TJMM
**10** (2018), No. 2, 113-120

# ANALYSIS OF THE EVOLUTION AND CORRELATION BETWEEN GROSS NET SALARY AND CONSUMER PRICE INDEX

ANAMARIA POPESCU

Abstract. This paper presents an econometric analysis of the ratio between the gross net salary and the consumer price. The work begins with a historical overview. The calculation statistics for the two variables will be presented and the relation between the variables will be further studied. Next we will find the regression equation by applying the regression function in EXCEL, Regression. Also, the hypothesis of homoscedasticity will be verified by using the graphical method and applying the Durbin-Watson test. If the relationship is directly proportional (the increase of the variable 1 involves the increase of the variable 2, the decrease of the variable 1 involves the decrease of the variable 2), or inversely proportional (the increase of the variable 1 leads to the decrease of the variable 2). Finally, we will present the conclusions and references.

## 1. Introduction

1. The term salary was initially used in the Roman army to designate the amount of money the Romanian soldiers were getting to buy salt (salary). The owner of the work resource, used in a certain human activity, is rewarded by salary, which is the amount paid for labor factor services, that is the price paid in exchange for the work performed. Given that demand and supply of work are equal, equilibrium salary is formed. In relation to the position of the two partners - the user and the owner of the labor resource - the way of interpreting the salary is different. For the former, salary is a component of the cost of an economic asset, and the second is the salary as the price of his work, which rewards his work in a certain human activity and takes the form of income.

In general, the salary is considered the income of most of the employed population in countries with a consolidated economy and which holds the main share between the market-specific fundamental incomes. According to their forms of existence and their practical significance, two forms of salary are distinguished: nominal salary (gross and net) and real salary.

The nominal salary ($Sn$) is the amount of money that the employee receives from his employer, expressed in current market prices. Gross salary includes all income from work (basic salary and other benefits). The net salary is based on the reduction of the gross nominal salary with the salary tax, with the social insurance contributions and other obligations stipulated by the law.

In countries with market economy, the minimum wage (minimum guaranteed wage) is set by law. It consists of a certain amount of money needed for subsistence insurance, being a measure of social protection of the employees. Its size varies from one country to another and from one period to the next.

2010 *Mathematics Subject Classification.* 62P20, 91B38, 62J02.

*Key words and phrases.* descriptive statistics, ANOVA, correlation, regression equation, Durbin-Watson test, Data Analysis.

The real salary ($Sr$) expresses, in the form of money, the quantity of goods, materials and services that can be acquired by an employee or a family of employees at a given time with the net nominal salary earned. The size of the real wage was influenced by two factors: the net nominal salary size and the price level of the goods for the consumption of the population.

If the nominal salary ($Sn$) grows faster than the prices increase, then the real wage increases; if prices rise more than the nominal salary, then the real wage drops. In other words, the real wage is nominal wage index adjusted consumer prices (IPC):

$$S_r = S_n/\text{IPC} \cdot 100$$

In practice, the evolution of nominal and real salary is calculated by indices. The nominal salary index ($I_{sn}$) is determined by reporting the current salary nominal salary ($S_{n1}$) to the nominal salary in the base period ($S_{n0}$); the real salary index ($I_{sr}$) is determined by comparing the actual salary in the current period ($S_{r1}$) during the actual base salary ($S_{r0}$) or by dividing the nominal wage index to the consumer price index (IPC), as follows:

$$I_{sn} = S_{n1}/S_{n0}; \quad I_{sr} = S_{r1}/S_{r0} \cdot 100; \quad I_{sr} = I_{sn}/\text{IPC} \cdot 100.$$

The tax notions used to define the salary in different stages are: gross income, net income, income tax base, earnings.

Salary negotiation and setting can be made in local or in foreign currency, the "gross" or "net". This term reflects the amount you would have to receive the employee after the employer has charged mandatory contributions and tax on monthly salaries. Unlike the meaning accorded to various terms in terms of gross or net tax, in terms of identifying negotiation with or without salary.

Gross amount includes revenue of each employee's income realization.

By place of realization income is understood to mean the individual times of legal employer, or the entity employer re-presented.

Net salary represents the difference between gross salary and expenses deductible (mandatory contributions and professional expenses granted only at the place where the basic function of the employee).

Salary earning is the salary payable to and paid to the employee after the employer has withheld compulsory contributions, wage tax and other charges.

The simplest situation for the employer implies the establishment of the salary in lei, the monthly salary earnings after the deduction and transfer of the compulsory contributions and the monthly salary tax.

