COMPLETE ECONOMETRIC ANALYSIS OF THE CORRELATION BETWEEN THE TOTAL REAL AVERAGE MONTHLY INCOMES AND EXPENSES OF A ROMANIAN HOUSEHOLD

NADIA ELENA STOICUȚA [*]

ABSTRACT: *It is known that the monthly expenses of a family in Romania are dependent on their income, starting from the fact that the expenses can be much better managed if we take into account the level of income that family has. As expected, according to a statistic made by the European Union, in 2021, Romania has the lowest cost of living among EU countries. On the other hand, in 2021, the annual inflation in Romania reached the highest level in the last 10 years, of almost 8%. Also, the large increases in gas and electricity prices have led to a decrease in the living standards of Romanians.This econometric study substantiates the connection between the average total monthly expenses, and the total average monthly incomes of a Romanian household, in the period 1997-2020. Based on the linear regression model, forecasts will be made in the end.*

KEY WORDS: *econometric analysis, correlation, estimation, Eviews, least squares method, forecast.*

JEL CLASSIFICATIONS: *C12, C13, C53, C87.*

1. INTRODUCTION

This study regarding the effect of the average total monthly income on the total average monthly expenses is a very useful one in the economic-financial analysis of the planning of the budgets of the Romanian households. In this sense, in order to optimize the budgetary planning of the Romanian households, they can predict based on the validated model, the level of expenditures according to the revenues they are going to realize.

According to data published by the National Institute of Statistics, the average total monthly income of a Romanian household in the second quarter of 2021 was in

[*] *Lecturer, Ph.D., University of Petrosani, Romania, stoicuta_nadia@yahoo.com*

nominal terms of 5573 lei, and the average total monthly consumption of a Romanian household in the same period was 4709 lei. In addition to wage income, which accounts for a considerable percentage of almost 70%, and income from social benefits (19,3%), income in kind (6,1%), income from agriculture (1%) contributed to the formation of total household income (1,8%), incomes from independent non - agricultural activities (1,6%), as well as those from property and sale of assets from the patrimony of the household (1,1%) (https://insse.ro).

Regarding the total expenses, according to the same press release of the National Institute of Statistics, over 60% of Romanians' income goes to consumption-related expenses, of which a percentage of approximately 35% is consumption on food and non-alcoholic beverages. A considerable percentage of over 33% of total household expenditure is tax and contribution, and only 0,5% goes to investment expenditure (https://insse.ro).

2. LITERATURE REVIEW

Recent studies on the correlation between monthly income and monthly expenditure in Romania were conducted by Babucea and Bălănescu. The authors of the article make a statistical analysis on the total income and expenditure of households in Romania by components, taking into account the consumer price index and the real earnings index, in the period 1990-2010 (Babucea & Bălănescu, 2001). Anghelache studies in his paper, the correlation between the monthly income and expenditure of the population, based on relevant indicators, using a linear econometric model (Anghelache, et. al., 2016). Țițan analyzes the interdependencies between the social insurance budget and the main macroeconomic indicators that characterize the Romanian economy (such as gross national product, average monthly income, total expenditures, unemployment, etc.) in the period 2000-2009 (Țițan, et. al., 2011).

3. ECONOMETRIC ANALYSIS

In this paragraph, is realised the econometric analysis performed between the average total monthly average real expenditures and the total average monthly real incomes of the Romanian households, for a period of 24 years (1997-2020). Thus, in table 1 are introduced the data series of the average total incomes and of the average total expenses of the Romanian households, in the analyzed period. The data were collected after consulting the website of the National Institute of Statistics (http://statistici.insse.ro) and that of the World Bank (https://data.worldbank.org).

Both data series were calculated according to the calculation methodology of the European System of Accounts of the Integrated Economy - SEC 2010. Both the average total monthly income and the average total monthly expenses of a Romanian household were deflated using the Consumer Price Index (2010 = 100) (https://data.worldbank.org).

The graph of the dependence between the average total monthly expenses $Y=(y_i)_{i=\overline{1,24}}$ and total average monthly incomes $X=(x_i)_{i=\overline{1,24}}$, of the households in

Romania (data diagram), in the period 1997-2020, is represented in Figure 1 (graph made in Eviews1.10).

