MODEL FOR MACROECONOMIC - ANALYSE BASED ON THE REGRESSION FUNCTION

PhD Professor Constantin ANGHELACHE
Universitatea „Artifex” /Academia de Studii Economice - Bucureşti
PhD Professor Mario G.R. PAGLIACCI
Universita degli Studi di Perugia - Italia
PhD Candidate Ligia PRODAN
Universitatea Creştină „Dimitrie Cantemir”- Bucureşti

Abstract

Regression function is the key to making many micro and macroeconomic analysis. After studying logical variables to be analyzed, graphical representation of data series and primary interpretation of information we pass to fundament the econometric model to be used.

Key words: *gross domestic product, final consumption, simple regression, correlation*

***

The linear regression model involves identifying variables for defining the model and specifying the residual variable, the context in which the regression model is used. For the analysis chronological series (of time) we use a temporal function which, essentially, is also a regression, with a variable time (t). The goal of using the regression model is to obtain the parameters corresponding to the set of variables formulated by analyzing dependence between variables, where data series are recorded at the level of population statistics for a period or a moment, and to highlight the dependence between variables in a given time horizon.

In theoretical analysis, the dependence between variables is stochastic. The consideration of the residual variable in such a model is required. Other factors that influence outcome variable are grouped in residual variable.

Single-factorial nonlinear models are made linear by transformations that are applied to the variables regression model. So, for example, a model of form $y_i = a \cdot x_i^b$ is transformed into a linear model through the logarithms of the two terms above equality resulting linear function $\log y_i = \log a + b \cdot \log x_i$.

This model is recommended when points $(\log x_i, \log y_i)_{i=1,n}$ are located like cloud of points around a straight line.

Sometimes, for estimation the parameter we using other estimation techniques, which cannot be made linear by elementary transformations, estimation of parameter is done by numerical methods.

Linear regression model is based on data series for the two characteristics. These are represented by vectors x (variable factor) and y (variable outcome).

This requires to define the methods used to estimate the two parameters; specify the methods used for testing properties of estimators of regression model and establish how to use the regression model in conducting predictions.

In defining the linear regression function most commonly four hypotheses are considered, namely:

• data series are not affected by registration errors.

• for each fixed value of the characteristic factor, average residual variable is zero, namely:

$E[\varepsilon_i | X = x_i] = 0$, for any i,

• the lack of correlation between residues expresses that the residuals do not show the phenomenon of covariance, which implies

• hypothesis of gap between residual with independent variable, which implies that $\mathrm{cov}(X, \varepsilon_j) = 0$,for any j, showing a growth of factor variable values do not automatically lead to an increase of the residual variable values.

Based on four hypothesis we define linear regression model by function:

$y_i = b + a \cdot x_i + \varepsilon_i$, i = 1,..,n

Hhypothesis are made on the residual variable, namely:

$$\begin{cases} E(\varepsilon_i) = 0 \\ \mathrm{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, i \neq j \\ \sigma_\varepsilon^2, i = j \end{cases} \\ \varepsilon_i \rightarrow N(0, \sigma_\varepsilon^2) \end{cases}$$

$y_i = b + a \cdot x_i + \varepsilon_i$, i = 1,..,n

Hhypothesis are made on variable result:

$$\begin{cases} E(y_i|X = x_i) = b + a \cdot x_i \\ \text{cov}(yi, yj) = \begin{cases} 0, i \neq j \\ \sigma_\varepsilon^2, i = j \end{cases} \\ y_i \rightarrow N(b + a \cdot x_i \cdot \sigma_\varepsilon^2) \end{cases}$$

When between the two variables there is a linear dependence, using data sets (y_i,x_i), $i = \overline{1,n}$, the values of results variable is estimated by the relation $\hat{y}_i = \hat{b} + \hat{a}x_i$, and the residues series is estimate of equality: $e_i = y_i - \hat{y}_i = y_i - (\hat{b} + \hat{a}x_i)$.

Determination of the linear model parameters is usually made using the method of least squares or maximum likelihood.

If we use the method of least squares the values of resulting characteristics are estimated based on the relationship: $\hat{y}_i = \hat{b} + \hat{a}x_i$, where $\hat{a}$ and $\hat{b}$ are estimators of the regression parameters.