## 2. Research Methodology

The analysis of the gross salary correlation with the consumer price index was performed, comparing their evolution between 2004 and 2016. The study of the data series, which includes the evolution of the two variables, is sufficiently suggestive [1, 2].

The data series was taken from `www.insse.ro` from the online tempo database of the National Institute of Statistics [10].

The variables are as follows:

X - the independent variable, annual gross net income (%),

Y - the dependent variable, the consumer price index (%).

## 3. Results

*Calculation of values and statistical coefficients*

| AN | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IPC - % | 111,9 | 109 | 106,56 | 104,84 | 107,85 | 105,59 | 106,09 | 105,79 | 103,33 | 103,98 | 101,07 | 99,41 | 98,45 |
| VNB - % | 95,8 | 97,1 | 96,7 | 97,3 | 98,1 | 99,1 | 99,1 | 99 | 98,7 | 98,1 | 99 | 97,9 | 97,6 |

TABLE 1. X Independent variable: VNB for 2004 - 2016; Y dependent variable: IPC for 2004 – 2016. Source: `www.insse.ro`

| X = VNB - % | | Y = IPC - % | |
|---|---|---|---|
| Mean | 97,96154 | Mean | 104,91231 |
| Standard Error | 0,287683 | Standard Error | 1,0429151 |
| Median | 98,1 | Median | 105,59 |
| Standard Deviation | 1,037255 | Standard Deviation | 3,760284 |
| Sample Variance | 1,075897 | Sample Variance | 14,139736 |
| Kurtosis | -0,22217 | Kurtosis | -0,042234 |
| Skewness | -0,66051 | Skewness | -0,103755 |
| Cv | 0,010588 | Cv | 0,0358422 |
| Count | 13 | Count | 13 |

FIGURE 1. Descriptive statistics indices of X independent variable as VNB and Y dependent variable as IPC (2004-2016).

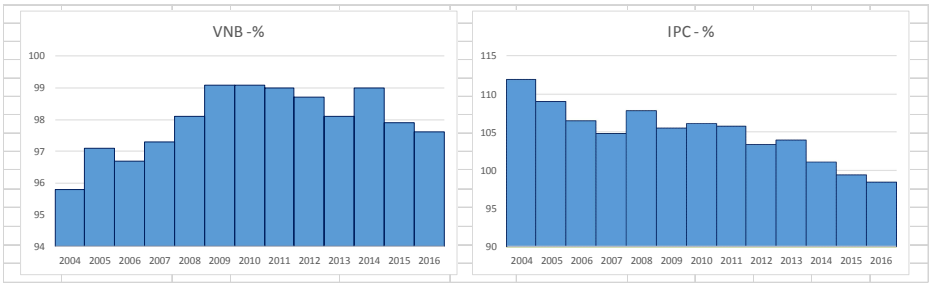


FIGURE 2. Histograms of X independent variable as VNB and Y dependent variable as IPC (2004-2016). Source: INS, `http://statistici.insse.ro/shop/`

The values of the two coefficients of variation take values in the range $0 < CV \leq 35\%$, which denotes that the data variations are small, the significant meanings, the series are homogeneous and the groups are well realized [5, 7].

The coefficient of variation can be used as a mean significance test of the average representativeness, and because the data variations are small, we note that the averages are strictly representative.

Coefficients of vaulting are negative for both data series, which results in slightly platikurtic distributions. A sharp, platikurtic distribution shows that data is very clustered and close to average, data series having a high degree of data homogeneity.

Since the asymmetry coefficient for the variable X is negative and very different from zero (-0.66), the VNB data series is moderately negative asymmetric, the curve being elongated to the left, and the IPC data series is approximately asymmetric negative because the asymmetric coefficient for the Y variable is, and it is negative (-0,10) [9].

*The validity of the regression model*

Using the regression function of EXCEL, namely by selecting Tools - Data Analysis – Regression, we have obtained the following results, which will be interpreted individually [3, 8].

| SUMMARY OUTPUT | |
|---|---|
| | |
| *Regression Statistics* | |
| Multiple R | 0,4239 |
| R Square | 0,1797 |
| Adjusted R Square | 0,1051 |
| Standard Error | 3,5572 |
| Observations | 13 |

TABLE 2. Statistics of the regression equation

| ANOVA | | | | | |
|---|---|---|---|---|---|
| | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 1 | 30,4846 | 30,4846 | 2,4091 | 0,1489 |
| Residual | 11 | 139,1922 | 12,6538 | | |
| Total | 12 | 169,6768 | | | |

TABLE 3. ANOVA (analysis of variance) test

A first table of results, shown in the figure below 3, contains general statistics of the regression equation.

