TJMM
9 (2017), No. 1, 43-50

CORRELATION ANALYSIS OF CONSUMER PRICE INDICES BY MULTIPLE REGRESSION

ANAMARIA POPESCU

Abstract. The analysis of these correlations between the consumer price indices, by multiple regression, supplements the information and the conclusions drawn by the implementation of models of the regression type simple. Additional information obtained through the use of multiple regression represents a support additional information for decision-makers and analysts. This article describes a correlation between the annual index of consumer prices, total annual index of consumer prices of foodstuffs, annual index of consumer prices of non-food goods and the annual index of consumer prices of services, by means of a regression model in multiple. The model explains the influence of the three types of consumption on the evolution of the annual index of consumer prices total and allows the creation of forecasts.

1. Introduction

In the field of economic phenomena and processes, arise a series of links, interdependencies caused by several different causes and conditions that affect more or less existing phenomena.

The complexity of economic and social phenomena, their quantitative and qualitative characterization determines the combined use of different sciences to investigate causal relationships that underlie the emergence and development.

Some of the methods and models have been imposed in the study of the interdependence of the most commonly used are the correlation and regression statistics.

The use of these methods is justified by the necessity of increasing the reflection in a numerical form appropriate to the interdependence objective of the phenomena social and economic in nature, the direction and the intensity of the connections, which manifests itself in a certain period of time or dynamics.

Between quality and price of products and services are closely interdependent. This relationship is studied in this paper.

Consumer Price Index (CPI) is intended to measure dynamic changes in the general level of prices of goods and services purchased for consumption by households in the country. CPI is an index monthly and is calculated only for items entering the population direct consumption, excluding consumption of goods and services from their own production of household expenditure as investment and accumulation, interest paid on loans, insurance rates fines, taxes, etc. and expenditures paid for production (agricultural etc.) households. CPI measures while the general trend in consumer prices.

Information obtained by using simple linear regression model are not always sufficient to characterize the evolution of an economic phenomenon and, especially, to identify possible further evolution.

2010 *Mathematics Subject Classification.* 62P20, 91B84, 62M10.
Key words and phrases. consumer prices, multiple regression, correlation, data series, Data Analysis.

A significant argument in this sense can be considered to be of great value as possible of the term freely (as the image of the factors which have not been included in the model) that appears in each of the simple linear regressions analysed in previous work

To remedy these shortcomings, the simple linear regression method can extend from sets of two variables to multiple variables through the multiple linear regression method, in which case we have a dependent variable and several factorial variables.

2. Case Study: Consumer Price Indices

The multiple linear regression model can be used in the Romanian economy, which complement analysis using simple linear models. In this sense, we consider as variable resultant value consumer price index total (CPT) and the factorial variables annual index of consumer prices of food products, the annual index of consumer prices of non-food index's annual consumer price services during 1990-2015.

Consumer Price Index (CPI) measures the evolution of prices for purchased goods and tariffs for services used by the population in a given period (current period), compared to a previous period (base or reference period).

The consumer price index shall be calculated only for the elements included in the direct consumption of the population, being excluded: consumption from own resources, the expenditures for investment and accumulation, interest paid on loans, insurance rates, fines, taxes, etc. as well as expenditures paid for agricultural production to individual households.

The four indices may be presented in synthetic form as follows [8] - 1:

On the basis of this information, we will analyze the existence of a possible dependence between the value of the annual index of total consumer prices (the variable Y) and the annual index of consumer prices of food commodities (causal variable X_1), consumer prices Non-food goods (causal variable X_2) and consumption prices of services (causal variable X_3)

Description econometric relationship between the three variables can be done using four models: [1, 2, 4]

- A unifactorial model to explain the annual variation of the consumer price index level changes based on the total annual consumer prices of food products made in our country:

$$y_i = f(x_{1i}) + \varepsilon_{1i} \tag{1}$$

- A unifactorial model to explain the variation in the annual index of consumer prices total on the basis of changes in the level of the annual index of consumer prices of non-food goods in our country:

$$y_i = f(x_{2i}) + \varepsilon_{2i} \tag{2}$$

- A unifactorial model to explain the variation in annual index of consumer prices change based on total consumer price level services:

$$y_i = f(x_{3i}) + \varepsilon_{3i} \tag{3}$$

- A multifactorial model explaining the change in the annual index of total consumer prices based on the simultaneous influence of the three above mentioned indicators.

$$y_i = f(x_{1i}, x_{2i}, x_{3i}) + \varepsilon_i \tag{4}$$

Regarding the first three models presented above, relationships between indicators can be reflected by linear regression models.