Real values of the characteristic result are equal to the estimate obtained using the regression model, adjusted by residual error, respectively:

$$y_i = \hat{y}_i + e_i$$

In estimation of parameters we start on the condition that the sum of squared differences between the actual and expected to be minimal, achieving equality:

$$\min_{\hat{a},\hat{b}} \phi(\hat{a}, \hat{b}) = \min_{\hat{a},\hat{b}} \sum_i^n e_i^2 = \min_{\hat{a},\hat{b}} \sum_i^n (y\hat{i} - \hat{b} - \hat{a}x_i)^2 .$$

The optimal conditions of the function lead to a system of two equations, respectively:

$$\begin{cases} \frac{\partial(\hat{a}, \hat{b})}{\partial(\hat{b})} = -\sum_i 2(y\hat{i} - \hat{b} - \hat{a}x_i) = 0 \\ \frac{\partial(\hat{a}, \hat{b})}{\partial(\hat{a})} = -2\sum_i (y\hat{i} - \hat{b} - \hat{a}x_i) \cdot x_i = 0 \end{cases}.$$

Equations are determined by applying the method of moments.

- The first equation follows from the condition $E(\varepsilon_i) = 0$ defining equality:

$$\frac{1}{n}\sum_i e_i = 0 \text{ or } \sum_i e_i = 0$$

- The second equation of the system is determined starting by hypothesis of mismatch of series of the values of variable factor with values of the residual variable values ($\text{cov}(X, \varepsilon) = 0$), satisfying the equality:

$$\frac{1}{n}\sum_i x_i e_i = 0 \cdot$$

In order to determine the two estimators we solve the linear system of results equations:

$$\begin{cases} n\hat{b} + \hat{a}\left(\sum_i x_i\right) = \sum_i y_i \\ \left(\sum_i x_i\right) \cdot \hat{b} + \hat{a}\left(\sum_i x_i\right) = \sum_i y_i \end{cases}$$

Testing if the solution satisfies the conditions of second order is done by determining the second order derivatives of the function[1]. The resulting matrix has two properties: it is defined positive and the matrix determinant is positive, respectively:

$$\Delta X = 4n\sum_i x_i^2 - 4\left(\sum_i xi\right)^2 = 4n\left[\sum_i (x_i - \bar{x})^2\right] > 0$$

The calculation formulas of the two estimators, $\hat{a}$ and $\hat{b}$ results from solving the linear system of equations.

Slope coefficient is obtained from the relationship:

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

1. Bardsen, G., Nymagen, R., Jansen, E. (2005) – *„The Econometrics of Macroeconomic Modelling”,* Oxford University Press

Using the least squares method has some disadvantages, such as:
- does not provide acceptable results if the hypothesis formulated are not satisfied;
- denoting by $\hat{a}^n, \hat{b}^n$ the estimators calculated by the series (x_i, y_i), $i = \overline{1,n}$ and by $\hat{a}^{n+1}, \hat{b}^{n+1}$ those evaluated for a series of values (x_i, y_i), $i = \overline{1,n+1}$, it follows that between the two pairs of estimators there is no simple relationship recurrence;

- estimators are distorted if data series have major changes.

The use of maximum likelihood in estimation of the parameters is based on the specification of residual distribution function.

Residual variable has the property:

$$\varepsilon_i \in N(0, \sigma_\varepsilon) \Leftrightarrow f(\varepsilon_i) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{\varepsilon_i^2}{2\sigma_\varepsilon^2}}$$

and from this we get $yi \in N(\tilde{b}, \tilde{a}x_i, \tilde{\sigma}_\varepsilon)$. The regression model is becoming specified when the parameters are solved $\tilde{a}, \tilde{b}$ and $\tilde{\sigma}_{\varepsilon,\varepsilon}$.

For linear regression model, the likelihood function is given by:

$$\ell(\tilde{a}, \tilde{b}, \tilde{\sigma}_\varepsilon^2) = \prod_{i=1}^{n} f(y_i / x_i).$$

Determining the form estimators is done using maximum conditions for log likelihood function.

$$L(a, b, \sigma_\varepsilon^2) = \ln l(a, b, \sigma_\varepsilon^2) = -\frac{n}{2}\left[\ln(2\pi) - \ln \sigma_\varepsilon^2\right] - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}(y_i - b - ax_i)^2$$

Based on the property of the logarithm function, we get:

$$\max_{\tilde{a}, \tilde{b}, \tilde{\sigma}_\varepsilon^2} \ell(\tilde{a}, \tilde{b}, \tilde{\sigma}_\varepsilon^2) \Leftrightarrow \max_{\tilde{a}, \tilde{b}, \tilde{\sigma}_\varepsilon^2} L(\tilde{a}, \tilde{b}, \tilde{\sigma}_\varepsilon^2).$$

We find that the by the maximum likelihood method we obtain the same set of estimators for parameters model as in the case of the least squares method.

