STATISTICAL STUDY ON THE CORRELATION BETWEEN „GROSS DOMESTIC PRODUCT" AND „NET INVESTMENTS" IN ROMANIA, IN THE PERIOD 2008-2022

Assoc. Prof. Ilie Murarita
Ph.D
University of Craiova
Faculty of Economics and Business Administration
Craiova, Romania

Abstract: "Gross domestic product" is the main indicator for evaluating the economic growth of a country, reflecting the final result of the production activity of resident production units. "Net investment" shows how money is allocated to economic growth. Net investments are found in various expenses (for construction, installation and assembly works, for the purchase of machinery, means of transport, other expenses intended for the creation of new fixed assets, etc.). The increase or decrease of the Gross Domestic Product is directly related to the volume of investments made. In this article, the analysis of the correlation between these two indicators is carried out in Romania, in the period 2008-2022, using the results of this analysis to make a forecast for the year 2023.

JEL classification: C82, O11

Key words: Gross Domestic Product, Net Investment, statistic, correlation, prediction

1. INTRODUCTION

The data specific to the indicators under study were taken from the National Institute of Statistics. Also here are their definitions, which we will reproduce in the following.

Macroeconomic aggregates are synthetic values that measure the output of the activity of the total economy seen from a particular perspective (eg production, value added, disposable income, final consumption, etc.).

The gross domestic product (GDP), the main macroeconomic aggregate of national accounting, represents the final result of the production activity of resident productive units. It is equal to the sum of the gross added values of the different institutional sectors or different branches of activity, to which taxes are added and subsidies on products (which are not distributed by sectors and branches of activity) are subtracted. It also represents the balance of the production account of the total economy.

Gross disposable income measures the income available to the nation to carry out final consumption and gross economy operations.

The financing capacity or need is the net amount of resources that the total economy makes available to the rest of the world (if positive) or receives from the rest of the world (if negative).

Net investments represent the expenses incurred for construction, installation and assembly works, for the purchase of machinery, means of transport, other expenses

intended for the creation of new fixed assets, for the development, modernization, reconstruction of existing ones, as well as the value of services related to the transfer of ownership of existing fixed assets and land (notary fees, commissions, transport, loading-unloading expenses, etc.). Net investments are presented on homogeneous activities (by destination) of the national economy. Net investments include the following structural elements:

a) *constructions* - represents the set of works through which new constructions are carried out, reconstruction, development, modernization of buildings with industrial, agricultural purpose, etc. It also includes the assembly work of the technological and functional equipment with the entire complex of operations through which the assembly of their components on the site is carried out, as well as their fixing on the foundations (including the value of the samples and mechanical running-in of the equipment and technological lines);

b) *machines* (with and without assembly) - represent the machines, equipment, lines and technological installations that can work only after ensuring the assembly works, as well as those that can work independently. It also includes the means of transport;

c) *geological and drilling works* - the set of works for: identifying new reserves of useful mineral substances, in fluid or solid state, within the perimeters in operation, in new structures and under research; promotion of reserves of useful mineral substances in higher categories; specification of some characteristics of the reserves of useful mineral substances in order to include them in the group of balance reserves; obtaining design data for energy, hydrotechnical, industrial constructions, etc., to determine the structure and composition of the soil, the groundwater regime, for the lands on which the investment objectives are to be located;

d) *the works executed for the purpose of extracting useful mineral substances in a fluid state* (exploitation drilling), as well as the works executed for the purpose of increasing the recovery factor and the rate of exploitation of crude oil, by injecting fluids into the deposit;

e) *other net investments* - represents the expenses incurred for the purchase of work, production and reproduction animals, vineyards, tree plantations, forestry, the purchase of household inventory items of the nature of fixed assets, the payment of research and design studies for investment objectives, the banks' commission for operations carried out in connection with investments.

The population's investments refer to investments for housing, household constructions, vineyards and tree plantations, animals transferred to the basic herd, purchases of machinery and agricultural machinery taken over with payment from other units.

The gross domestic product represents the value of goods produced in society during a year and reached the last stage of the economic circuit. This value is determined by branches of the national economy and is summed up at its level. In principle, at the level of the national economy, GDP is calculated as the difference between the gross global product and intermediate consumption; at branch level, the difference is made between its overall gross production and the related intermediate consumption. Gross global output expresses the value of goods created, regardless of whether or not they are commodities. This method of determining GDP at the branch or economy level is known as the production method [Ciurlău Constantin, 2008].