The financial series generates greater uncertainty because of econometric modelling or depreciating significantly intensifies and correlation of data series that characterize the

TABLE 1. Annual consumer price index in Romania, between 1990 and 2015

The reference period	TOTAL IPC (%)	The CPI Goods (%)	The CPI non-food goods (%)	The IPC Services (%)
Year	Y	X_1	X_2	X_3
1990	392914,9	267109,5	443139,75	752047,2
1991	145416,32	93329,65	165474,14	319069,7
1992	46842,5	27725,71	56243,15	113774,2
1993	13154,16	7946,85	15241,79	33431,75
1994	5556	3363,97	6545,83	13329,92
1995	4200,77	2551,36	5049,74	9340,81
1996	3026,29	1870,97	3630,54	6356,69
1997	1187,88	744,33	1437,72	2299,22
1998	746,64	501,48	897,18	1197,17
1999	512,09	392,25	589,02	650,51
2000	351,55	272,93	409,05	422,79
2001	261,44	201,14	307,4	312,23
2002	213,35	170,03	244,92	246,31
2003	185,08	148,29	210,97	214,48
2004	165,43	135,45	186,37	186,95
2005	151,76	127,66	167,49	169,14
2006	142,41	122,93	154,41	156,32
2007	135,84	118,33	147,07	146,6
2008	125,95	108,35	138,27	135,03
2009	119,29	104,93	130,18	123,91
2010	112,44	102,54	118,58	118,25
2011	106,29	96,72	111,7	113,22
2012	102,86	94,92	107,64	107,75
2013	98,92	92,19	102,33	104,42
2014	97,87	93,72	99,89	101,23
2015	98,45	97,43	98,96	99,21

Sursa: INSSE – Baze de date statistice – TEMPO-Online serii de timp

TABLE 2. Matrix of correlation applied to independent variables (1990 - 2015)

	X_1	X_2	X_3
X_1	1		
X_2	0,999545	1	
X_3	0,99689	0,99875779	1

financial markets, as it becomes apparent from the correlation matrix applied to the series of consumer price indices [5, 6].

The correlation coefficients resulting from the comparison of independent variables X_1, X_2 and X_3 is close to 1, it is clear that there is a correlation between the significant variables, i.e. they are more linear dependent from each other.

Dependence of independent variables and Y is positive: an increase in the variables X_1, X_2 and X_3 involves an increase in the total consumer price index, Y. Regression lines have an uptrend, and dispersion diagrams indicate an increasing trend. Interpretation