When we use maximum likelihood method we obtain directly the estimator of residual variable dispersion.

The aim of the simple regression is to highlight the relationship between a dependent variable explained (endogenous, result) and an independent variable (explanatory factors, exogenous, predictors).

Exemple:

To build a linear regression model we defined the final consumption as an independent variable, while GDP value was considered as a dependent variable.

To determine the parameters of the linear regression model we considered a range of data on the evolution of the results of the two macroeconomic indicators.

Evolution of GDP and final consumption of Romania during 1998-2011

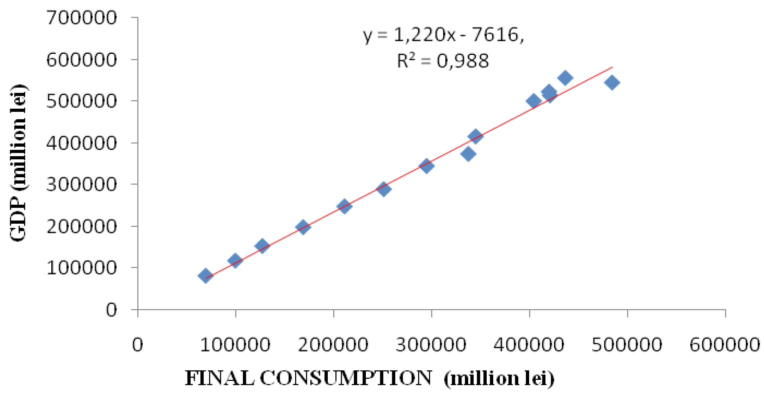
million lei

Year	GDP Y_X	FINAL CUNSUMPTION X
1998	373798.2	337468.6
1999	545730.2	484361.5
2000	80377.3	69253.3
2001	116768.7	99473.7
2002	152017.0	127118.8
2003	197427.6	168818.7
2004	247368.0	211054.6
2005	288954.6	251038.1
2006	344650.6	294867.6
2007	416006.8	344937.0
2008	514700.0	420917.5
2009	501139.4	404275.5
2010	523693.3	419801.2
2011	556708.4	436485.0

Source: Romanian Statistical Yearbook, Gross domestic product, by type of use, NIS, Bucharest, 2008, 2009, 2010, 2011, 2012

In order to identify the typology of the regression function we performed a graphical representation of the pairs of points that includes GDP values and the corresponding final consumption.

Correlation GDP - final consumption



Based on the graphical representation we can say that between GDP and final consumption, there is a direct and linear form, respectively $Y_x = a + bX + \varepsilon$.

where: Y_x is the dependent variable (explained, endogenous, result),

a is the Y intercept (constant term),

b is the slope,

X is the vector of independent variable (explanatory, exogenous),

ε is a variable, interpreted as an error (disturbance, measurement error).

Based on the graph it is reasonable to assume that the average variable Y depends on X through a linear relationship. The calculations performed using linear regression model function we get parameters

a = -7616.882095 and a = b = 1.220180278. Consequently, regression function becomes: Y_X = -7616,882095 + 1,220180278 X..

Based on the above data, using Excel / Data Analysis, the following results were obtained:

Regression model estimation results in Excel

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9942589
R Square	0.9885508
Adjusted R Square	0.9875967
Standard Error	18784.648
Observations	14

ANOVA

	df	SS	MS	F	ignificance F
Regression	1	365605382319.68	365605382319.68	1036.1114	0.00
Residual	12	4234356045	352863003.7		
Total	13	369839738364.30			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-7616.882	12109.51335	-0.628999851	0.5411382	-34001.2	18767.4809	-34001.25	18767.4809
CF	1.2201803	0.037907119	32.18868441	5.106E-13	1.137588	1.30277279	1.1375878	1.30277279

Multiple R is the multiple correlation coefficient, in this case, the simple correlation between x and y. We note that between the GDP and the value of final consumption recorded in Romania during 1998 - 2011 there is a direct and very strong link, conclusion drawn from the value of Multiple R (0.9942).

R Square, R^2 is the coefficient of determination, which shows the validity of the chosen model for explaining the variation of y; Multiple R is obtained from R Square: $r=\sqrt{R^2}$, and in this example $\mathbf{R^2 = 0.9885}$ is a value close to 1 indicating that the model is chosen correctly, the final consumption, x, explain variation gross product, y, at the rate of 98.85%.

Adjusted R Square is a coefficient of determination corrected with degrees of freedom and has the same meaning as R^2.

Standard Error is the standard error and it shows how average observed values y_i deviate from the theoretical values that are on regression straight, $\hat{y}_i$ (in this case to ± 18784.648). This value raised to the power 2 is the dispersion residues.

