Modeling the amount of cash in circulation in a country using linear multiple regression

Marina Batova

candidate of economic sciences (PhD in Economics), Department of Informatics and Management, Military University of the Ministry of Defense of the Russian Federation, 14, Bolshaya Sadovaya str., Moscow, Russian Federation

e-mail: batova_m_m@mail.ru

Abstract. The problem of modeling and forecasting the amount of cash in circulation in the country based on the example of the Russian Federation was formulated. Based on the methods of mathematical statistics and econometrics the dominant factors having the most significant effect on the amount of cash in circulation found. The problem in the form of an econometric model represented as a linear multiple regression localized. The quality of the resulting model justified.

1. Introduction

Money circulation in a country is a money flow in the performance of their functions in cash and non-cash form. This money flow serves the sale of goods, as well as non-commodity payments and settlements in the country’s economy. [1]

Despite the rather positive dynamics in the development of the infrastructure of non-cash payments, cash has been and remains the preferred payment instrument in Russia. At the same time, there is an annual increase in the circulation of banknotes and coins [2]. The cash circulation is organized by the Central Bank of the Russian Federation in Russia.

The development of money circulation, the optimal ratio of cash and non-cash money in any country depends on many factors. And the amount of cash in circulation, in turn, affects other important economic indicators, such as, for example, GDP, inflation, etc.

Thus, one of the pressing economic problems is as follows: to identify the factors that have a significant impact on the amount of cash money, model and forecast the amount of cash in circulation in the country.

2. Problem statement and solution methods

The main objective of this work was the study of cash circulation and the creation of a mathematical model suitable to forecast the amount of cash in a country, such as Russia.

For the most complete description, research and modeling of the economic situation under study, various factors shall be considered, and then the most essential and significant of them shall be taken into account and included in the model. In order to solve the problem of modeling the dependence of such an economic variable as “the amount of cash in circulation” on several factors, the author simultaneously used the methods of mathematical statistics and econometrics [3, 4, 5].
To do this, the data of economic and financial statistics for 19 years, covering the period from 1998 to 2016, were first collected, processed and studied. The source of statistics was the official website of the Central Bank of the Russian Federation https://www.cbr.ru/statistics.

And then the modeling of dependence of the amount of cash on several factors was carried out by creating an econometric regression model, namely, a linear multiple regression [3, 4], which has the following form in general population:

\[ Y = \alpha_0 + \alpha_1 \cdot X_1 + \alpha_2 \cdot X_2 + \cdots + \alpha_m \cdot X_m + \varepsilon. \]  

(1)

where \( Y \) – studied economic indicator, dependent variable;
\( X_1, X_2, \ldots, X_m \) – factors affecting \( Y \);
\( m \) – number of factors in a linear multiple regression model;
\( \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_m \) – linear multiple regression parameters;
\( \varepsilon \) – random component that expresses random, insignificant and unaccounted factors in the model.

Then, the task is aimed at finding the equation of multiple regression according to the statistics sample, having the form of:

\[ \hat{Y}(x_1, x_2, \ldots, x_m) = \hat{\alpha}_0 + \hat{\alpha}_1 \cdot x_1 + \hat{\alpha}_2 \cdot x_2 + \cdots + \hat{\alpha}_m \cdot x_m. \]  

(2)

where \( \hat{Y} \) – estimate modeled by the values of the studied economic indicator \( Y \);
\( x_1, x_2, \ldots, x_m \) – statistics of each factor in the sample;
\( \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_m \) – linear multiple regression coefficients, which are the corresponding estimates of parameters of the linear multiple regression model (1).

The least square method is used to create the linear multiple regression equation. Also, it should be noted that due to the multiplicity of factors, certain problems may arise during the model creation. For example, selection of factors and their multicollinearity.

3. Problem solution and modeling results of the amount of cash in circulation

3.1. Four-factor model

Let us introduce the variable \( Y \) - the amount of cash in circulation in Russia, billion roubles - in the problem of modeling the amount of cash in the country. There are many factors effecting \( Y \) in Russia, as in any country. Each of these factors contributes to the amount of cash in circulation in the country: more or less.

