

ANALYSIS OF CONSUMER PRICES IN ROMANIA BY MULTIPLE REGRESSION

ANAMARIA POPESCU

ABSTRACT. This article describes a correlation between the annual index of total consumer prices, the annual index of consumer prices of food, the annual index of consumer prices of non-food products and the annual index of consumer prices of services, using a multiple regression model. The use of statistical-econometric models in macroeconomic analyzes can be used successfully by using the simple linear regression model, but multiple linear regression is preferred, because several factors influence the evolution of the resulting variables. In the case of multiple linear regression, the factors we have in mind must be identified and they must be included in the reconnected model following the same graph representation procedure, establishing the correlation chart to inventory the point cloud and the evolution of each variable, following as their basis to perform data interpretation and moreover by establishing the value of regression parameters to identify quantified intensity, direction of influence, in other words, the intensity of the correlation between the factors considered. In order for the inference based on the valid results of the linear regression to be valid, a set of hypotheses was tested, known as the normal classical model of multiple regression.

1. INTRODUCTION

The model explains the influence of the three types of consumption on the evolution of the annual index of total consumption prices and allows the realization of forecasts. The consumer price index is an economic indicator, which measures the overall evolution of the prices of purchased goods and service tariffs used by the population between two given periods of time [10].

In this study we used four sets of data on consumer price indices: the annual consumer price index, the total annual index on food consumption prices, the annual consumer price index of non-food products and the annual consumer service price index.

The purpose of this article is to determine the function that best describes the relationship between the four indices, respecting the links that exist between them and the estimation of a valid and statistically significant economic model. The econometric model will be estimated based on the data obtained from the website of the National Institute of Statistics of ROMANIA.

2. THEORETICAL METHODOLOGICAL CONSIDERATIONS

Many authors, Cao [3] and Cioran [4], have studied the dependence on the Consumer Price Index (CPI) from other variable economies and time series models. Multiple regression models are often used to model the relationship between a dependent variable, the response variable y , versus two or more independent explanatory variables, the economic variables x_i [2]. In a previous paper, Popescu [10] constructed a multiple regression model to explain the total CPI of food CPI, non-food CPI and non-food CPI (case study:

2010 *Mathematics Subject Classification.* 62P20, 91B84, 62J02, 62M10.

Key words and phrases. Multiple regression, correlation, econometric model, correlation, data analysis.

Romania). Falnic and Sipos [8] also propose a multiple regression model to explain CPI based on 7 independent variables (exchange rate, interest rate, output price etc.). Other authors work on inflation forecasting models Cogoljević et al. [5], DharmoEralda et al. [7]. Iyiolaand Adetunji[9] studied the effect of monetary policy on consumer price indicators (inflation rate, gross domestic product, private sector credit, broad money, net government credit and consumer price index) and showed that there is a dependence between them by estimating the multiple regression model. The validity of the estimated model is verified by testing for autocorrelation, multicollinearity and detection of multicollinearity. Also, Cruceru et al. [6] presented a model forecast by multiple linear regression with variables such as direct investigations, imported and exported, to represent graphs, establishing the correlation chart to make the inventory of point clouds and the evolution of each variable, data interpretation and, by estimating the regression parameters, they identified the intensity, the direction of influence, in other words, the intensity of the correlation between the considered factors. Other authors who analyze the econometric model of multiple regression are Anghelache et al. [1]. The paper includes the necessary steps to determine the model parameters, model testing and interpretation of the characteristics of the residual variable, depending on the values of these parameters. In the following through the established procedure and using existing data in the documents of the National Institute of Statistics and EUROSTAT, we proceeded to achieve a multiple linear regression. For the inference based on the results of the multiple linear regression to be valid, a set of hypotheses must be fulfilled, the regression based on this set of hypotheses being known as the normal classical model of multiple regression.

Hypotheses:

1. Defining the regression model;
 - form, variables and parameters of the regression model
 - graphical approximation of the model of the connection between the variables
2. Estimation of model parameters;
 - punctual estimation of the parameters
 - estimating the parameters by confidence interval
3. Testing the significance of the correlation and the parameters of the regression model;
 - testing the significance of the correlation
 - testing the parameters of a regression model
4. Testing the classical hypotheses on the regression model;
 - testing the linearity of the proposed model
 - testing the normality of errors
 - testing the homoscedasticity hypothesis
 - testing the hypothesis of autocorrelation of errors
5. The prediction of the value of the variable y in the hypothesis of the modification of the exogenous variables.

2.1. Defining the regression model. The regression model is developed using analytical tools in Microsoft Excel, which allow the automatic calculation of multiple linear regression using the Data Analysis option in the Tools menu of the Excel program.

2.1.1. Form, variables, parameters. Based on the data from National Institute of Statistics of Romania, can build a multifactorial econometric model of the form (1):

$$y = \beta_0 + \beta_i x_i + u_i, \quad i = \overline{1, 3}, \quad \hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (1)$$

that is, a multiple linear regression model with explanatory variables.

The variables of the model are:

- explained, dependent or endogenous variable y - TOTAL CPI (%)
- explanatory variables, independent or exogenous:
 - x_1 is the CPI of food;
 - x_2 is CPI Non-food goods;
 - x_3 is CPI Services.
- residual or error variable u_i , $i = \overline{1, 21}$.