Table 1. The data series of the two variables

Years	**Average monthly nominal total income [RON]**	**Average nominal total monthly expenses [RON]**	**CPI [%] 2010=100**	**Average total real monthly income [RON]**	**Average real monthly total expenses [RON]**
1997	144.38	144.38	9.464965	1525.415	1525.415
1998	215.48	215.48	15.05844	1430.958	1430.958
1999	280.88	280.88	21.95577	1279.299	1279.299
2000	388.82	388.82	31.98222	1215.738	1215.738
2001	521.79	516.52	43.00874	1213.219	1200.965
2002	658.51	651.65	52.70286	1249.477	1236.460
2003	795.08	781.44	60.75242	1308.722	1286.270
2004	1085.79	1049.94	67.96639	1597.540	1544.793
2005	1212.18	1149.33	74.0935	1636.014	1551.189
2006	1386.32	1304.66	78.95293	1755.882	1652.453
2007	1686.74	1541.96	82.77214	2037.811	1862.897
2008	2131.67	1915.19	89.27042	2387.879	2145.380
2009	2315.99	2047.33	94.25833	2457.067	2172.041
2010	2304.28	2062.95	100	2304.280	2062.950
2011	2417.26	2183.76	105.7893	2284.976	2064.254
2012	2475.04	2244.47	109.3172	2264.090	2053.172
2013	2559.05	2317.4	113.6732	2251.234	2038.651
2014	2500.72	2269.25	114.8876	2176.667	1975.191
2015	2686.77	2366.25	114.205	2352.585	2071.932
2016	2944.6	2559.25	112.4408	2618.800	2276.087
2017	3391.67	2874.14	113.9464	2976.549	2522.361
2018	4251.26	3666.59	119.2169	3565.988	3075.562
2019	4789.83	4091.83	123.7804	3869.619	3305.717
2020	5216.38	4371.86	127.037	4106.190	3441.407

Following the distribution of the pairs in the plane $(x_i, y_i), i=\overline{1,24}$, it is observed that they can be approximated by a line. Therefore, we can say that the econometric model that describes the connection between the two variables is a linear regression model, of the form:

$$y_i = ax_i + b + \varepsilon_i, i = \overline{1,24} \tag{1}$$

where a and b are the parameters of the model, and ε is the residual variable.

From the graphical representation, it is observed that the parameter a (coefficient of the slope of the regression line) must be greater than zero, which confirms the hypothesis from economic theory on a direct link between the two variables: revenue growth attracts expenditure growth. To estimate the parameters of the linear regression model a and b we will use the least squares method.

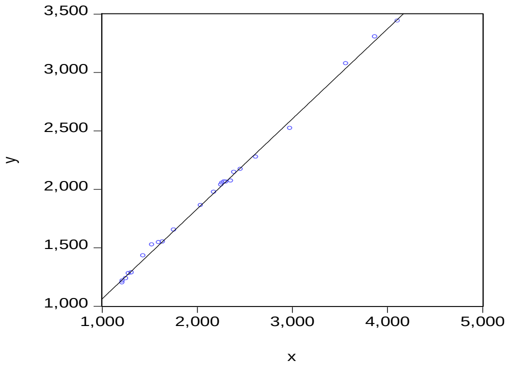


Figure 1. The graph of the connection between the expenses and incomes of a Romanian household in the period 1997-2020 with the highlighting of the regression line

Taking the above into account, the following linear regression model is obtained:

$$\hat{y}_i = 0,770050 x_i + 293,8228 \quad i = \overline{1,24} \tag{2}$$

Table 2. Table with coefficient values

Method: Least Squares (Gauss-Newton / Marquardt steps)				
Included observations: 24				
Y=a*X+b				
	Coefficient	Std. Error	t-Statistic	Prob.
a	0.770050	0.006884	111.8655	0.0000
b	293.8228	15.87893	18.50394	0.0000
R-squared	0.998245	Mean dependent var		1957.964
Adjusted R-squared	0.998165	S.D. dependent var		635.1174
S.E. of regression	27.20448	Akaike info criterion		9.524296
Sum squared resid	16281.85	Schwarz criterion		9.622467
Log likelihood	-112.2916	Hannan-Quinn criter.		9.550341
F-statistic	12513.88	Durbin-Watson stat		1.215557
Prob(F-statistic)	0.000000			

In the Figure 2 shows the real curve (empty circles) in tandem with the approximate curve (solid circles) by the least squares method of the linear regression model. As can be seen, the two graphs are very close, which means that the linear regression model approximates very well the data series of the two variables involved in the model.