The difference between the product or gross production and intermediate consumption takes on different forms of expression and characterization, namely consumption, expenditure or final demand and added value, a fact that led to the formulation of two specific methods for calculating the gross domestic product, namely the income method and the expenditure method.

According to the income method, which is the most widespread, the gross domestic product is calculated as the sum of the costs that form the added value in all the productive institutions in the country, namely: net indirect taxes, i.e. indirect taxes paid less subsidies received, consumption of fixed capital or depreciation, employee benefits and other personnel expenses, and operating surpluses. Indirect taxes are the amounts of money that units pay to the state or international bodies and include in their production costs. These are taxes on the production, sale, purchase or use of goods and import duties. Direct taxes on income or wealth paid by businesses are not included here, because they are included in value added as part of the profit from which it is borne (operating surplus). Subsidies are granted by the state to public or private enterprises to partially cover their costs. Subsidies are linked to the value of the goods produced, exported or consumed, to the labor force employed or to the areas used for production, as well as to the way this production is organized or executed. Amounts granted by the state to private enterprises to finance investment, cover damage or loss to real estate equipment and renewable assets are considered capital transfers and not subsidies. The consumption of fixed capital expresses the value of equipment-like goods consumed for production during a year, as a result of normal physical wear and tear. Destruction due to various causes and removal from use due to wear and tear are considered fixed capital losses. The operating surplus is the difference between the product or gross output, on the one hand, and intermediate consumption, net indirect taxes, consumption of fixed capital and employee remuneration, on the other hand; basically, it's about profit. Exploitation surpluses can only occur in enterprises that capitalize their production at the market price; this is not the case with public administrations that provide services to third parties at their expense. According to the expenditure or final use method, the gross domestic product is calculated by summing up the final destinations of production, namely: final public (governmental) consumption, final consumption of households (households), final consumption of non-profit private institutions, gross formation of fixed capital, change in inventories and net export.

2. STUDY OF THE CORRELATION BETWEEN "GROSS DOMESTIC PRODUCT" AND "NET INVESTMENTS" IN ROMANIA, IN THE PERIOD 2008-2022

The evolution of the "Gross Domestic Product" in Romania, during the period 2008-2022, is shown in figure 1. (data are taken from Annex 1). In most of the period we record its growth, which is a favorable situation. The exception is 2009 (when it decreases, compared to 2008, by 8,940.2 million lei).

Figure 2 shows the evolution of "Net Investments" in Romania, in the period 2008-2022 (data are taken from Annex 1). As we can see, the indicator shows an oscillating variation, with alternate increases and decreases.

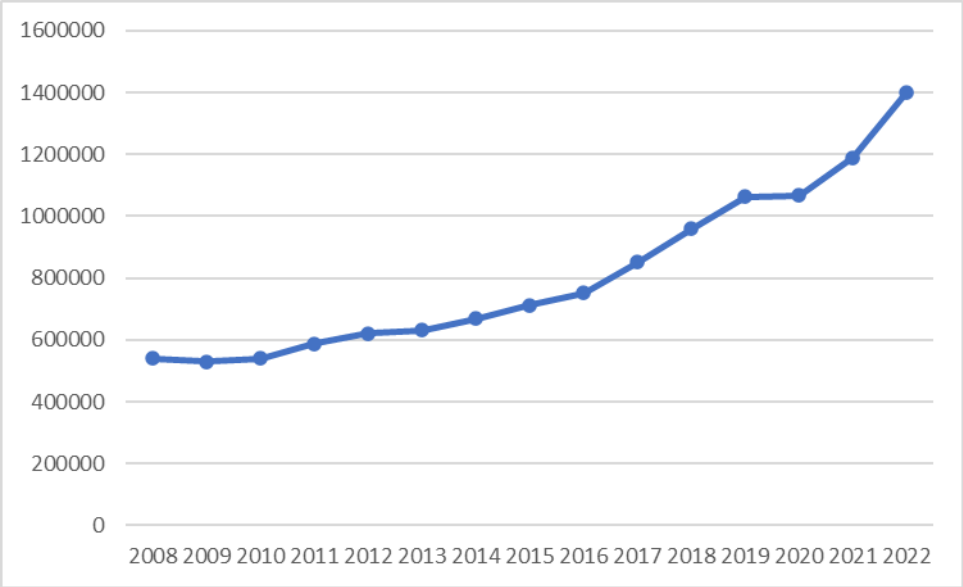

Figure 1. Evolution of the ”Gross Domestic Product” in Romania, in the period 2008-2022