The parameters of the model are:

- β_0 - point of intersection of the regression line with the Oy axis;
- $\beta_1, \beta_2, \beta_3$ - slope of the line, regression coefficient that shows how many units of measure y changes if x_1, x_2, x_3 changes one unit of measurement.

2.1.2. *Graphical approximation of the model of the connection between the variables.* Based on the collected data, the econometric model can be constructed for each independent variable, as seen in Figure 1, Figure 2, Figure 3:

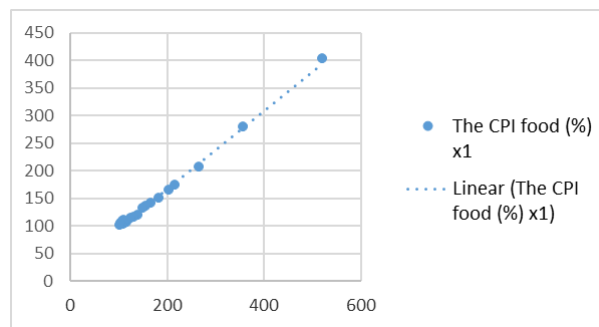


Figure 1. The regression model (y, x_1)

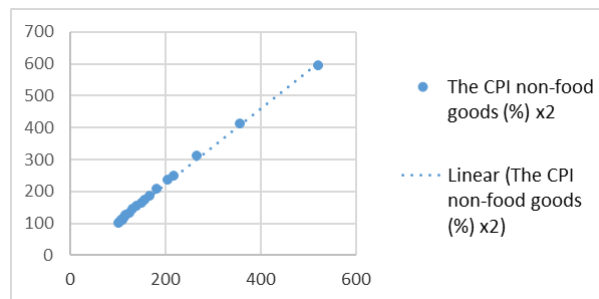


Figure 2. The regression model (y, x_2)

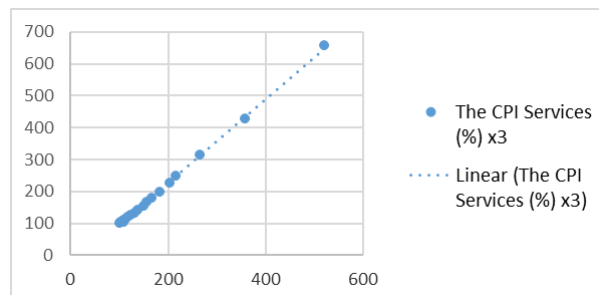


Figure 3. The regression model (y, x_3)

From the resulting graphs it can be seen that between the annual index of total consumer prices and the annual index of consumer prices of food, the annual index of consumer prices of non-food goods, respectively the annual index of consumer services prices there is a direct link that can be approximated by a line, therefore the model describing the connection is a linear one. With the help of this regression model, the relationship between the four variables is described.

2.2. Estimation of model parameters.

2.2.1. *Punctual estimation of parameters.* The estimates of the parameters resulting from Table 1 (Coefficients column) have the following values:

Table 1. Model parameters - Microsoft Excel/Data Analysis/Regression

ANOVA	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
	<i>Regression estimate model parameters (coefficients of the regression equation in the sample)</i>	<i>(Standard error or Mean deviation of estimators)</i>	<i>Calculated value of the test t</i>		<i>Lower limit of confidence interval</i>	<i>Upper limit of confidence interval</i>
Intercept	1.1726454	1.51882582	0.772074	0.450671	-2.0318	4.377088
x_1 CPI food	0.4094779	0.02893266	14.1528	7.76E-11	0.348435	0.470521
x_2 CPI non-food goods	0.4481466	0.01585169	28.27122	9.83E-16	0.414702	0.481591
x_3 CPI Services	0.1325711	0.0251131	5.278963	6.14E-05	0.079587	0.185555

$$\hat{\beta}_0 = 1.1726454, \hat{\beta}_1 = 0.4094779, \hat{\beta}_2 = 0.4481466 \text{ and } \hat{\beta}_3 = 0.1325711.$$

The estimated regression equation is:

$$\hat{y} = 1.1726454 + 0.4094779x_1 + 0.4481466x_2 + 0.1325711x_3. \quad (2)$$

2.2.2. *Estimation of parameters by confidence interval.* The confidence interval of the four parameters is estimated using the standard error related to them (3), *Standard Error* in Excel - SE (Table 1), and the statistical test, *t State*, from the same table.

$$\hat{\beta}_i - tStat \cdot SE(\hat{\beta}_i) \leq (\hat{\beta}_i) \leq \hat{\beta}_i + tStat \cdot SE(\hat{\beta}_i), \quad i = \overline{0,3}. \quad (3)$$

Thus, we can say, with a degree of confidence of 95 % that the intervals in which the parameters of the linear regression model are:

$\hat{\beta}_0 \in [-2.031796932; 4.37708781]$ - Intercept is the free term, so the estimated coefficient $\hat{\beta}_0$ is 1.1726454 and shows the average level of the dependent variable when the level of all explanatory variables is equal to 0 units (p.a). So the total CPI that would be obtained, if it were not any type of consumption, would be 1.1726454% (p.a.). Because the calculated value of the *t*-test statistic for testing the hypothesis: $H_0 : \hat{\beta}_0 = 0$ versus the hypothesis: $H_1 : \hat{\beta}_0 \neq 0$ is $t_{\hat{\beta}_0(\text{calc})} = 0.772074$, and the calculated (not imposed) significance threshold of the test, *P - value*, is $0.45067132 > 0.05 = \alpha$ means that the parameter β_0 is insignificant. Incidentally, the fact that the lower limit of the confidence

interval $CI_{95\%}(\hat{\beta}_0) = [-2.031796932; 4.37708781]$ for this parameter $\hat{\beta}_0$ is negative, and the upper limit is positive, it shows that parameter $\hat{\beta}_0$ in the general community is statistically insignificant, i.e. it does not differ significantly from zero.