Multiple R (multiple correlation coefficient or r) = 0.4239. We note that the value of "r" is $> 0$, which means that there is a direct link between the two variables considered, VNB and IPC.

R Square ($R^2$) (determination coefficient), expresses how much of the IPC frequency variation is explained by the change in VNB. It can take values in the range [0,1]. The closer the value of 1 to 1, the greater the part of the Y variation explained by X, and the link between them is stronger. In our case, R Square has the value of 0.1797, expressing a percentage that 17.97% of the change in the consumer price index can be explained by the gross net income variable, which indicates a very weak link between the two variables.

Adjusted R Square shows that 0.1051 of the total variation is due to the regression line, taking into account the number of degrees of freedom (n-k = 13-2 = 11).

Standard Error, It is calculated as the standard deviation of the residue and is the estimation of the standard deviation of the errors $\varepsilon$ (assuming their normality). In our case, it's worth 3.5572.

The number of observations in the sample is 13.

The second result table (ANOVA) 3 includes the variance analysis matrix associated with the estimated regression.

The ANOVA (analysis of variance) test is used to validate the regression model used.

The variation explained by the regression model is 30.4846, and the average of the variation explained, corrected by the number of degrees of freedom (2), is 30.4846. The residual variation (the unexplained variation of the regression model) is 139.1922 and the average of the residual variation corrected by the number of degrees of freedom (11) = 12.6538.

The F (Fisher) test is calculated in the table. Since F = 2.4091 and Significance F = 0.1489 (slightly higher than $\alpha = 0.05$) the built-in regression model is valid for a probability of up to 95% and can be used for analysis the dependence between the consumer price index and gross net income [1, 4, 6].

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Intercept | 255,4412 | 96,9869 | 2,6338 | 0,0233 | 41,9744 | 468,9080 |
| X = VNB % | -1,5366 | 0,9900 | -1,5521 | 0,1489 | -3,7156 | 0,6424 |

TABLE 4. Coefficients of regression equation

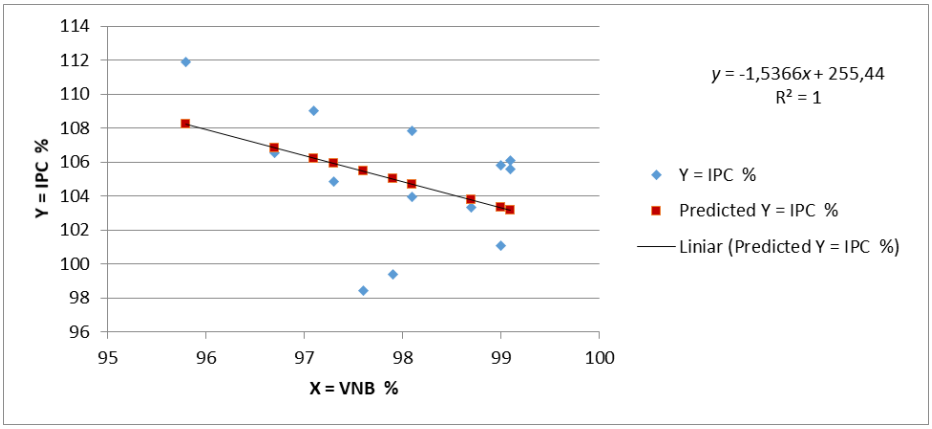


FIGURE 3. $XY$ (scatter), $X$-$Y$ diagram, function graph

| Correlation | | |
|---|---|---|
| | *X = VNB - %* | *Y = IPC - %* |
| X = VNB - % | 1 | |
| Y = IPC - % | r=-0,423866 | 1 |

TABLE 5. Correlation between (X, Y)

Intercept is the free term, so the coefficient a = 255,4412. The free term is the point where the explanatory variable is 0. Since t statistic = 2.6338 and P-value 0.0233 <0.05, this coefficient is significant. The free term of the regression equation is 95% probable in the range: [41.9744; 468.9080] [9].

The coefficient corresponding to the independent variable (b) has a value of -1.5366, which means that on a one-unit increase in gross net income, the consumer price index will decrease by -1.5366 and vice versa. Because the significance threshold P-value = 0,1489> 0,05 means that this coefficient is insignificantly statistically, ie the independent variable (VNB) has a certain influence on the dependent (IPC), but not totally. The confidence interval for the "gross net income"-VNB parameter [-3.7156; 0.624].