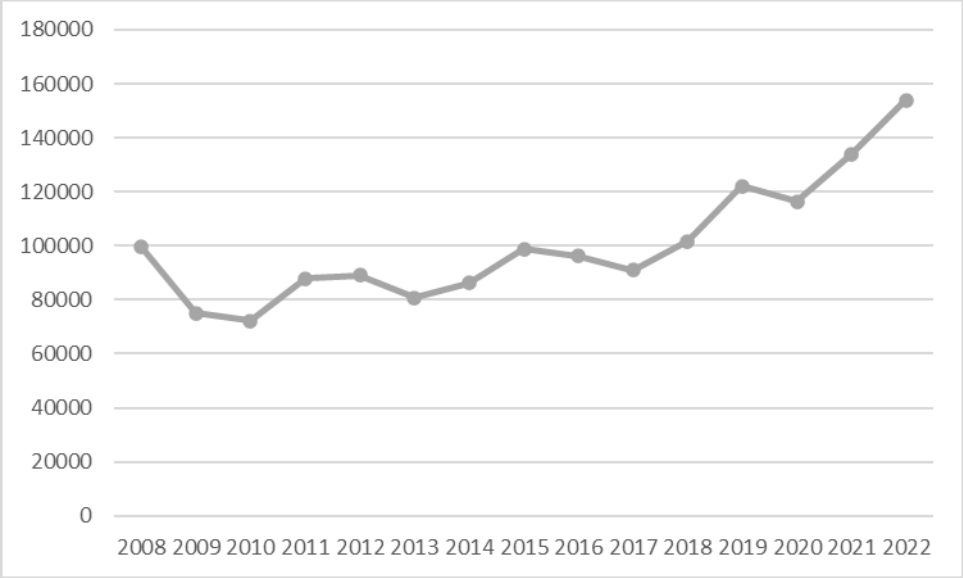

Figure 2. Evolution of ”Net Investments” in Romania, in the period 2008-2022

To study the link between ”Gross Domestic Product” and ”Net Investments” we used a unifactorial linear regression model, whose form is:

$$y_i = \alpha + \beta \cdot x_i + \varepsilon_i,\ i = 1, 2, \ldots$$

and the adjustment equation

$$y_i = a + b \cdot x_i + e_i,\ i = 1, 2, \ldots$$

where: - y represents the dependent variable - in the present case, it is the ”Gross Domestic Product”;

- x is the independent variable - that is, ”Net Investments”;

- α is the free term and, mathematically, represents the point where the regression line intersects the OY axis. It has no economic significance;

- β is called the regression coefficient and, mathematically, it represents the slope of the regression line showing how many units the variable y changes when the variable x increases by one unit;

- ε is called the residual or error variable and records the influences of all other unrecorded factors.

The graphical approximation of the model of the link between the two variables is made using the scatter diagram in figure 3.

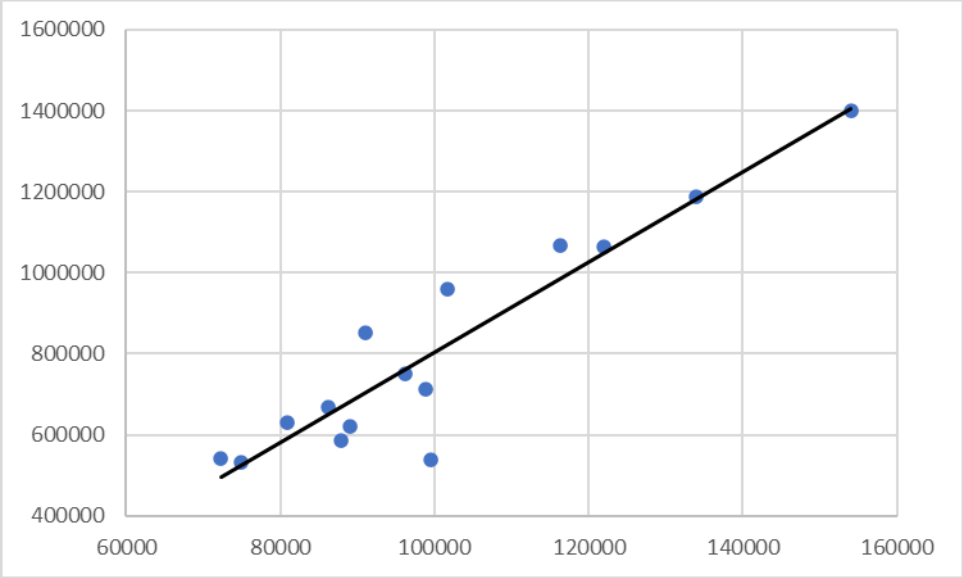

Figure 3. Scatter diagram of the variables "Gross domestic product" and „Net investments" in Romania, in the period 2008-2022