$\hat{\beta}_1 \in [0.348435383; 0.47052051]$ - The coefficient $\hat{\beta}_1$ is 0.4094779, which means that when increasing an increase of the variable x_1 with a unit of measure (% p.a.), given that the variables x_2 and x_3 remain constant, the variable y will increase, on average by 0.4094779%. Because the calculated value of the t -test statistic for testing the hypothesis: $H_0 : \hat{\beta}_1 = 0$ versus the hypothesis: $H_1 : \hat{\beta}_1 \neq 0$ is $t_{\hat{\beta}_1(calc)} = 14.15279555$, and the calculated (not imposed) significance threshold of the test, P -value, is $7.76035E - 11 < 0.05$ which means that the parameter $\hat{\beta}_1$ is statistically significant (for a probability of $(100 - 7.76035E - 11)\% = 99.99\% > 95\%$). The 95% confidence interval for this parameter $CI_{95\%}(\hat{\beta}_1) = [0.348435383; 0.47052051]$ shows that with an increase in the CPI for foodstuffs by 0.41 pa, the Total CPI will increase, on average, by a value between about 0.35 and 0.47 pa, guaranteed interval with 95% probability.

$\hat{\beta}_2 \in [0.414702498; 0.48159079]$ - The coefficient $\hat{\beta}_2$ is 0.4481466, which means that when increasing an increase of the variable x_2 with a unit of measure (% p.a.), given that the variables x_1 and x_3 remain constant, the variable y will increase, on average by 0.4481466%. Because the calculated value of the t -test statistic for testing the hypothesis: $H_0 : \hat{\beta}_2 = 0$ versus the hypothesis: $\hat{\beta}_2 \neq 0$ is $t_{\hat{\beta}_2(calc)} = 28.2712182$, and the calculated (not imposed) significance threshold of the test, P -value, is $9.83294E - 16 < 0.05$ which means that the parameter $\hat{\beta}_2$ is statistically significant (for a probability of $(100 - 9.83294E - 16)\% = 99.99\% > 95\%$). The 95% confidence interval for this parameter $CI_{95\%}(\hat{\beta}_2) = [0.414702498; 0.48159079]$ shows that with an increase in the CPI for non-food foods by 0.45 pa, the Total CPI will increase, on average, by a value between about 0.41 and 0.48 pa, guaranteed interval with 95% probability.

$\hat{\beta}_3 \in [0.079587121; 0.18555516]$ - The coefficient $\hat{\beta}_3$ is 0.4481466, which means that when increasing an increase of the variable x_3 with a unit of measure (% p.a.), given that the variables x_1 and x_2 remain constant, the variable y will increase, on average by 0.1325711%. Because the calculated value of the t -test statistic for testing the hypothesis: $H_0 : \hat{\beta}_3 = 0$ versus the hypothesis: $\hat{\beta}_3 \neq 0$ is $t_{\hat{\beta}_3(calc)} = 5.278962694$, and the calculated (not imposed) significance threshold of the test, P -value, is $6.13886E - 05 < 0.05$ which means that the parameter $\hat{\beta}_3$ is statistically significant (for a probability of $(100 - 6.13886E - 05)\% = 99.99\% > 95\%$). The 95% confidence interval for this parameter $CI_{95\%}(\hat{\beta}_3) = [0.079587121; 0.18555516]$ shows that at an increase in the CPI for services by 0.13 pa, the Total CPI will increase, on average, by a value between about 0.08 and 0.19 pa, guaranteed interval with 95% probability.

Thus, we can write the mean square deviation or the standard deviation (*Standard Error* from ANOVA) of the estimators $\hat{\beta}_0$ and $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$.

The mean square deviation of the estimator $SE(\hat{\beta}_0) = 1.518825818$.

The mean square deviation of the estimator $SE(\hat{\beta}_1) = 0.028932655$.

The mean square deviation of the estimator $SE(\hat{\beta}_2) = 0.015851692$.

The mean square deviation of the estimator $SE(\hat{\beta}_3) = 0.025113104$.

Comments:

- $0.028932655 > 0$ x_1 increases. If x_1 increases by one unit, y is modified with 0.028932655 units.
- $0.015851692 > 0$ x_2 increases. If x_2 increases by one unit, y is modified with 0.015851692 units.
- $0.025113104 > 0$ x_3 increases. If x_3 increases by one unit, y is modified with 0.025113104 units.

2.3. Testing the significance of the correlation and the parameters of the regression model.

2.3.1. *Testing the significance of the correlation.* A comparison of *Significance F* and the significance threshold $\alpha = 0.05$ (Table 2) shows that:

Significance F = $1.93E - 40 < \alpha$ results that the model is reliable.