From the coefficients analysis, we deduce that the regression model is:

$$y = -1,5366x + 255,4412$$

The relationship between the two variables is inverse. As previously pointed out in the X-VNB variable unit growth, the Y-IPC variable decreases by -1.5366 and vice versa.

It can be seen from the graph and from the negative correlation coefficient (r = -0,423866) that we have an inverse correlation, the two correlated variables vary in opposite order (one increases, the other decreases).

| RESIDUAL OUTPUT | | |
|---|---|---|
| Observation | Predicted Y = IPC % $\hat{y}_i = 255{,}44 - 1{,}5366 \cdot x_i$ | Residuals $e_i = y_i - \hat{y}_i$ |
| 1 | 108,2338 | 3,6662 |
| 2 | 106,2362 | 2,7638 |
| 3 | 106,8508 | -0,2908 |
| 4 | 105,9288 | -1,0888 |
| 5 | 104,6995 | 3,1505 |
| 6 | 103,1629 | 2,4271 |
| 7 | 103,1629 | 2,9271 |
| 8 | 103,3166 | 2,4734 |
| 9 | 103,7776 | -0,4476 |
| 10 | 104,6995 | -0,7195 |
| 11 | 103,3166 | -2,2466 |
| 12 | 105,0069 | -5,5969 |
| 13 | 105,4679 | -7,0179 |

TABLE 6. Residual output table in excel

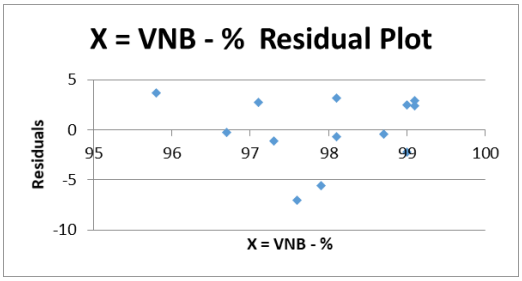


FIGURE 4. Graphical representation of residues ($e_i$) according to (predicted Y)

*Testing of error normality*

The points in the figure 3 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 residual dispersion for the entire range of the independent variable X.

*Testing the homoscedasticity Hypotheses*

If the chart falls into a band, then the errors are homoscedasticity. However, the graphical procedure is approximate, being suggestive only for large volume samples.

*Durbin-Watson test. Assumption of independence of errors*

To apply the Durbin Watson test , first time we make the next assumptions:

H0: errors are independent

H1: errors are dependent

In our case: DWcalc = 0.3873. This is the case for positive autocorrelation, and the errors are not independent, are in the position of positive autocorrelation. So if the errors

| VNB -% | ei² |
|---|---|
| 95,8 | 13,44136 |
| 97,1 | 7,638821 |
| 96,7 | 0,084566 |
| 97,3 | 1,185563 |
| 98,1 | 9,92536 |
| 99,1 | 5,890651 |
| 99,1 | 8,567717 |
| 99 | 6,117732 |
| 98,7 | 0,200327 |
| 98,1 | 0,517746 |
| 99 | 5,047189 |
| 97,9 | 31,32494 |
| 97,6 | 49,25025 |

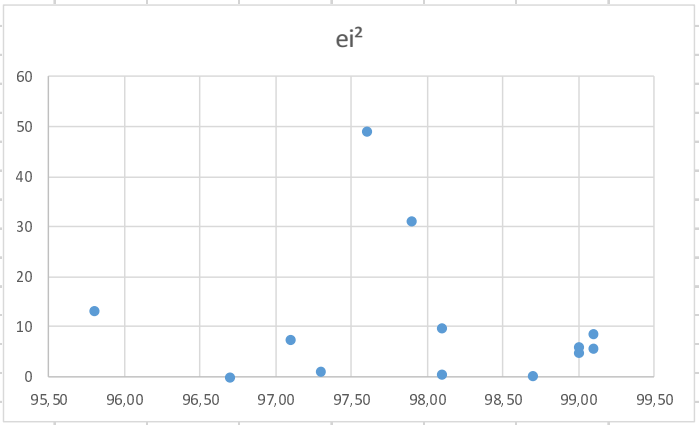


FIGURE 5. Graphical representation of $e_i^2$ according to X-VNB

| | |
|---|---|
| Sum of Squared Difference of Residuals | 53,9142 |
| Sum of Squared Residuals | 139,1922 |
| | |
| **Durbin-Watson Statistic** | **0,3873** |