For *the point estimation of the model parameters*, we will apply the least squares method, through the system of equations:

$$\begin{cases} na + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 = \sum x_i y_i \end{cases}$$

After solving the system (calculations needed to estimate the parameters are in Annex 2), the adjustment equation becomes:

$y_i = -309700.51 + 11.14 \cdot x_i$

In what follows, we will estimate the parameters through confidence intervals (the necessary calculations can be found in Annex 3).

The confidence interval for the parameter α is obtained using the relation:

$a - t_{\alpha/2,n-k} \cdot s_a \leq \alpha \leq a + t_{\alpha/2,n-k} \cdot s_a$

where: - n represents the number of terms ($n=15$);

- k represents the number of model parameters ($k=2$).

$t_{\alpha/2,n-k} = t_{0.025,13} = 2,1604$

$$s_a = s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}$$

$$s = \sqrt{\frac{\sum(y_i - Y_i)^2}{n-2}} = \sqrt{\frac{134869875229.7}{15-2}} = 101855.81$$

$$s_a = 101855.81 \cdot \sqrt{\frac{1}{15} + \frac{100319.9^2}{7142437084.9}} = 123733.63$$

-309700.51 – 2.1604·123733.63 ≤ α ≤ -309700.51 + 2.1604·123733.63

-577010.76 ≤ α ≤ -42390.26

The confidence interval for the parameter β is obtained using the relation:

$b - t_{\alpha/2,n-k} \cdot s_b \leq \beta \leq b + t_{\alpha/2,n-k} \cdot s_b$

$$s_b = s \cdot \sqrt{\frac{1}{\sum(x_i - \bar{x})^2}}$$

$$s_b = 101855.81 \cdot \sqrt{\frac{1}{7142437084.9}} = 1.2052$$

11.14 - 2.1604·1.2052 ≤ β ≤ 11.14 + 2.1604·1.2052

8.54 ≤ β ≤ 13.74

These results lead us to the observation that we can guarantee with a probability of 95% that the free term (α) falls between the values -577010.76 and -42390.26, and the regression coefficient (β) falls between the values 8.54 and 13.74. Because none of the intervals contain 0, we can guarantee that the parameters are statistically significant.

The linear correlation coefficient is determined using the formula (the necessary calculations are in Annex 3):

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

$$r_{xy} = \frac{79559275178,5}{\sqrt{7142437084,9 \cdot 1021076948764,0}} = 0,9316$$

To test the significance of the linear correlation coefficient, the following statistical hypotheses are issued:

- H0: ρ = 0 (not statistically significant)
- H1: ρ ≠ 0 (it is statistically significant)

The t-test will be:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,9316\sqrt{15-2}}{\sqrt{1-0,9316^2}} = 25.43$$

Since $t(25.43) > t_{0.025;13}(2,1604)$ the null hypothesis is rejected, the alternative hypothesis is accepted according to which the linear correlation coefficient is statistically significant.

To test the validity of the model, the following statistical hypotheses are issued:

- H0: the independent variable does not have a significant influence on the dependent variable, and the Coefficient of Determination (R^2) is not statistically significant.

- H1: the independent variable has a significant influence on the dependent variable, and the Coefficient of Determination (R^2) is statistically significant.

The F-test will be (the necessary calculations can be found in Annex 3):

$$F = \frac{\frac{\sum(Y_i - \bar{y})^2}{k-1}}{\frac{\sum(y_i - Y_i)^2}{n-k}} = \frac{\frac{886207073534,3}{2-1}}{\frac{134869875229,7}{15-2}} = 85.42$$

The coefficient of determination (R^2) will be:

$$R^2 = \frac{\sum(Y_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{886207073534,3}{1021076948764,0} = 0.8679$$

As $F > F_{\alpha,k-1,n-k}$, mean $F\ (85.42) > F_{0.05;1;13}\ (4,67)$, the null hypothesis is rejected and the alternative hypothesis that R^2 is statistically significant is accepted.