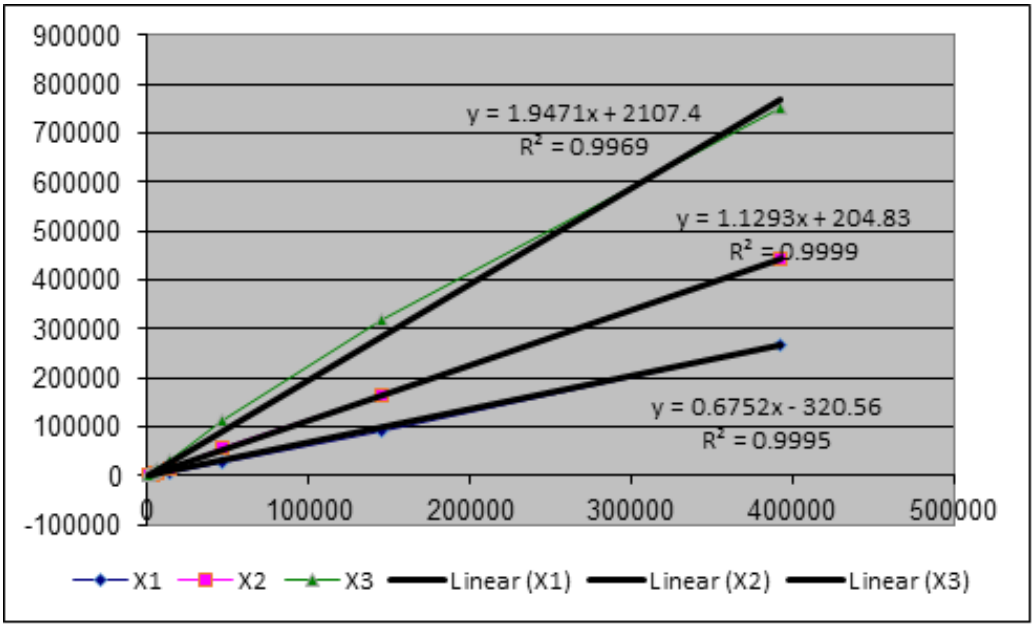

FIGURE 1. The correlation between Y and X_1, X_2, X_3

of R^2: 99% of the variance in Y can be explained by the linear relationship of X_1, X_2 and X_3. The coefficient of the variable X_1 from the equation for the regression Y shall be construed as follows: for each increase in the X_1 with a unit of measure (%p.a.), Y increases with 1.95 percentage points, the coefficient of the variable X_2 from the equation for the regression Y shall be construed as follows: for each increase in the X_2 with a unit of measure (%p.a.), Y increases with 1.13 percentage points and the coefficient of the variable X_3 from the equation for the regression Y shall be construed as follows: for each increase in the X_3 with a unit of measure (%p.a.), Y increases with 0.68 [5, 7].

In this context it is particularly important to specify and analyze the relationship between the four macroeconomic indicators using a regression model multifactorial. From mathematical point of view it can be transcribed as follows: [2, 4]

$$y_i = b_0 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} + b_3 \cdot x_{3i} + \varepsilon_i \tag{5}$$

The multifactor linear identified above can be written in matrix form as follows:

$$Y = X \cdot B + \varepsilon \tag{6}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1X_{11} & \cdots & X_{1k} \\ 1X_{21} & \cdots & X_{2k} \\ & \vdots\vdots\vdots & \\ 1X_{n1} & \cdots & X_{nk} \end{pmatrix} \cdot \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{7}$$

where: $n = 26 \rightarrow$ the number of available observations;
$k = 3 \rightarrow$ the number of exogenous variables.

Regression function corresponding to the model considered, is written as a matrix equation is:

$$\hat{Y} = X \cdot \hat{B} \tag{8}$$

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0,9999999					
R Square	0,9999999					
Adjusted R Square	0,9999999					
Standard Error	28,076996					
Observations	26					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	3	1,6336E+11	54453300719	69075323	7,46941E-77	
Residual	22	17342,98977	788,3177166			
Total	25	1,6336E+11				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-6,521213	5,961909952	-1,093812768	0,285869	-18,88545765	5,8430312
X1	0,7651963	0,009972245	76,73260404	3,2E-28	0,744515166	0,7858775
X2	0,2666513	0,009432325	28,26994072	8,69E-19	0,247089829	0,2862127
X3	0,093574	0,002091383	44,74264243	4,23E-23	0,089236739	0,0979113

