characteristic we calculate and get one estimator. Based on these estimators, the existence, intensity and the direction of the dependency between the rezultative and factorial characteristic is interpreted. Also, in this model, the econometrico mathematical estimation process is the same as in simple linear regression. Although, for the multiple linear regression we analyze an increased number of factors, still a number of influences that include the residual variable remaines.

As the regression equation is unknown, it must be estimated on the basis of a sample. If we provide a sample of n observations, each containing values for both variable resultant, Y and each of the variables factor X, we write X_{it} value of variable i in recording t, where $i = 1, ..., k$ and $t = 1, ..., n$.

If you accept the hypothesis that the relationship between Yt and factorial variables is linear, then for the sampled recordings we can write:

$$\begin{cases} Y_1 = a_0 + a_1 X_{11} + a_2 X_{21} + \cdots + a_k X_{k1} + \varepsilon_1 \\ Y_2 = a_0 + a_1 X_{12} + a_2 X_{22} + \cdots + a_k X_{k2} + \varepsilon_2 \\ \vdots \\ Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + \cdots + a_k X_{kt} + \varepsilon_t \\ \vdots \\ Y_n = a_0 + a_1 X_{12} + a_2 X_{2n} + \cdots + a_k X_{kn} + \varepsilon_n \end{cases}$$

The above system of equations can be written, concentrated as follows:

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + \cdots + a_k X_{kt} + \varepsilon_t$$

The above formula can be synthesized in a simple matrix form, as follows:

$$Y = X \cdot A + \varepsilon$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ . \\ . \\ . \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1\ X_{11}\ X_{21}\ ...\ X_{k1} \\ 1\ X_{12}\ X_{22}\ ...\ X_{k2} \\ 1\ X_{13}\ X_{23}\ ...\ X_{k3} \\ . \\ . \\ . \\ 1\ X_{1n}\ X_{2n}\ ...\ X_{kn} \end{pmatrix}, \quad A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ . \\ . \\ . \\ a_k \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

where: *Y* is a column vector, *n x 1* shape, which has the *n* values of the resultive variable;

X is a matrix of the form *n x (k+1)* containing in the first column the constant 1, and in other columns *k* values for each of the *k* factorial variables;

A is a column vector of the *(k+1) x 1*, which contains *k + 1* parameters of the model;

ε is a column vector, shaped *n x 1*, which has as components the *n* values of the residual variable.

We consider that the data were used to estimate by means of a regression equation of the sample, of the form:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 + ... + \hat{a}_k X_k \ ,$$

in which: $\hat{a}_i$ is estimators to a_i, and $\hat{Y}$ is known as expected value of Y.

Please note that the actual values of *Y* are not the same as the predicted values of *Y*. As in the case of the two variable regression, the differences between them are known as residual values (u_t), as: $Y_t = \hat{Y}_t + u_t$, for all values of *t = 1, ... , n* or in the form of matrix

$$Y = X \hat{A} + u_.$$

On residual values, we need to remember two things. Thus, regardless of the method used to estimate the regression equation, we also obtain residual values - one for each sample observations. On the other hand, after estimation, $\hat{a}_i$ becomes known and the equation can be used.

The estimation method best known for multiple regression is least squares method. In this case we choose the sample regression function and minimize the sum of squared residuals values.

4. Results
Data

To analyze the correlation between observed variables we used data on Romania, on a yearly basis, covering the period 1990-2015 that were collected from the National Statistics Institute publications and to ensure comparability of data we deflated, taking as a basis reference 1990.

The econometric analysis

In order to achieve the analysis of the corelation between the seven indicators is necessary to identify the particularities of the evolution of each size considered in the specified time frame. The graphical representation of the comparison favors the evolution of the seven sets of data considered, but also as a means of electing the statistical calculation methods and procedures.

The graphic method is more suitable than the number method for visual identification of data trend.

In this regard, by using Eviews 7.1 we studied in a first stage development of the seven indicators. Thus, studying of the seven indicators evolution of Romania during 1990-2015, allowed to obtain the following information and graphical material:
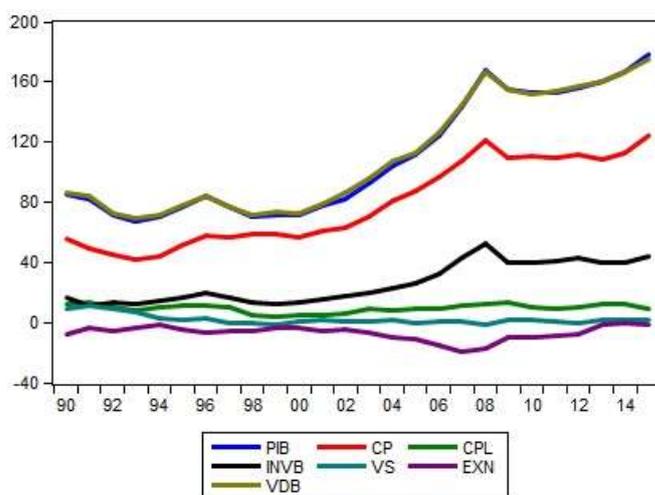
The evolution of GDP, households consumption, public consumption, gross investments, changes in inventories, net exports variation, gross disposable income, in Romania during 1990-2015

From the representation of indicators analyzed a clear link is resulted, at least in theory, manifested in economic cycles: when the economy is growing, the level of consumption, investment, national income increases and vice versa, during the recession, with the decreasing rate of the gross domestic product growth also the analyzed indicators fall.