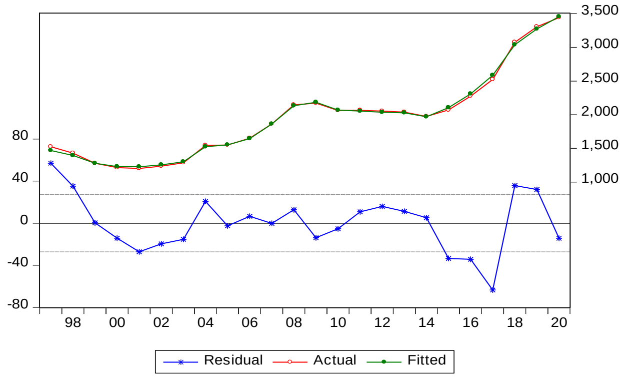


Figure 2. Graph of the real curve (empty circles), in tandem with the graph of the approximate straight line of the linear regression model (solid circles) with the highlighting of the residue (stars)

The correlation coefficient has the role of determining the meaning of the dependence between the variables x and y of the regression model, as well as the intensity of the linear connection between the two variables. The calculation relation of the correlation coefficient is the following:

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n \cdot \sigma_x \cdot \sigma_y} = 0{,}957 \qquad (3)$$

Since the value of the correlation coefficient is very close to 1, we can say that between the two variables x and y of the regression model, there is a direct, very strong linear dependence. The following table calculates the values of the main descriptive indicators of the two variables *x* and *y*.

Table 3. The table with the values of the descriptive indicators

	X	**Y**
Mean	2161.083	1957.964
Median	2213.951	2006.921
Maximum	4106.190	3441.407
Minimum	1213.219	1200.965
Std. Dev.	824.0506	635.1174
Sum	51866.00	46991.14
Sum Sq. Dev.	15618365	9277605.
Observations	24	24

The determination ratio is calculated using the relation:

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = 0,998245 \tag{4}$$

As the determination ratio is $R^2 = 0,998$, it results that the intensity of the connection is very strong, respectively that 99,8% of the variation of the total average expenses is explained by the variation of the average registered total incomes. In fact, the R-squared statistic measures the "success" with which the estimated regression equation manages to explain the value of the dependent variable in the sample.

The adjusted determination ratio is calculated using the relation:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2) = 1 - \frac{23}{22}(1-0,998245) = 0,998165 \tag{5}$$

where *k* represents the number of input variables in the model. *Adjusted R-squared* statistics are an alternative, which has the advantage of "penalizing" the addition of regressors that do not contribute to the explanatory power of the model.

3.1. Testing the significance of estimators

The Student's test is applied to test the significance of the linear regression model. Thus, the estimators of the parameters of the linear regression model will have to be significantly different from zero. To test whether the slope parameter of the regression line differs significantly from zero, the inequality is checked $t_{\hat{a}} = \left|\frac{\hat{a}}{\sigma_{\hat{a}}}\right| = 111,8655 > t_{\alpha,n-2} = t_{0,05;22} = 1,717$. For a threshold of significance $\alpha = 5\%$ and $n = 24$ data, $\sigma_{\hat{a}}$ the mean square deviation of the parameter estimator $\hat{a}$ is calculated by the relation:

$$\sigma_{\hat{a}} = \sqrt{\frac{\sigma_{\hat{\varepsilon}}^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} = 0,006884 \tag{6}$$

where $\sigma_{\hat{\varepsilon}}^2$ is the variance (dispersion) of the residual variable estimator:

$$\sigma_{\hat{\varepsilon}} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-k-1}}, k = 1 \tag{7}$$

As this inequality is verified, we can say that the estimator of the parameter of the slope of the regression line differs significantly from zero.