FIGURE 2. Regression model estimation results in Excel

For the estimation of the parameters we will use the method of least squares (MCMMP - the least squares). For the linear multifactorial apply this method involves the minimization of the function:

$$\begin{aligned} F\left(\hat{B}\right) &= \min \sum_{t=1}^{n} \varepsilon_t^2 = \min\left(Y - X\hat{B}\right) = \\ &= \min\left(Y^T Y - 2\hat{B}^T\left(\hat{X}TY\right) + \hat{B}^T\left(X^T X\right)\hat{B}\right) \end{aligned} \tag{9}$$

which involves the determination of the derivation of the function in relation to the estimator $\hat{B}$ and canceling it:

$$\left(X^T X\right)\hat{B} = X^T Y \tag{10}$$

To facilitate estimation model multiple regression we used the software Microsoft Excel, where we defined equation is variable resultant annual index of consumer prices Total (Y) and the factorial variables annual index of consumer prices of food products (X_1), annual consumer price index of non-food goods (X_2) and annual consumer price index of services (X_3). I also thought that this regression model will contain free term b_0 will reflect the influence of the terms that were not considered when building the model. The estimation method defined in the program is least squares - Least Squares.

3. The Estimation Results of Regression Model Parameters

In the above model, the multiple regression determined above which describes the relationship between the consumer price indices which are the subject of this research can be transcribed in the form of the equation as follows:

$$y_i = b_0 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} + b_3 \cdot x_{3i} + \varepsilon_i$$

$$Y = -6,521213 + 0,7651963 \cdot X_1 + 0,2666513 \cdot X_2 + 0,093574 \cdot X_3 + \varepsilon$$

As you can see, using the model multiple regression complement the conclusions that can be drawn during analysis using linear model simple namely that the annual index

of consumer prices of food products, annual index of consumer prices of non-food goods and annual index of consumer prices of services are significant influencing factors for the evolution of annual consumer price index total.

It is worth mentioning that, in the model considered, the influence of free term as image factors were not included in the model is significant. Thus, if the three variable factorial, X_1, X_2 and X_3 have the value 0, the average Y value is estimated to about -6.521213. We can say that factors not taken into account when building econometric model causes a reduction in the value of the annual consumer price index total.

From the point of view of statistical tests that verify the accuracy of the econometric model considered, it can be seen that the respective values of R^2 and R^2 adjusted tests are very close to the maximum (R^2 = 99.99%, and the adjusted R^2 = 99.99%) which allows us to conclude that the model under consideration is correct and a minimum risk for economic analysis. Also, we can say that the introduction into the model of several variables factorials leads to an increase of the degree of probability of it in comparison with the models linear regression simple.

The coefficient of determination indicates that 99.99% of the variance of the dependent variable explained by the simultaneous variation of the three indices of annual consumer price of food, non-food goods and services in the period 1990 to 2015, i.e. a strong link between at enhancing the variable and those three variables exogenous, confirmed and the coefficient of determination adjusted (adjusted R squared = 0,9999), which takes into account and the number of sightings and the number of exogenous variables. The ratio of the correlation (R = 0.9999) tending towards 1 shows that the estimated regression model approximates the observation data very well, with a high reliability.

bi parameters are statistically significant at the 5% significance threshold (i = 0,1,2,3), so the model is valid. This is reinforced by the very small values of P-value for each parameter of the model.

We can see that the statistical F-test value is greater than the reference value of the table, which leads to the idea that considered econometric model is correct, which can then be used in analysis and forecasting macro consumer price index of the total yearly (Y).

Test value Significance F is zero, confirming statements made previously, whereby an econometric model regression using the variable resultant annual index of consumer prices overall and variables factorial annual index of consumer prices of food products, the annual index of non-food consumer price index and consumer price annual service public and private consumption is correct and can be used in forecasting price trends of total consumption.

Quality analysis is facilitated by the graphics model built Regression automatic procedure.

In the figure 3 there are presented the residual-type variables - independent variables X_1, X_2, and X_3. The points in the figure can be considered as a horizontal band region which does not contradict the assumptions of normality of the errors. The uniform band shape reflects the constant of residue dispersion for the entire range of independent variables X_1, X_2, and X_3.

4. Conclusions

Based on the observations made on the analysis of price index of total consumption using regression models we conclude that this indicator is influenced by changes in the annual index of consumer prices of food products (X_1), the annual index of consumer prices of non-food goods (X_2) and annual consumer price index of services (X_3).