The Multiple Correlation Coefficient R, $R_{y/x_1, x_2, x_3} = 0.9999908 \in (0.95; 0.91]$ denotes a strong bond.

The coefficient of determination R Square, $R^2 = 0.9999908$ shows us that the independent variables x_1 (Food Goods), x_2 (CPI Non-Food Goods) and x_3 (CPI Services) influence the dependent variable y (TOTAL CPI) in proportion of 99%, the remaining 1% being other factors.

The regression is only based on a statistical assumption that expresses the average trend of the relationship between the dependent variable y and independent variables and measuring the intensity of the first step to the link, which is performed by the correlation method.(Table 2)

Table 2. Regression Statistics - Microsoft Excel/Data Analysis/Regression

Regression Statistics					
Multiple R	0.99999081	R			
R Square	0.99998161	R ²	r =	0.999991	
Adjusted R Square	0.99997837				
Standard Error	0.47572889				
Observations	21				
ANOVA					
	df	SS	MS	F	Significance F
Regression	3	209212.477	69737.49	308139.4	1.93E-40
Residual	17	3.84740559	0.226318		
Total	20	209216.324			

Mean square deviation of errors 0.4757289. If this indicator is zero it means that all points are on the regression line.

The time in the financial series generates a greater uncertainty of the econometric modeling, because it intensifies or significantly depreciates the correlations of the data series that characterize the financial markets, as found in the correlation matrix applied to the data series of CPI.

Table 3. Correlation matrix applied to the data series variables 1999–2019

	y	x1	x2	x3
y	1			
x1	0.998163	1		
x2	0.999274	0.995232	1	
x3	0.999555	0.998454	0.998170	1

The correlation coefficients resulting from the comparison of the dependent variable y și the independent variables x_1 , x_2 and x_3 being close to 1, result that there is a significant correlation between these variables, i.e. they are more strongly linearly dependent on each other.

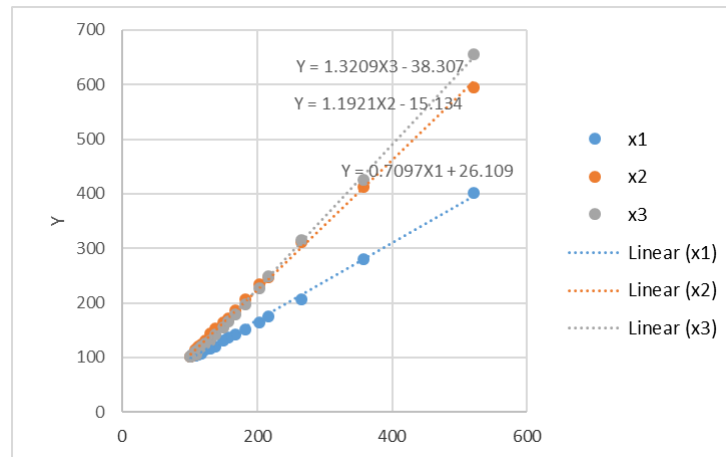


Figure 4. The correlation between y and x_1 , x_2 , x_3

The dependence between the independent variables and y is positive: an increase of the variables x_1 , x_2 and x_3 implies an increase of the total consumer price index, y . The regression lines have an upward trend, and the scatter plots indicate an upward trend, as seen in Figure 4.

2.3.2. Testing the parameters of the multiple regression model. Parameter testing is performed through *the Student test*. We need to check if the exogenous variable is independent or not, practically if it depends on the endogenous variable. We can see this if the resulting values of the coefficients are different or not from 0. This is done by comparing the P – value probability obtained with a threshold of 0.05. If the P – value obtained is greater than or equal to 0.05, the null hypothesis will be accepted, the coefficient being statistically different from zero, otherwise the null hypothesis will be rejected.

Testing the parameters of a multiple regression model is made with the level of significance value - P – value for each parameter in Excel - or by using *tStat* calculated (Table 4) as follows:

Table 4. Significance level of parameters–Microsoft Excel/Data Analysis/Regression

	<i>Coefficients</i>	<i>t Stat</i>	<i>P-value</i>
β_0	1.1726454	0.772074	0.450671
β_1	0.4094779	14.1528	7.76E-11
β_2	0.4481466	28.27122	9.83E-16
β_3	0.1325711	5.278963	6.14E-05

Testing the meanings of the parameters for $\hat{\beta}_1$

H_0 : if $\hat{\beta}_1 = 0$ - is not statistically significant

H_1 : if $\hat{\beta}_1 \neq 0$ - is statistically significant, significantly different from 0

$t_{\hat{\beta}_1} = 14.152$, and $t_{\alpha/2, n-2} = t_{0.05, 19} = 2.093$

$14.152 > 2.093$ H_0 is rejected, H_1 is accepted for $\hat{\beta}_1$.

Testing the meanings of the parameters for $\hat{\beta}_2$

H_0 : if $\hat{\beta}_2 = 0$ - is not statistically significant

H_1 : if $\hat{\beta}_2 \neq 0$ - is statistically significant, significantly different from 0

$t_{\hat{\beta}_2} = 28.271$, and $t_{\alpha/2, n-2} = t_{0.05, 19} = 2.093$

$28.271 > 2.093$ H_0 rejects, H_1 accepts for $\hat{\beta}_2$.