Next, we will *test the parameters of the regression model*.

To test the parameter α, the following statistical hypotheses are issued:

- H0: $\alpha = 0$ (not statistically significant)
- H1: $\alpha \neq 0$ (is statistically significant)

The t-test will be:

$$t = \frac{a-0}{s_a} = \frac{-309700.51}{123733.63} = -2.50$$

Since $|t| > t_{0.025;13}$ ($2.50 > 2,1604$) the null hypothesis is rejected, the alternative hypothesis that the parameter is statistically significant is accepted.

To test the parameter β, the following statistical hypotheses are issued:

- H0: $\beta = 0$ (not statistically significant)
- H1: $\beta \neq 0$ (is statistically significant)

The t-test will be:

$$t = \frac{b-0}{s_b} = \frac{11.14}{1.2052} = 9.24$$

Since $t\ (9.24) > t_{0.025;13}\ (2.1604)$ the null hypothesis is rejected, which implies that the parameter β is significantly different from 0.

We will determine *the prediction interval for variable y* (GDP) if "Net Investments" will be $x^{2023} = 134000$ million lei (estimated level using a linear trend function). The point estimate will be:

$y^{2023} = -309700.51 + 11.14 \cdot 134,000 = 1182919.27$ *million lei*

This result tells us that if the "Net Investments" will be $x^{2023} = 134000$ million lei, then "GDP" could reach *1182919.27* million lei in 2023.

The prediction interval is:

$$y^{2023} = (a + bx^{2023}) \pm t_{\lambda;n-2} \cdot \sqrt{s^2 \cdot \left(1 + \frac{1}{n} + \frac{(x^{2023} - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}$$

$$y^{2023} = 1182919.27 \pm 2,1604 \cdot 101855.81 \cdot \sqrt{\left(1 + \frac{1}{15} + \frac{(134000 - 100319,9)^2}{7142437084.9}\right)}$$

$y^{2023} = (939324.63;\ 1426513.91)$

Consequently, we can guarantee with a probability of 95% that the prediction interval (for GDP) is between *939324.63* and *1426513.91* million lei.

3. CONCLUSIONS

The evolution of the ”Gross Domestic Product” reveals that, for most of the period, we are recording growth, which is a positive situation. In contrast, the evolution of ”Net Investments” is oscillating, with increases alternating with decreases.

To study the link between ”Gross Domestic Product” and ”Net Investments” we used a unifactorial linear regression model, where the adjustment equation is:

$y_i = -309700.51 + 11.14 \cdot x_i$

After estimating the confidence intervals of the parameters, we can guarantee with a probability of 95% that α falls between -577010.76 and -42390.26, and β between 8.54 and 13.74. Since none of the intervals contain 0, we can guarantee that the parameters are statistically significant.

The linear correlation coefficient through the high value obtained (0.9316) highlights a very strong link between GDP and investments. Testing the significance of the coefficient leads to the conclusion that the hypothesis that the linear correlation coefficient is statistically significant is accepted.

Testing the validity of the model leads us to the conclusion that the alternative hypothesis is accepted according to which the coefficient of determination (R^2) is statistically significant. *Testing the parameters of the regression model* shows that the alternative hypothesis that the model parameters (α and β) are statistically significant is accepted.

The point estimate for ”Gross Domestic Product” is *1182919.27* million lei, and the prediction interval, given that ”Net Investments” will be *134000* million lei, is (*939324.63; 1426513.91*) million lei.

REFERENCES

1. Ciurlău, C., Enea-Smarandache, I., Murăriţa, I., Ciurlău, C.F., Ciobanu, A.M. *ECONOMIC FORECASTING. Theory. Choice test. Applications,* Universitaria Publishing House, Craiova, 2008
2. Dimian, G.C., Dragan, I.M. *The basics of econometrics. Theory. Applications. Case studies*, ASE Publishing House, 2018
3. Georgescu, V. *Econometrics Course*, Sitech Publishing House, Craiova, 2014
4. Murăriţa, I., *Statistics Software Packages*, PIM Publishing House, Iasi, 2018
5. Murăriţa, I., Radu, C., Ionaşcu, C.M. *STATISTICS. Theory and Applications solved in Excel*, PIM Publishing House, Iasi, 2014
6. Radu, C., Ionaşcu, C.M., Murăriţa, I. *THEORETICAL STATISTICS*, Second Edition, Universitaria Publishing House, Craiova, 2009
7. *** www.insse.ro