TABLE 7. Durbin-Watson Calculations

| t-Test: Paired Two Sample for Means | | |
|---|---|---|
| | | |
| | *X = VNB - %* | *Y = IPC - %* |
| Mean | 97,96153846 | 104,9123077 |
| Variance | 1,075897436 | 14,1397359 |
| Observations | 13 | 13 |
| Pearson Correlation | -0,423866485 | |
| Hypothesized Mean Difference | 0 | |
| df | 12 | |
| t Stat | -5,823168319 | |
| P(T<=t) one-tail | 4,0862E-05 | |
| t Critical one-tail | 1,782287556 | |
| P(T<=t) two-tail | 8,17239E-05 | |
| t Critical two-tail | 2,17881283 | |

TABLE 8. The results of the t Test: Paired Two Sample for Means

are in a positive autocorrelation situation, the model considered can not be used in the forecast.

The results of the "t Test: Paired Two Sample for Means" procedure are described in the following table.

It can be seen that the average of the variable X is lower than the average of the Y variable (97.96 vs. 104.91), the difference being relatively important, 6.95. Comparing averages wants to show if this difference can be accepted for the entire population, or is the effect of the survey (incidentally in the second sample there are several large values).

Because Pearson's coefficient is p = -0.42, then the two random variables vary inversely and indicate an acceptable degree of association.

In this case, P (T <= t) two-tail (0,0000812) gives the probability that the absolute value of t-Stat (5,823) is observed, which is greater than t Critical two-tail ). Since p is less than $\alpha = 0.05$, we reject the null hypothesis that there is no significant difference in the mean of each variable.

## 4. Conclusions

The relation between the two variables Y and X is an inverse relationship between the variables. The logic of the linear regression equation best represents the relationship between the two variables: IPC = -1.5366VNB + 255.4412. The equation is determined by using the Excel software, the Data Analysis package. The model considered is homoscedastic. Regarding errors, we can say that in 2006, 2007, 2012 to 2016 most errors are negative and positive in 2004, 2005, 2006-2011. At any time, errors are correlated, not independent. Errors are not normally distributed. Such errors are placed in the positive aucorelate model. They can not be used in the forecast.

In the case of a negative correlation coefficient we have an inverse correlation, the two correlated variables vary in opposite order (when one increases, the other decreases).

The notion of correlation refers to the tendency of simultaneous growth or decrease of two parameters no matter how strong or weak this tendency is.

In the studied case the two variables are correlated by the increase of one accompanied by a decreasing tendency of the other. Thus, the rise in VNB is accompanied by a drop in IPC. These two indices are an example of an inverse correlation.

## References

[1] Anghel, M.G., Mirea, M., Olteanu, C., *Evolution and correlation of gross salary with inflation*, Revista Română de Statistică - Supliment nr. 2 / 2018.

[2] Anghel, M.G., Lillea, F.P.C., Mirea, M., *Analysis of the interdependence between GDP and inflation*, Revista Română de Statistică - Supliment nr. 3 / 2017.

[3] Bowerman, B.L., O'Connell, R.T., Koehler, A.B., *Forecasting, time series and regression: An applied approach*, Thomson Brooks/Cole, Belmont CA, 2005.

[4] Gjika (Dhamo), E., Puka, L., Zaçaj, L., *Forecasting consumer price index using time series models and multiregression models*, 10th International Scientific Conference "Business and Management 2018" May 3–4, 2018, Vilnius, LITHUANIA, doi: 10.3846/bm.2018.51.

[5] Marcellino, M., *A comparison of time series models for forecasting GDP growth and inflation*, IEP-Università Bocconi, IGIER and CEPR, 2007.

[6] Norbert, H., Wanjoya, A., Waititu, A., *Modeling and forecasting consumer price index*, American Journal of Theoretical and Applied Statistics **5** (3), 2016.

[7] Popescu, A., Stoicuța, N.E. , Annals of the University of Petrosani, Economics 2016.

[8] Popescu, A., *Monetary policy indicators analysis based on regression function*, Transylvanian Journal of Mathematics and Mechanics, Vol. **8** (2), Edyro Press Publisher, 2016.

[9] Popescu, A., *Statistical analysis of consumer price indices*, Transylvanian Journal of Mathematics and Mechanics, Vol. **8** (1), Edyro Press Publisher, 2016.

[10] `http://statistici.insse.ro/shop/`

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