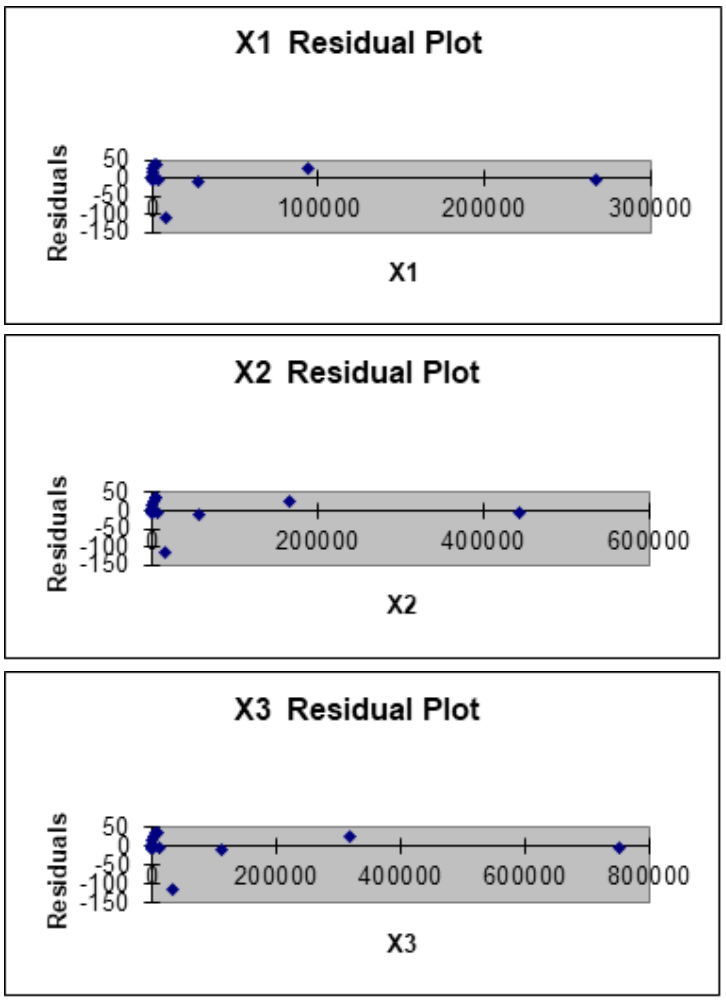

FIGURE 3. Graphics allocated to residues

The methodological, it can be seen that the use of multifactor regression model allows to obtain conclusive results in macroeconomic analysis, without this meaning that single factor regression model does not allow for relevant research on the development of the national economy.

REFERENCES

[1] Anghelache, C., Anghel, M.C., *Metode de analiză a PIB prin utilizarea modelelor statistico-econometrice*, Revista / Journal ,,Economica" nr. 1 (91), 2015.
[2] Anghelache, C., Manole, Al., Anghel, M.C., *Analiza corelației dintre PIB, consumul privat și public prin regresie multiplă*, Revista Română de Statistică - Supliment nr. 8 / 2015.
[3] Craiu, M., *Mathematical Statistics*, Matrix Rom, Bucharest, 1998.
[4] Iyiola, R.O. and Adetunji, A.A., *The effect of monetary policy on consumer price index*, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6, August 2014.
[5] Popescu, A., *Statistical analysis of consumer price indices*, Transylvanian Journal of Mathematics and Mechanics, Vol. 8 (2016), No. 1, Edyro Press Publisher.
[6] Popescu, A., *Monetary policy indicators analysis based on regression function*, Transylvanian Journal of Mathematics and Mechanics, Vol. 8 (2016), No. 2, Edyro Press Publisher

[7] Popescu, A., Stoicuța, N.A., *Statistical analysis of monetary policy indicators variability*, The Annals of the University of Petroșani, vol. XVI - 2016.
[8] Sursa: INSSE – Baze de date statistice – TEMPO-Online serii de timp

Department of Mathematics & Computer Science
Faculty of Science, University of Petroșani
20 University Street, Petroșani, Romania
E-mail address: am.popescu@yahoo.com