Regarding the free term estimator, the Student test statistic is defined based on the relationship $t_{\hat{b}} = \left| \frac{\hat{b}}{\sigma_{\hat{b}}} \right| = 18,50394 > t_{0,05,22} = 1,717$, the mean square deviation of the parameter estimator $\hat{b}$ is calculated by the relation:

$$\sigma_{\hat{b}}^2 = \sigma_{\hat{\varepsilon}}^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n} \left(x_i - \overline{x} \right)^2} \right] = 252,14039 \tag{8}$$

As this inequality is verified we can say that the estimator of the free term parameter differs significantly from zero. Regarding the probabilities associated with the two estimators $\hat{a}$, respectively $\hat{b}$ of the parameters, it can be said that a value as close to zero as possible will indicate a high significance of the parameters, otherwise, this confirming, together with the *t* test, that the respective parameters they are insignificant.

3.2. Verification of hypotheses specific to the linear regression model

A number of hypotheses are considered in the definition of linear regression. These are important in estimating and determining the properties of the linear regression model. The hypotheses presented below refer to the two variables that define the linear regression model, but also to the residual variable.

Hypothesis 1. Data series are not affected by measurement errors. It is considered that the values for the two variables are not affected by significant measurement errors that would distort the quality of the estimators of the parameters of the linear regression model. However, in order to verify this hypothesis, the so-called"rule of three sigma" applies, ie the variables *x* and *y* must verify the intervals:

$$\begin{matrix} x \in \left(\overline{x} \pm 3\sigma_x \right) \\ y \in \left(\overline{y} \pm 3\sigma_y \right) \end{matrix} \quad \text{or} \quad \begin{matrix} -311,068 < x_i < 4633,235 \\ 52,612 < y_i < 3863,316 \end{matrix} \tag{9}$$

As can be seen, each value of the input variable *x* verifies the double inequality above. The same can be said for each value of the variable *y*.

Hypothesis 2. The residual variable is zero mean. As can be seen in the figure below, the residues estimated by the least squares method are small enough, with the average of the errors having a very small value, which tends to zero. The values of the residual variable in each node are represented in the figure below.

$$\overline{\varepsilon} = \frac{\sum_{i=1}^{n} \varepsilon_i}{n} = \frac{\sum_{i=1}^{24} \varepsilon_i}{24} = \frac{1,346 \cdot 10^{12}}{24} = 5,61 \cdot 10^{14} \tag{10}$$

Hypothesis 3. The residue variance is constant (homoskedasticity hypothesis). Homoscedasticity describes the situation where the error is the same for all values of the output variable *y*. This hypothesis is verified either by the graphical method or by using a series of statistical tests. Since the point graph in the plan shows an oscillating distribution, this hypothesis can be accepted. If this property was not satisfied, then the model is said to be ***heteroskedastic***. In order to verify the homoscedasticity of the residual variable, in addition to the graphical method, the White test is applied. This test consists of checking ***Fisher statistics*** or checking LM (***Lagrange Multiplier***) statistics.

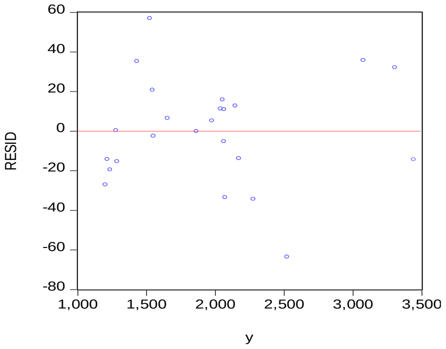


Figure 4. Graphic verification of the homoskedasticity hypothesis

Table 4. The White test generated in Eviews

Dependent Variable: RESID^2				
Method: Least Squares				
Included observations: 24				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	75.10590	1701.365	0.044144	0.9652
(X)^2	-2.17E-05	0.000286	-0.075894	0.9402
X	0.332689	1.464071	0.227235	0.8224
R-squared	0.032965	Mean dependent var		678.4103
Adjusted R-squared	-0.059133	S.D. dependent var		1018.086
S.E. of regression	1047.755	Akaike info criterion		16.86316
Sum squared resid	23053598	Schwarz criterion		17.01041
Log likelihood	-199.3579	Hannan-Quinn criter.		16.90222
F-statistic	0.357937	Durbin-Watson stat		1.044165
Prob(F-statistic)	0.703300			