Annex 1. Evolution of "Gross Domestic Product" and "Net Investments" in Romania, in the period 2008-2022

-millions of lei-

Year	Gross Domestic Product	Net Investments
2008	539834,6	99525,6
2009	530894,4	74939,3
2010	540336,3	72294,7
2011	587203,3	87815,8
2012	621268,7	89092,3
2013	631602,9	80849,0
2014	668876,4	86160,0
2015	712543,6	98888,0
2016	752116,4	96162,9
2017	851619,7	91045,5
2018	959058,6	101702,5
2019	1063794,6	121992,9
2020	1066780,5	116305,5
2021	1189089,8	134019,6
2022	1401345,4	154004,2

Sursa: http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table

Annex 2. Calculations necessary to estimate the parameters of the regression model

Anul	***x_i - Inv***	***y_i - Pib***	***x_i^2***	***y_i^2***	***$x_i * y_i$***	***Y_i***
2008	99525,6	539834,6	9905345055,4	291421395357,2	53727362465,8	798910,5
2009	74939,3	530894,4	5615898684,5	281848863951,4	39784854709,9	525044,9
2010	72294,7	540336,3	5226523648,1	291963317097,7	39063450707,6	495586,8
2011	87815,8	587203,3	7711614729,6	344807715530,9	51565727552,1	668475,6
2012	89092,3	621268,7	7937437919,3	385974797599,7	55350257401,0	682694,5
2013	80849,0	631602,9	6536560801,0	398922223288,4	51064462862,1	590872,7
2014	86160,0	668876,4	7423545600,0	447395638477,0	57630390624,0	650031,7
2015	98888,0	712543,6	9778836544,0	507718381901,0	70462011516,8	791808,3
2016	96162,9	752116,4	9247303336,4	565679079149,0	72325694161,6	761453,6
2017	91045,5	851619,7	8289283070,3	725256113428,1	77536141396,4	704451,1
2018	101702,5	959058,6	10343398506,3	919793398234,0	97538657266,5	823158,9
2019	121992,9	1063794,6	14882267650,4	1131658950989,2	129775388258,3	1049172,7
2020	116305,5	1066780,5	13526969330,3	1138020635180,3	124072439442,8	985821,1
2021	134019,6	1189089,8	17961253184,2	1413934552464,0	159361339360,1	1183137,6
2022	154004,2	1401345,4	23717293617,6	1963768930101,2	215813077250,7	1405745,1
Total	**1504797,8**	**12116365,2**	**158103531677,2**	**10808163992748,7**	**1295071254975,6**	**12116365,2**

Annex 3. Calculations necessary to determine the total variance, the explained variance and the residual variance

Anul	$(x_i-xm)^2$	$(y_i-Y_i)^2$	$(x_i-xm)(y_i-ym)$	$(y_i-ym)^2$	$(Y_i-ym)^2$
2008	630838,3575	67120336979,2	212798799,4	71782776796,7	78272081,1
2009	644172487,5	34216949,3	7026943244,1	76653275812,4	79926530428,9
2010	785409219,4	2002517948,5	7494525179,1	71514194481,1	97450659703,1
2011	156351349,8	6605188419,1	2757823730,4	48644234537,2	19399494944,0
2012	126057953,9	3773127006,1	2093814969,0	34778139661,4	15640802859,2
2013	379114129,5	1658945245,8	3429883885,3	31030506516,8	47039073536,1
2014	200501446,4	355121552,1	1966538555,5	19288009934,4	24877475007,5
2015	2050204,0	6282897767,3	136332597,8	9065721030,2	254381695,9
2016	17280261,0	87182739,2	231298204,4	3095952040,0	2144070625,0
2017	86013629,8	21658600157,4	-406791871,4	1923876798,5	10672251809,8
2018	1911711,8	18468723827,1	209195712,7	22891968392,8	237198102,5
2019	469720951,8	213798764,0	5549100115,5	65554904403,1	58281231621,2
2020	255540899,4	6554431530,8	4140647279,1	67092821280,8	31706565964,8
2021	1135672925,4	35428756,6	12850795839,9	145414185743,7	140910079814,9
2022	2882009077,0	19357587,2	31866368937,6	352346381334,8	357588985340,1
Total	**7142437084,9**	**134869875229,7**	**79559275178,5**	**1021076948764,0**	**886207073534,3**