Observations is n, the number of observations, here n = 14.

ANOVA is the analysis of version table. For the version due to factor x, Regression, the residual version, due to other factors unregistered, Residual, and total version, due to all factors, In total, we specify:

- df (degrees freedom), degrees of freedom: k - number of explanatory variables x (the simple regression k = 1), n-k-1 for residue (14-1-1= 12 degrees of freedom) and n-1 for total version (14-1 = 13); Sum df for Regression and Residual is equal to the Total df: k + (n - k - 1) = n - 1.

- SS stands for Sum Square, which is the sum of squares of deviations, called versions, as follows:

- Regression: $SS^2_{y/x} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$,

- Residual: $SS^2_{y/z} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$,

- Total: $SS^2_y = \sum_{i=1}^{n}(y_i - \bar{y})^2$.

Between these versions there is the relationship: Total = Regression + Residual, like: $SS^2_y = SS^2_{y/x} + SS^2_{y/z}$.

MS, short for Modified Sum, called modified amounts, actually, modified dispersion:

-Regression: $MS^2_{y/x} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{k}$, the version due to the regression model chosen,

-Residual: $MS^2_{y/z} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}$, dispersion of residue.

F, Fisher test of overall significance of the regression is the ratio of the two dispersions corrected by degrees of freedom $F = \frac{MS^2_{y/x}}{MS^2_{y/z}}$.

Intercept is the name for the constant term (constant) of model.

Coefficients - contains estimates of the coefficients a and b. From the values shown it follows that the estimated model in the example is:

GDP = -7616.882095 + 1.220180278 CF.

Also, the validity of this model is confirmed by regression tests F-statistic values (1036.1114-value table level higher than what is considered to be the benchmark in analysis validity of econometric models) and the degree of the neutral risk (reflected by the test Significance F value).

Lower 95%, Upper 95% - lower and upper limits of the confidence interval for the parameter. The 0.05 threshold limits are calculated automatically, regardless of the initialization procedure Regression. It can therefore be interpreted as the population linear model parameters are included in the following ranges:

-34001.24512 <a <18767.48093;

1.137587762 <b <1.302772794.

RESIDUAL OUTPUT			
Observation	*Predicted PIB*	*Residuals*	*Standard Residuals*
1	404155.648	-30357.448	-1.682066841
2	583391.4675	-37661.26752	-2.086762013
3	76884.62874	3492.671264	0.193524387
4	113758.9648	3009.735197	0.166765526
5	147490.9706	4526.0294	0.250781422
6	198372.3662	-944.7661657	-0.052348269
7	249907.7784	-2539.77836	-0.140725827
8	298694.8565	-9740.256497	-0.539694986
9	352174.748	-7524.147982	-0.416903286
10	413268.4424	2738.357618	0.151728846
11	505978.35	8721.650022	0.483255322
12	485672.1098	15467.29021	0.8570225
13	504616.2627	19077.03727	1.057033905
14	524973.5065	31734.89355	1.758389313

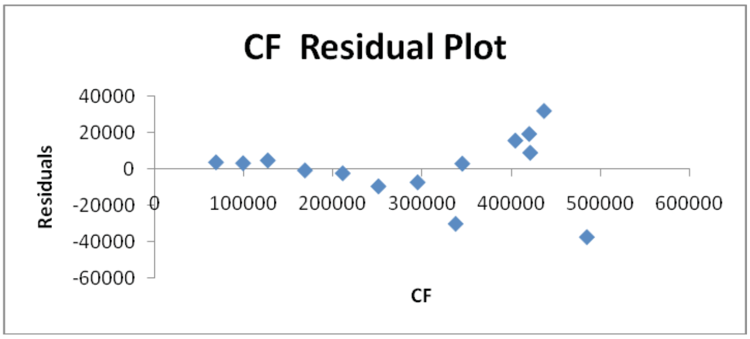
Predicted y - predicted y value for that observation, it is obtained by replacing the X's observation in the model estimated $\overline{Y_X}$ = -7616,882095 + 1,220180278X. Note that the sum of adjusted values $\overline{Y_X}$ is equals with the sum of the empirical Y_X which allows us to say that the estimate parameters of regression equation is correct.

Residuals – the value of the prediction error (difference between the observed and predicted value).

Reziduals standard - standard error value. It is obtained by dividing the residue to standard deviation of residue.