Testing the meanings of the parameters for $\hat{\beta}_3$

H_0 : if $\hat{\beta}_3 = 0$ - is not statistically significant

H_1 : if $\hat{\beta}_3 \neq 0$ - is statistically significant, significantly different from 0

$t_{\hat{\beta}_3} = 5.278$, and $t_{\alpha/2, n-2} = t_{0.05, 19} = 2.093$

$5.278 > 2.093$ H_0 rejects, H_1 accepts for $\hat{\beta}_3$.

For the significance threshold $\alpha = 0.05$, the null hypothesis of the coefficients $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ can be rejected ($7.76E - 11$, $9.83E - 16$ and $6.14E - 05$ are less than 0.05).

2.4. Testing classical hypotheses on the regression model.

2.4.1. *Testing the linearity of the proposed model.* ANOVA linear correlation coefficient from Table 2 is: $r_{x_1, x_2, x_3, y} = R^2 = 0.99998161$.

We calculate the correlation ratio $R_{x_1, x_2, x_3, y} = \sqrt{R^2} = 0.999990805$.

Since $r_{x_1, x_2, x_3, y} = R_{x_1, x_2, x_3, y}$ it can be said that the connection between the variables is linear.

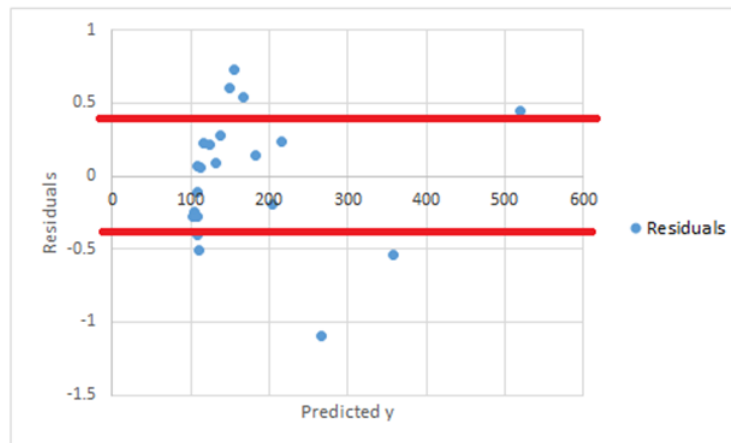
Check the significance of the correlation ratio $R_{x_1, x_2, x_3, y}$. It is significant if $F_{calculat} \geq F_{tabelar}$.

Because $F_{calculat} = 308139.43$ and $F_{tabelar} = F_{0.05; 2; 21-2} = 3.52$, $\implies F_{calculat} > F_{tabelar}$. The correlation ratio $R_{x_1, x_2, x_3, y}$ is significant. Because $R_{x_1, x_2, x_3, y} \neq 0$, $\alpha = 0.05$, the model correctly describes the dependence between The CPI food, The CPI non-food goods and The CPI Services.

2.4.2. *Testing the normality of errors.* The analysis of the quality of the model is facilitated by the graphs built automatically by the regression procedure. Based on the data in Table 5, we can construct the residue diagram versus the independent variable y , as seen in Figure 5.

Table 5. Residual output table from Excel

Observation	Predicted y	Residuals
1	519.6863211	0.443678866
2	357.6140927	-0.544092697
3	266.6367931	-1.096793118
4	216.4613124	0.238687613
5	203.9390456	-0.189045554
6	181.9783953	0.141604703
7	166.5180583	0.541941663
8	156.0473317	0.73266828
9	148.9324224	0.60757759
10	138.3792899	0.280710119
11	131.2277615	0.092238469
12	123.5660088	0.213991229
13	116.7821471	0.22785287
14	113.1726329	0.05736707
15	108.8171044	0.072895645
16	107.8502623	-0.110262325
17	108.7933439	-0.403343904
18	110.5956607	-0.505660701
19	108.9082461	-0.278246053
20	104.0777156	-0.247715627
21	101.9360541	-0.276054138

Figure 5. Graphical representation of residues (u_i) according to (predicted y)

Calculate the interval $[-t_{tab} \cdot SE, t_{tab} \cdot SE]$, where SE = Standard model error (Table 3)

$$t_{tab} = t_{\alpha/2, n-3} = t_{0.25; 18} = 0.688364, \quad SE = 0.47573 \implies \text{The range is } [-0.327475; 0.327475].$$

The points in the Figure 5 can be considered in a horizontal band region which does not contradict the assumptions of normality of errors. The uniform band shape reflects the residue dispersion constancy for the whole range of the dependent variable y . It can be considered that the errors come from a distribution that is normal.

To estimate the error distribution we will apply the Jarque-Bera test.

The Jarque-Bera test considers both the asymmetry and the flattening coefficient and verifies the extent to which the empirical distribution can be approximated by a normal distribution. The null hypothesis of this test assumes that the data sample comes from a normal distribution. The JB statistic has an asymptotic distribution χ^2 .

According to the χ^2 distribution, the critical value of the Jarque-Bera test for a degree of statistical significance of 5% is 7.82. In other words, if the JB statistic calculated for a series of residues is greater than 7.82 we reject the null hypothesis.