The White test verified by Fisher statistic (statistics F) is calculated using the relation:

$$F_{calculat} = \frac{R^2}{1-R^2}(n-2) = 0,3579 \tag{11}$$

The calculated value of this statistic is compared to the tabular value of the statistic $F_{\alpha,k,n-k-1} = F_{0,05;1;22} = 4,747$, for a significance threshold of 5%, and *k* represents the number of output variables in the model. If $F_{calculat} = 0{,}357937 < F_{0{,}05;1;22} = 4{,}301$, then the regression model is correctly specified, ie the homoskedasticity hypothesis is verified (the dispersion is the same for all values of *x*). ***White test verified by LM statistics*** whose calculation formula is:

$$LM = n \cdot R^2 = 24 \cdot 0{,}032965 = 0{,}79116 \qquad (12)$$

The calculated value of this statistic is compared to the tabular value of the statistic $\chi^2_{2,\alpha} = \chi^2_{2;0{,}05} = 12{,}338$, for a significance threshold of 5%. If $LM = 0{,}79116 < \chi^2_{2;0{,}05} = 12{,}338$, then the regression model is correctly specified, ie the homoskedasticity hypothesis is verified. In addition to the White test, the Goldfeld-Quandt test or the Glesjer test can also be applied.

Hypothesis 4. Residues are not correlated, ie there is no phenomenon of autocorrelation of residues. Basically, this hypothesis verifies that the values of the residual variable are independent of each other. This hypothesis can be tested with the help of the relation:

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad \forall i \neq j, \quad i < j \qquad (13)$$

If this assumption is not met, then we say that the errors are self-correlated.

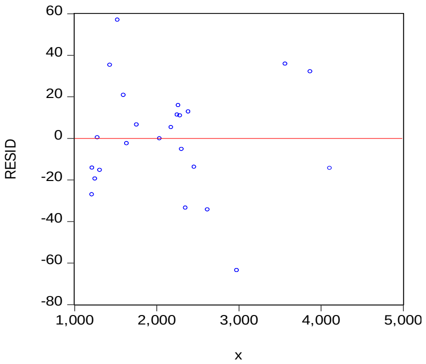


Figure 5. Graphic method for verifying the residue autocorrelation hypothesis

Since the point graph shows an oscillating evolution in all nodes, we can accept that this hypothesis is verified. To detect residual autocorrelation, certain statistical tests (Durbin-Watson test or Breush-Godfrey test) are applied.

The Durbin-Watson test is the most widely used analysis of residue autocorrelation. This test detects first-order autocorrelation. The test statistic are defined by the relation:

$$DW = \frac{\sum_{i=2}^{n} \left(\varepsilon_i - \varepsilon_{i-1}\right)^2}{\sum_{i=1}^{n} \varepsilon_i^2} = \frac{19791{,}5184}{16281{,}8461} = 1{,}215557 \qquad (14)$$

The value of the Durbin-Watson statistic can also be found in the last line of Table 3. On the other hand, the calculated value of this test is compared with the tabulated value of the Durbin-Watson statistic, for a significance threshold $\alpha = 1\%$. Specifically, the value of the DW statistic is compared to two tabelated values $d_1 = 1,035$ and $d_2 = 1,119$ from the Durbin-Watson table, for a number of input variables $k = 1$ and $n = 24$ date.

Accepting or not accepting the hypothesis of autocorrelation of errors implies the verification of the following inequalities:

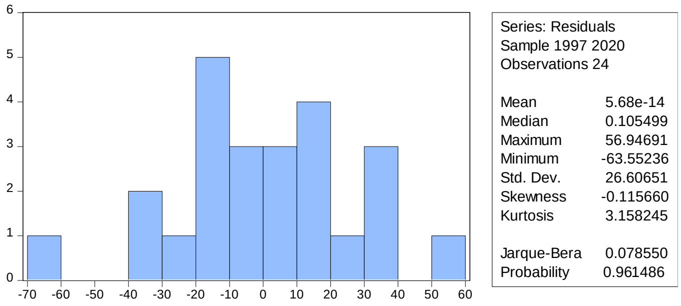
- ✓ $0 < DW < d_1$ there is a significant positive linear correlation;
- ✓ $d_1 < DW < d_2$ the test does not provide a relevant answer regarding the correlation of the residue and it is recommended to apply another statistical test;
- ✓ $d_2 < DW < 4 - d_2$ there is no significant first-order linear correlation at the residue level, ie they are independent;
- ✓ $4 - d_2 < DW < 4 - d_1$ the test does not provide a relevant answer regarding the correlation of the residue and it is recommended to apply another statistical test;
- ✓ $4 - d_1 < DW < 4$ there is a significant negative linear correlation;