In the early years there is a period characterized by economic instability, followed by a longer period from 1995 to 2007 where the growth rate of indicators analyzed had an upward trend. 2008 was the year in which the effects of the global economic and financial crisis arrived in our country having a strong impact on Romania's macroeconomic indicators (Penu, D., 2010). The next year brought a drop in indicators over the previous year. Since 2010 the indicators began to improve gradually due to a slight economic recovery, a trend explained by some positive domestic economic phenomena.

Therefore, we believe that the chosen period could be representative and used for this analysis because there is a match in the dynamics of the indicators where the quality of the econometric model and the coefficients of the variables are consistent and therefore more feasible in econometric modeling approach. From the evolutions of the indicators we can see that, in the period 1990 – 2015, between GDP and factorial variables chosen, it may be a direct linear relationship. It can be considered that there is a mathematical model to describe the relationship between several causative factors x and y effect factor.

Econometric analysis

The objective of this review is to ascertain primordial factors in the evolution of GDP, in Romania, between 1990 - 2015 and to estimate the action exercised by them using multiple linear regression model. The first step consisted in selecting variables describing the evolution of GDP namely, households consumption, public, gross investments, changes in inventories, net exports and available national income.

 The statistical data used in the regression model were collected from the National Statistics Institute publications and covers the period 1990 - 2015. Particular attention was paid to the values expressed in current prices. A deflationary data was necessary in order to ensure comparability of data, the basis of comparison is the year of commencement analysis, in 1990.

Based on this information, we analyze the existence of any addiction link between:
- the dependent or resultative variable - *Y*, the Gross Domestic Product, on the one hand and
- the explanatory, independent or causal variables:

- x_1 shows households consumption;

- x_2 shows the level of public consumption;

- x_3 shows the level of gross investments;

- x_4 shows changes in inventories;

- x_5 shows variation in net exports;

- x_6 shows the national income available.

By processing the data using the software Eviews 7.1, in which we defined the equation witch has the gross domestic product (GDP) as resultative variable and the factorial variables households consumption (CP), the public consumption (CPL), the gross investment (INVB), the change in inventories (VS), net exports (EXN) and available national income (VND), we have determined estimators of the equation using the method of least squares. We also considered that this regression model will contain free term C that contains the contribution of the factors not included in the model.

Results obtained by using Eviews 7.1 are as follows:

Dependent Variable: PIB
Method: Least Squares
Date: 04/10/16   Time: 22:07
Sample: 1990 2015
Included observations: 26

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.147736	0.125343	-1.178656	0.2531
CP	0.996288	0.014252	69.90309	0.0000
CPL	1.003705	0.017072	58.79076	0.0000
INVB	0.987696	0.013235	74.62480	0.0000
VS	0.993101	0.015140	65.59590	0.0000
EXN	0.994228	0.012891	77.12566	0.0000
VDB	0.006226	0.013413	0.464158	0.6478

R-squared	0.999997	Mean dependent var		109.8500
Adjusted R-squared	0.999996	S.D. dependent var		39.35390
S.E. of regression	0.077935	Akaike info criterion		-2.041073
Sum squared resid	0.115404	Schwarz criterion		-1.702354
Log likelihood	33.53395	Hannan-Quinn criter.		-1.943534
F-statistic	1062417.	Durbin-Watson stat		2.381546
Prob(F-statistic)	0.000000			

The results of the regression model parameter estimates

From the above, multiple regression model describing the relationship of the seven indicators is as follows:

$$PIB = -0.147736 + 0.996288 \cdot CP + 1.003705 \cdot CPL + 0.987696 \cdot INVB + 0.993101 \cdot VS + 0.994228 \cdot EXN + 0.006226 \cdot VDBGP$$

The analysis goes on with statistical significance of parameters included in the model, as well as the test of the validity of the model and its degree of reliability:
- C is the constant free, so a_0 coefficient is -0.147736 and consequential indicates the average level

of the resultative variable when all variables factor is zero. So, the average GDP that would be obtained, if there would be no consumption, gross investments would be void, we did not have inventories, net exports would not realize and did not have income would be -0.147736 millions lei. Since the t- statistical test probability attached is far superior of the threshold of 5% it means that the coefficient is insignificant;

- a_1 coefficient is 0.996288, which means that the households consumption growth by one million lei, while maintaining unchanged the other explanatory variables in the model, GDP will grow by an average of 0.996288 millions lei. Working at the level of relevance 5% the probability attached to the t- statistical test is below this level for households consumption exogenous variable;