This test first calculates the asymmetry coefficient (Skewness) and the vaulting coefficient (Kurtosis) (5) for the residues obtained by LSM. The hypotheses to be tested are:

H_0 : Residues have normal distribution ($S=0$ and $K=3$);

H_1 : Residues do not have a normal distribution.

The linear regression model also contains the free term, then the sum of the residues is zero, i.e. the average is 0 and the Jarque-Bera statistic (4) is:

$$JB = n \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right] \chi_{\alpha, k-1}^2 \quad (4)$$

$$S^2 = \frac{\left(\frac{\sum \hat{u}_i^3}{n} \right)^2}{\left(\frac{\sum \hat{u}_i^2}{n} \right)^3}, \quad K = \frac{\frac{\sum \hat{u}_i^4}{n}}{\left(\frac{\sum \hat{u}_i^2}{n} \right)^2} \quad (5)$$

where: n = number of observations, S = asymmetry coefficient (Skewness), K = vaulting coefficient (Kurtosis).

Calculating these coefficients with the help of Excel explanation (Table 5), we obtained the following values (Table 6):

Table 6. Table with coefficients in Residual output table from Excel

RESIDUAL OUTPUT					
Observation	Predicted Y	Residuals	u_i^2	u_i^3	u_i^4
1	519.6863211	0.443678866	0.196850936	0.087339	0.03875
2	357.6140927	-0.544092697	0.296036863	-0.16107	0.087638
3	266.6367931	-1.096793118	1.202955144	-1.31939	1.447101
4	216.4613124	0.238687613	0.056971777	0.013598	0.003246
5	203.9390456	-0.189045554	0.035738221	-0.00676	0.001277
6	181.9783953	0.141604703	0.020051892	0.002839	0.000402
7	166.5180583	0.541941663	0.293700766	0.159169	0.08626
8	156.0473317	0.73266828	0.536802809	0.393298	0.288157
9	148.9324224	0.60757759	0.369150527	0.224288	0.136272
10	138.3792899	0.280710119	0.078798171	0.022119	0.006209
11	131.2277615	0.092238469	0.008507935	0.000785	7.24E-05
12	123.5660088	0.213991229	0.045792246	0.009799	0.002097
13	116.7821471	0.22785287	0.05191693	0.011829	0.002695
14	113.1726329	0.05736707	0.003290981	0.000189	1.08E-05
15	108.8171044	0.072895645	0.005313775	0.000387	2.82E-05
16	107.8502623	-0.110262325	0.01215778	-0.00134	0.000148
17	108.7933439	-0.403343904	0.162686305	-0.06562	0.026467
18	110.5956607	-0.505660701	0.255692744	-0.12929	0.065379
19	108.9082461	-0.278246053	0.077420866	-0.02154	0.005994
20	104.0777156	-0.247715627	0.061363032	-0.0152	0.003765
21	101.9360541	-0.276054138	0.076205887	-0.02104	0.005807
		Total	3.85	-0.81561	2.207777

$S^2 = 0.245292114$ ($S = 0.202193024$), $K = 3.132120981$, $JB = 0.873796$.

If the Jarque-Bera statistic $JB < \chi_{0.05;3}^2 = 7.82$, it results that the hypothesis of normality of H_0 errors is accepted, i.e. the errors are distributed normally.

2.4.3. *Testing the homoscedasticity hypothesis.* The residual variable is of zero mean $M(\hat{u}) = 0$, and its dispersion S_u^2 is constant and independent of x - the homoscedasticity hypothesis, based on which it can be admitted that the connection between the variables x and y is stable. The acceptance of the homoscedasticity hypothesis can be done using the graphical method and the analysis of the variation through Anova. The graphical method (the correlation between the factorial variable x and the residual variable u) can be observed in Figure 6, Figure 7, Figure 8.