In this case, the double inequality $d_2 = 1,119 < DW = 1,215557 < 4 - d_2 = 2,801$ shows us that the residues are not correlated, ie this hypothesis is verified.

Hypothesis 5. The residual variable is normally distributed. In this hypothesis, the residual variable must have a normal distribution, more precisely it must be of zero mean and dispersion σ_ε^2, meaning $\varepsilon \to N\left(0, \sigma_\varepsilon^2\right)$.

Figure 6 shows the histogram of the residual variable. For the residue series, two indicators used in the descriptive statistics are used to analyze the asymmetry and flattening of the residue series. This hypothesis is verified either by the asymmetry and vaulting coefficient or by the Jarque-Bera test.

The asymmetry coefficient (skewness) is denoted by *S*, and ***the coefficient of vaulting or flattening (kurtosis)*** is denoted by *K*.



```
Series: Residuals
Sample 1997 2020
Observations 24

Mean          5.68e-14
Median        0.105499
Maximum       56.94691
Minimum      -63.55236
Std. Dev.     26.60651
Skewness     -0.115660
Kurtosis      3.158245

Jarque-Bera   0.078550
Probability   0.961486
```

Figure 6. Histogram and characteristics of the estimated residue

The value of the asymmetry coefficient shows us that the residue distribution curve has a slightly voluminous “tail” on the left. The value of this indicator should be as close to zero as possible.

The value of the flattening coefficient must be equal to 3. If the value of the flattening coefficient is less than 3, then the residue distribution is ***platikurtoric***, and if the value of the flattening coefficient is greater than 3, then the residue distribution is ***lipokurtoric***. This value of the asymmetry coefficient shows us that the residue distribution curve is very easily lipocurtoric.

The Jarque-Bera (JB) test is calculated based on the following statistic:

$$JB=\frac{n}{6}\cdot S^2+\frac{n}{24}\cdot\left(K-3\right)^2=\frac{24}{6}\cdot(-0{,}115660)^2+\frac{24}{24}\cdot\left(3{,}158245-3\right)^2=0{,}078549883 \qquad (15)$$

The calculated value of this test is compared to the tabulated value of the statistic $\chi^2_{2,\alpha}$, for a significance threshold of $\alpha=5\%$, namely $\chi^2_{2;0,05}=12,338$. If $JB=0,078<\chi^2_{2;0,05}=12,338$ then the residue normalization hypothesis is accepted.

4. TESTING THE VALIDITY (LIKELIHOOD) OF THE REGRESSION MODEL

Variance analysis is applied to test the validity and quality of the linear regression model. For this, Table 5 is compiled. The calculated value of Fisher-Snedecor statistics (which can also be found in Table 2), is compared with the tabular value of the statistic $F_{\alpha,k-1,n-k-1}=F_{0,05;1;22}=4,301$, for a significance threshold of 5%.

If $F=12513,88299>F_{0,05;1;22}=4,301$, then the regression model is correctly specified, ie the linear regression model is a valid one, ie it is a correctly estimated one. The high value of the F statistic shows that, through the input variable, ie the average total monthly income, the dynamics of the average total monthly expenses in Romania for the analyzed period are appreciated to a large extent.