- a_2 coefficient is 1.003705, meaning that at the increase of public consumption by one million lei, while maintaining unchanged the other explanatory variables in the model, GDP will grow by an average of 1.003705 millions lei. This coefficient is significant because the probability attached to the t-statistical test lies below 5%;

- a3 Coefficient is 0.987696, which means that the gross investment increased by one million lei, while maintaining unchanged the other explanatory variables in the model, GDP will grow by an average of 0.987696 millions lei. This coefficient is significant because the probability attached to t- statistical test lies below 5%;

- a_4 coefficient is 0.993101 which means that the increase in the stock of one million lei, while maintaining unchanged the other explanatory variables in the model, GDP will grow by an average of 0.993101 millions lei. Working at the level of relevance of 5%, the probability attached to t- statistical test is lower this level for exogenous variable change in inventories;

- a_5 coefficient is 0.994228, which means that net exports increase by one million lei, while maintaining unchanged the other explanatory variables in the model, GDP will grow an average of 0.994228 millions lei. Working at the level of relevance 5%, the probability attached to t- statistical test is below this level for exogenous variable net exports.

- a_6 coefficient is 0.006226, which means that to increase gross disposable income of households with one million lei, while maintaining unchanged the other explanatory variables in the model, GDP will grow, by an average of 0.006226 millions lei. Working at the level of relevance 5% the attached probability of the t-statistical test is superior to that level for the variable gross disposable income.

From the standpoint of statistical tests that check the quality of the econometric model considered, it can be seen that the values of the tests R^2, respectively adjusted R^2 are very high, tend towards 1, which lead

us to claim that the model used has a degree of minimal risk analysis of economic forecasting. The report of determination shows that the variables households consumption, public, gross investments, changes in inventories, net exports and gross disposable income exerts a decisive influence on all the variation in GDP, the share of all factorial variables simultaneously influence the total variation of the variable result is 99.99 %.

We can safely say that the model is statistically significant after applying the test F-statistic, whose value is greater than the value in the table, which is reckoned landmark in establishing the validity of econometric models and is valid for a significance level Prob (F-statistic) is zero, lower than 5%.

Given the above observations, we can say that the model chosen can be considered representative to describe the influence that households consumption, public, gross investments, change in inventories, net exports and gross disposable income have on GDP growth.

The multiple regression analysis followed the evolution of GDP under households and public consumption, gross investments, stock variation, net exports and gross disposable income, in Romania during 1990-2015. A linear relationship between the variables subject to the inquiry was identified. Estimated multiple regression model was shown to be a precise determination by having a determination coefficient close to 1.

5. Conclusions and recommendations

The econometric approach taken in the paper was a dominant component of statistical thinking, identifying their determinants and measurement factor.

The classic statistical regression, be it single or multiple, is always followed by successive testing of the models and aafter their validation, the economic correlations can become viable econometric models following a rigorous and stepwise procedure described from which the statistical tests can not miss (Durbin - Watson is a more important argument, even than the value of R-squared test "t", "F" and so on canceling apparently correctly specified models). The main restrictions that were overcome are related to ensuring comparability, gathering a more consistent database.

Based on the results obtained in this study and other research conducted recently, one can conclude that the country's economy increased based on imports consumption. Romania also needs to increase consumption but to be anchored in domestic demand, not imports.

The main measures that can be adopted for a more balanced economic growth, which should not be based only on consumption as before the crisis, are increasing

the value of gross fixed capital through investment, including foreign and balancing foreign trade balance through coherent stimulus export.

References

Anghelache, C. (2008) - „*Tratat de statistică teoretică şi economică*", Editura Economică, Bucureşti

Anghelache, C., Isaic-Maniu, AL., Mitruţ, C., Voineagu, V. (2007) - „*Sistemul conturilor naţionale*", Ediţia a II-a, Editura Economică, Bucureşti

Anghelache, C., Capanu, I. (2003) – „*Indicatori macroeconomici – calcul şi analiză economică*", Editura Economică, Bucureşti

Anghelache, C. şi colaboratorii (2007) – „*Analiză macroeconomică*", Editura Economică, Bucureşti

Biji, M., Lilea, E., Roşca, E., Vătui, M. (2010) - „*Statistica pentru economişti*" Editura Economică, Bucureşti

Capanu, I., Wagner, P., Mitruţ, C. (2004) – „*Sistemul Conturilor Naţionale şi Agregate macroeconomice*", Editura ALL, Bucureşti

Penu, D., (2010) - „*Current Directions of Fiscal Policy in Romania*", Revista Romana de Statistica – Supliment, pg.196- 201

Stancu, S., Andrei, T., Iacob, A.I., Tusa, E., (2008) - „*Introducere în econometrie utilizând Eviews*", Editura Economica, Bucureşti

*** Institutul Naţional de Statistică – *Anuarul Statistic al României*, Ediţiile 2000- 2015