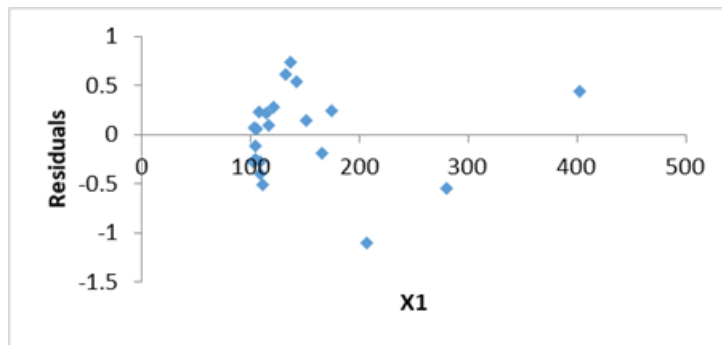


Figure 6. Residue diagram – variable x_1

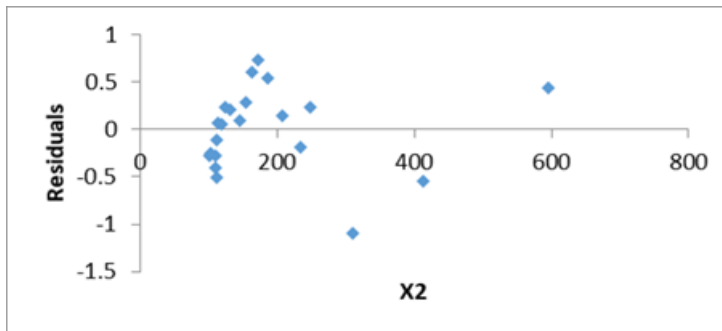


Figure 7. Residue diagram – variable x_2

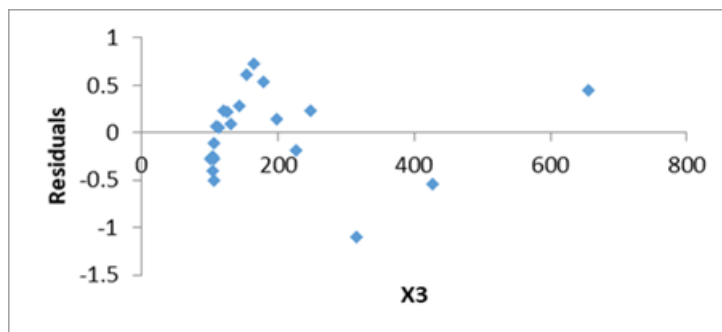


Figure 8. Residue diagram – variable x_3

It can be seen that the graph of empirical points shows an unordered (oscillating) distribution. Thus, it can be accepted that the three variables are independent and not correlated. The graph falls into a band, then the errors are homoscedastic.

Acceptance or rejection of the homoscedasticity hypothesis can be done with the help of variation analysis (Anova). (Table 2)

Significance $F = 1.93E - 40$ (very low value, close to 0) < 0.05 . The model is plausible.

2.4.4. *Testing the hypothesis of autocorrelation of errors.* From Table 5 Residual Output we take the estimated errors (residuals) and we will apply Durbin Watson (DW) statistics. (Table 7)

Statistical hypotheses:

H_0 : errors are not autocorrelated ($\rho = 0$)

H_1 : errors are autocorrelated ($\rho \neq 0$)

Table 7. Durbin Watson statistics (DW)

Observation	Predicted Y	Residuals	ui-1	ui^2
1	519.6863211	0.443678866		0.196850936
2	357.6140927	-0.544092697	0.443679	0.296036863
3	266.6367931	-1.096793118	-0.54409	1.202955144
4	216.4613124	0.238687613	-1.09679	0.056971777
5	203.9390456	-0.189045554	0.238688	0.035738221
6	181.9783953	0.141604703	-0.18905	0.020051892
7	166.5180583	0.541941663	0.141605	0.293700766
8	156.0473317	0.73266828	0.541942	0.536802809
9	148.9324224	0.60757759	0.732668	0.369150527
10	138.3792899	0.280710119	0.607578	0.078798171
11	131.2277615	0.092238469	0.28071	0.008507935
12	123.5660088	0.213991229	0.092238	0.045792246
13	116.7821471	0.22785287	0.213991	0.05191693
14	113.1726329	0.05736707	0.227853	0.003290981
15	108.8171044	0.072895645	0.057367	0.005313775
16	107.8502623	-0.110262325	0.072896	0.01215778
17	108.7933439	-0.403343904	-0.11026	0.162686305
18	110.5956607	-0.505660701	-0.40334	0.255692744
19	108.9082461	-0.278246053	-0.50566	0.077420866
20	104.0777156	-0.247715627	-0.27825	0.061363032
21	101.9360541	-0.276054138	-0.24772	0.076205887
		Total		3.85

The Durbin-Watson test (d) consists in calculating the empirical term (6):

$$d = \frac{\sum_{i=2}^{21} (u_i - u_{i-1})^2}{\sum_{i=1}^{21} u_i^2} \quad (6)$$

The value of the Durbin-Watson test is $d = 1.023887$.

Comparison of this quantity d with two theoretical values d_1 and d_2 , taken from the Durbin-Watson tables according to an arbitrarily chosen significance threshold α , the number of exogenous variables (k) and the observed values n ($n \geq 15$). The values read from the table ($\alpha = 0.05$, $n = 21$, $k = 3$) for d_1 and d_2 :

$d_1 = 1.026$ and $d_2 = 1.669$.

The acceptance or rejection of the hypothesis of independence of errors is based on a certain rule, which consists in:

$0 < d < d_1 \iff 0 < d < 1.026 \implies$ positive autocorrelation;

$d_1 \leq d \leq d_2 \iff 1.026 \leq d \leq 1.669 \implies$ indecision;

$d_2 < d < 4 - d_2 \iff 1.669 < d < 2.331 \implies$ errors are independent;

$4 - d_2 < d < 4 - d_1 \iff 2.331 < d < 2.974 \implies$ indecision;

$4 - d_1 < d < 4 \iff 2.971 < d < 4 \implies$ negative autocorrelation;

It is observed that: positive autocorrelation.

It can be concluded that there is a positive autocorrelation of the errors, and the H_0 hypothesis is rejected.

The correlation coefficient of order 1 is given by (7):

$$r_1 = \frac{\sum_{i=2}^{21} u_i \cdot u_{i-1}}{\sum_{i=2}^{21} u_i^2} \quad (7)$$

The value of the correlation coefficient of order 1 is $r_1 = 0.98$. The data can be found in Excel, in the Regression Sheet. (Table 7)

There is a relationship between the degree 1 correlation coefficient and the Durbin-Watson variable (8):

$$d = 2(1 - r_1) \implies r_1 = 1 - \frac{d}{2} \quad (8)$$

Knowing:

$r_1 = -1 \implies$ strictly negative autocorrelation $d = 4$

$r_1 \in (-1, 0) \implies$ indecision $d \in (2, 4)$

$r_1 = 0 \implies$ independence $d = 2$

$r_1 \in (0, 1) \implies$ indecision $d \in (0, 2)$

$r_1 = 1 \implies$ strictly positive autocorrelation $d = 0$

$r_1 = 0.98 \in (0, 1) \implies d = 1.024 \in (0, 2) \implies$ indecision.