Table 5. Analysis of variance

Source of variance	The sum of the squares of the deviations	Degrees of freedom	Variance	F - Statistic
Dependent variable	$SPE=\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2=$ $=9261323,51440954$	*2-1*	*SPE/1*	$F_{calculat}=\frac{SPE/1}{SPR/(n-2)}=$ $=\frac{9261323,514409}{9277605,3605/22}=$ $=12513,88$
Residual variable	$SPR=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=$ $=9277605,3605325$	*n-2*	*SPR/(n-2)*	
Total variance	$SPT=\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2=$ $=9277605,3605325$	*n-1*	-	

5. CALCULATION OF THE LOGARITHM OF THE LIKELIHOOD FUNCTION AND THE CRITERIA BASED ON INFORMATION THEORY

The logarithm of the maximum likelihood function, is a function that is determined taking into account the estimated values of the two parameters and is calculated using the relation:

$$L=-\frac{n}{2}\left(1+\ln(2\pi)+\ln\frac{\sum_{i=1}^{n}\varepsilon_i^2}{n}\right)=-\frac{24}{2}\left(1+\ln(2\pi)+\ln\frac{16281,846123}{24}\right)=-112,2916 \qquad (16)$$

This indicator is used to compile statistical tests for omitted variables in an econometric model, as well as tests for redundant variables in an econometric model, such as the LR test or Likelihood Ratio.

The three criteria based on information theory are CA - ***Akaike criterion***, CS - ***Schwartz criterion*** and CHQ - ***Hannan-Quin criterion***. These criteria are calculated using the following relations (Andrei & Bourbonnais, 2008):

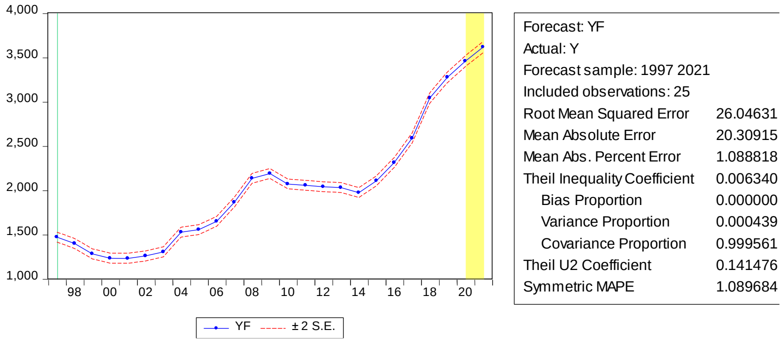
$$\begin{aligned}
CA&=-\frac{2L}{n}+\frac{2k}{n}=-\frac{2\cdot(-112,2916)}{24}+\frac{2\cdot 2}{24}=9,524296\\
CS&=-\frac{2L}{n}+\frac{k\cdot\ln(n)}{n}=\frac{2\cdot(-112,2916)}{24}+\frac{2\cdot\ln(24)}{24}=9,622467\\
HQ&=-2\frac{L}{n}+\frac{2\cdot k\cdot\ln(\ln(n))}{n}=-2\frac{(-112,2916)}{24}+\frac{2\cdot 2\cdot\ln(\ln(24))}{24}=9,550341
\end{aligned} \qquad (17)$$

The closer the value of the three indicators is to zero, the better the regression model chosen. As can be seen, the values of the three indicators are small enough to say that the linear regression model is one that gives very good results.

5. FORECAST BY CONFIDENCE INTERVAL

In this paragraph we will make a forecast for the next year, 2021. To make forecasts we must first make sure that the model parameters remain unchanged and for the period for which the forecast is made, ie at the level of the evolution of the analyzed phenomenon no special phenomena occurred . We will consider that the level of the average total real monthly income in 2021 will increase by 5% compared to the previous year, ie they will increase by 205,3 lei reaching 4312,2 lei. To make the prediction, we will use the mathematical model that describes the linear dependence between the two variables under analysis given by the relation (2).

The following figure (made in Eviews10.1) shows the graph of the evolution of the average total real monthly monthly expenses, with the highlighting of the value predicted in 2021.



```
Forecast: YF
Actual: Y
Forecast sample: 1997 2021
Included observations: 25
Root Mean Squared Error        26.04631
Mean Absolute Error            20.30915
Mean Abs. Percent Error        1.088818
Theil Inequality Coefficient   0.006340
     Bias Proportion           0.000000
     Variance Proportion       0.000439
     Covariance Proportion     0.999561
Theil U2 Coefficient           0.141476
Symmetric MAPE                 1.089684
```

Figure 7. The evolution over time of the predicted values

The predicted value of real expenditures on a household in Romania, in average values, in 2021 is $\hat{y}_{25} = 3614,437$ lei. The forecast value for 2021 shows that the level of average monthly total expenditures will also increase by about 5% compared to the previous year. In determining the prediction by the confidence interval, we will define in this case the following prediction interval, for a significance threshold $\alpha = 5\%$ and $t_{\alpha,n-k-1} = t_{0,05;22} = 1,321$to check if the value obtained from the average total monthly expenses per household can be real. So the confidence interval, for a level of the average total monthly income on a Romanian household of RON 4312,209, is $3572,868725 \le y_{25} \le 3656,00527$.