The quality analysis of the analysis model that we choose is facilitated by the graph below:

The independent variable vs. residue diagram

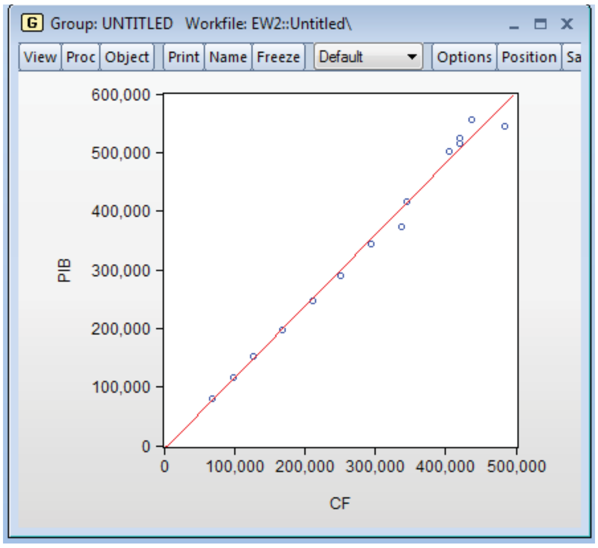


The graph leads to the conclusion, by the shape of the cloud of points, that there is no correlation between independent variables x and residue, that we can say that the model is well chosen.

To analyze the correlation between the evolution of GDP and the final consumption we have researched the data series that includes the values of the two indicators from 1998 to 2011. These were processed using the software package Eviews7.

A first step in this research was the identification of the type of econometric model that reflects the evolution of the phenomenon studied. To this end, we generated pairs of points chart GDP – CF.

Diagram final consumption vs. GDP



As you can see from the chart above, pairs of points follow a straight path, so it is possible to analyze the phenomenon investigated using simple linear regression model.

In this research we used GDP as dependent variable, while the final consumption is the independent variable. Also, in the model, I entered the free term C.

Results obtained using Eviews program is as follows:

The regression model's characteristics

Equation: UNTITLED Workfile: EW2::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PIB
Method: Least Squares
Date: 03/30/13 Time: 22:19
Sample: 1998 2011
Included observations: 14

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7616.882	12109.51	-0.629000	0.5411
CF	1.220180	0.037907	32.18868	0.0000

R-squared	0.988551	Mean dependent var	347095.7
Adjusted R-squared	0.987597	S.D. dependent var	168668.9
S.E. of regression	18784.65	Akaike info criterion	22.65103
Sum squared resid	4.23E+09	Schwarz criterion	22.74233
Log likelihood	-156.5572	Hannan-Quinn criter.	22.64258
F-statistic	1036.111	Durbin-Watson stat	0.519236
Prob(F-statistic)	0.000000		

From the analysis results that the simple linear regression model reflects the correlation between the value of GDP and of final consumption is as follows:

GDP = -7616.8 + 1.22 CF.

Conclusions

As we can see, the final consumption is an extremely important factor for the evolution of GDP. Thus, for an increase by a monetary unit of final consumption we will get an increase of 1.22 monetary units of GDP.

Also, it is noted that the value of the free term C is very high, which allows us to state that the factors that were not considered in building the design have a significant influence on the evolution of gross domestic product. Negative value of the free term shows that the variables which were not included in the econometric model have a negative effect on the evolution of GDP.

Selective bibliography:

- Anghelache C., (2008) – „Tratat de statistică teoretică şi economică", Editura Economica, Bucureşti
- Anghelache, C (coordonator) şi alţii (2012) – „Modele statistico– econometrice de analiză economic – utilizarea modelelor în studiul economiei României", Revista Română de Statistică (Supliment), ISBN 1018—046X
- Anghelache, C. şi alţii (2012) – „*Econometrie*", Editura Artifex, Bucureşti
- Anghelache, C. şi alţii (2012) – „*Elemente de econometrie teoretică şi aplicată*", Editura Artifex, Bucureşti
- Anghelache, C., Mitruţ, C. (coordonatori), Bugudui, E., Deatcu, C. (2009) – „*Econometrie: studii teoretice şi practice*", Editura Artifex, Bucureşti
- Bardsen, G., Nymagen, R., Jansen, E. (2005) – *„The Econometrics of Macroeconomic Modelling",* Oxford University Press
- Turdean, M.S., Prodan L., (2012) – „Statistică pentru afaceri", Editura ProUniversitaria, Bucuresti, ISBN 978-606-647-312-5
- Voineagu, V., Ţiţan, E. şi colectiv (2007) – "*Teorie şi practică econometrică*", Editura Meteor Press
- *** Institutul Naţional de Statistică – Anuarul Statistic al României, Ediţiile 2008, 2009, 2010, 2011, 2012