In this case, it is not known for sure whether the hypothesis of error independence is accepted or rejected. This result does not fall within any of the decision intervals, so no decision can be made regarding the correlation of errors. It can be concluded that the errors are not autocorrelated, i.e. u_i does not depend on u_{i-1} .

2.5. Prediction of the value of the variable y in the hypothesis of modification of exogenous variables. Prediction of the value of the variable y if the variable x increases by 10% compared to the last recorded value (including confidence interval) for all known variants.

The last records made for this data series are for 2019, which have the value: The CPI food (x_1) with 102.53% of the Total CPI, The CPI non-food goods (x_2) with 101.05% of the Total CPI and The CPI Services (x_3) with 101.79% of total CPI. It is assumed that by 2020 they will increase by 10% compared to the previous value.

So, the values of x_1 , x_2 and x_3 in 2020 will be:

$$102.53 \cdot 10\% = 10.253 \implies 102.53 + 10.253 = 112.783 = x_1$$

$$101.05 \cdot 10\% = 10.105 \implies 101.05 + 10.105 = 111.155 = x_2$$

$$101.79 \cdot 10\% = 10.179 \implies 101.79 + 10.179 = 111.969 = x_3$$

Equation (4)

$$\hat{y} = 1.1726454 + 0.4094779x_1 + 0.4481466x_2 + 0.1325711x_3$$

becomes:

$$\hat{y} = 1.1726454 + 0.4094779 \cdot 112.783 + 0.4481466 \cdot 111.155 + 0.1325711 \cdot 111.969 = 110.66$$

Therefore, a 10% increase in the indices of consumer prices of foodstuffs, non-food goods and services, will lead to an increase in the index of total consumer prices by 8.85%.

3. CONCLUSION

The estimated multiple regression model proved to be accurate - it has a high coefficient of determination $R^2 = 0.99998161$, i.e. Total CPI (y) is explained to the extent of almost 100% by the independent variables included in the model. In addition, the least squares method (LSM) assumptions are perfectly verifiable - the errors are homoscedastic, not autocorrelated, and the variables are not collinear. The value of the F - test is high enough to determine the overall validity of the model for a significance threshold of at least $Significance\ F = 1.93E - 40$, much lower than the chosen α .

REFERENCES

- [1] Anghelache, G.V., Anghelache, C., Prodan, L., Dumitrescu, D., Soare, D.V., Theoretical elements regarding the use of the model econometric of multifactorial regression, *Romanian Statistical Review Supplement, National Institute of Statistics*, 60 (2012), 3, 221-231.
- [2] Book, E., Ekelöf, L., *A Multiple Linear Regression Model To Assess The Effects of Macroeconomic Factors On Small and Medium-Sized Enterprises*, Degree work in technology, Royal Institute of Technology School of Engineering Sciences KTH SCI SE-100 44 Stockholm, Sweden, 2019.
- [3] Cao, T., *Paradox of Inflation: The Study on Correlation between Money Supply and Inflation in New Era*, Arizona State University, 2015.
- [4] Cioran, Z., Monetary Policy, Inflation and the Causal Relation between the Inflation Rate and Some of the Macroeconomic Variables, 21st International Economic Conference, *Procedia Economics and Finance* 16, 2014, 391-401.
- [5] Cogoljević, D., Gavrilović, M., Roganović, M., Matić, I., Piljan, I., Analyzing of consumer price index influence on inflation by multiple linear regression, *Physica A: Statistical Mechanics and its Applications*, Volume 505, Issue C, 2018, 941-944.
- [6] Cruceru, D., Anghel, M. G., Diaconu, A., The multiple Linear regression used to analyse the correlation between variables, *Romanian Statistical Review Supplement, National Institute of Statistics*, 64 (2016), 10, 114-117.
- [7] DhamoEralda, G., Llukan, P., Začaj, O., Forecasting consumer Price index (CPI) using time series models (Albania case study), *Proc. of the 10th International Scientific Conference "Business and Management 2018"*, Vilnius, LITHUANIA Section: Financial Engineering, 2018, 466-472.
- [8] Falnita, E., Sipos, C., A multiple regression model for inflation rate in Romania in the enlarged EU, *Munich Personal RePEc Archive*, Paper No. 11473, 2007.
- [9] Iyiola, R.O., Adetunji, A.A., The effect of monetary policy on consumer price index, *International Journal of Innovative Science, Engineering and Technology*, 1 (2014), 6, 140-146.
- [10] Popescu, A., Correlation Analysis of Consumer Price Indices by Multiple Regression, *Transylvanian Journal of Mathematics and Mechanics*, 9 (2017), 1, 43-50.

UNIVERSITY OF PETROȘANI
 DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
 UNIVERSITĂȚII 20, 332006, PETROȘANI, ROMÂNIA
 Email address: am.popescu@yahoo.com