6. STATISTICAL MEASURES FOR ASSESSING THE QUALITY OF THE FORECAST

In order to be able to realize the quality of the forecast made in the previous paragraph we will calculate the following statistical indicators. The values of these indicators can also be found in the attached table Figure 7. These are the Root mean square deviation (σ_e), mean absolute error ($\bar{d}_e$), the mean of the absolute procent errors (c) and the Theil coefficient (CT).

The values of the indicators are determined as follows:

$$\sigma_e = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} = \sqrt{\frac{1}{24}\sum_{i=1}^{24}\left(\hat{y}_i - y_i\right)^2} = \sqrt{\frac{1}{24}\cdot 16281,84612} = 26,04631$$

$$\bar{d}_e = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| = \frac{1}{24}\sum_{i=1}^{24}\left|\hat{y}_i - y_i\right| = \frac{1}{24}\cdot 487,4194893 = 20,30915$$

$$c = \frac{1}{n}\cdot 100\cdot\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| = \frac{1}{24}\cdot 100\cdot\sum_{i=1}^{24}\left|\frac{\hat{y}_i - y_i}{y_i}\right| = 1,088818302$$

$$CT=\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i-y_i\right)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i^2}+\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_i^2}}=\frac{\sqrt{\frac{1}{24}\sum_{i=1}^{24}\left(\hat{y}_i-y_i\right)^2}}{\sqrt{\frac{1}{24}\sum_{i=1}^{24}\hat{y}_i^2}+\sqrt{\frac{1}{24}\sum_{i=1}^{24}y_i^2}}=0,00634 \qquad (18)$$

Theil coefficient takes values between $\left[0,1\right]$. The closer the values of the five statistical measures are to zero, the better the quality of the predictions made by the model.

7. CONCLUSION

Following the analysis of the dependence between the average real monthly average income and the total average monthly real expenditures in Romania, it was shown that the dependence between the two variables is a direct linear one. The statistical results obtained show that there is a strong link between the two variables analyzed. It was also observed that the monthly expenses of a family in Romania are dependent on their income, starting from the fact that the expenses can be much better managed if we take into account the level of income that that family has.

On the other hand, the incomes made by Romanians are mainly obtained from salaries, therefore their expenses must fall within these amounts, otherwise they will have to borrow. Analyzing the results obtained after making forecasts, it can be seen that a 5% increase in total average monthly real income in 2021, lead to an increase in total average monthly real expenditure by the same percentage. The descriptive indicators that measure the quality of the forecast have very low values close to 1, which shows us that the forecast made is one that can be in line with reality.

REFERENCES:

[1]. Andrei, T.; Bourbonnais R. (2008) *Econometrics*, Economic Publishing House, Bucharest

[2]. Anghelache, C.; Manole, A., Anghel, M.G., Pârţachi, I. (2016) *Analysis of the correlation between household income and expenditure*, Informatics, Statistics and Economic Cybernetics, Economics, 4(98), Bucharest

[3]. Babucea, A.G.; Bălăcescu, A. (2011) *Statistical analysis of the dynamics of household income and expenditure in the period 1990-2010,* Annals of Constantin Brâncuşi University of Târgu Jiu, Economics Series, 1, pp. 9-16

[4]. Ţiţan, E.; Boboc, C.; Ghiță, S.; Todose, D. (2011) *Statistical-econometric analysis of the correlations between the social security budget and the main macro-aggregates in Romania*, Theoretical and Applied Economics, vol. XVIII, no. 2(555), pp. 117-126

[5]. https://data.worldbank.org/indicator/FP.CPI.TOTL?locations=RO

[6]. http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table

[7]. https://insse.ro/cms/sites/default/files/com_presa/com_pdf/abf_tr2r21.pdf