However, after studying the statistics of the Central Bank of the Russian Federation for the previous 19 years and based on the economic and mathematical considerations, it was identified 4 main factors effecting \( Y \) for the study:

- \( X_1 \) – volume of precious metals in circulation in Russia, billion roubles.
- \( X_2 \) – volume of securities in circulation, billion roubles.
- \( X_3 \) – amount of funds in the accounts opened with the Central Bank of Russia, billion roubles.
- \( X_4 \) – capital of the Central Bank of Russia, billion roubles.

Then, it was studied the pairwise correlation of \( Y \) with each of four factors. For this purpose, it was created the graphs, being pairwise correlation fields of \( Y \) with each factor (Figure 1).

Then, it was calculated the pairwise correlation coefficients of \( Y \) with each factor, as well as compiled a correlation matrix consisting of pairwise correlation coefficients of \( Y \) with each factor and the correlation coefficients between the factors [3, 4] (Table 1).

Based on economic considerations and pairwise correlation coefficients, as well as other information, it was selected the factors advisable to be included in the equation.

The basic principle is: the greatest factor correlation with the effective attribute \( Y \), but less correlation of this factor with others. Otherwise, it would be the problem of factor multicollinearity, which would distort the regression results and violate the premises of the least square method (conditions of the Gauss-Markov theorem) [3, 4].
Figure 1. Pairwise correlation fields of $Y$ with factors $X_1, X_2, X_3, X_4$.

In this case, the correlation of $Y$ with each factor is very high, which is undoubtedly a plus. But the correlation between the factors is also very high, which is a minus, as it leads to the multicollinearity problem. (Table 1)

Table 1. Correlation matrix of $Y$ with factors $X_1, X_2, X_3, X_4$.

|       | $Y$    | $X_1$  | $X_2$  | $X_3$  | $X_4$  |
|-------|--------|--------|--------|--------|--------|
| $Y$   | 1      | 0.88   | 0.84   | 0.95   | 0.78   |
| $X_1$ | 0.88   | 1      | 0.89   | 0.86   | 0.78   |
| $X_2$ | 0.84   | 0.89   | 1      | 0.88   | 0.78   |
| $X_3$ | 0.95   | 0.83   | 0.86   | 1      |        |
| $X_4$ | 0.78   | 0.97   | 0.88   | 0.78   | 1      |

Let us calculate the first equation by including the factors $X_1, X_2, X_3, X_4$, and estimate the obtained coefficients of multiple linear regression (2) $a_j$ ($j$=1, 2, 3, 4) using the Student's t-test by calculating the observed value $t_{\text{calc}}$ for each of them.

$$Y(x_1, x_2, x_3, x_4) = 416.853 + 3.796 \cdot x_1 - 0.095 \cdot x_2 + 0.433 \cdot x_3 - 0.878 \cdot x_4$$

(3)

$$t_{\text{calc}} \text{ in absolute value } 0.67 \quad 7.9 \quad 0.04 \quad 8.41 \quad 6.25$$

It should be noted that inadequate results were obtained for factors $X_2$ and $X_4$: their coefficients are negative, whereas the pairwise correlation coefficients and the correlation field show a positive relationship. This discrepancy is due to the multicollinearity problem. It is required a correct set of factors. To this end, the statistical significance of each regression coefficient will be tested and it will be decided on the factor to be excluded.

Let us compare the calculated value of Student's t-test $t_{\text{calc}}$ with table at significance level $\alpha=0.05$ with $(n-m-1)=(19-4-1)=14$ freedom degrees: $t_{\text{tabl}}=2.14$. 

3
The coefficient at $X_2$ cannot be recognized as statistically significant, therefore factor $X_2$ is excluded from the study. Thus, the factors $X_1$, $X_3$, $X_4$ shall be left for further analysis.

3.2. Three-factor model

Then a similar procedure is carried out: it is calculated the second equation with factors $X_1$, $X_3$, $X_4$, as well as it is evaluated the obtained regression coefficients (2) of this equation:

$$y(x_1, x_2, x_4) = 390.444 + 3.797 \cdot x_1 + 0.431 \cdot x_3 - 0.88 \cdot x_4$$

It should also be noted in this equation that there are inadequate results on the factor $X_4$: its coefficient is negative, whereas the pairwise correlation coefficient and the correlation field show a positive relationship. It is required to repeat the factor selection process. To this end, the statistical significance of each regression coefficient will be tested and it will be decided on the factor to be excluded again.

Let us compare the calculated values $t_{calc}$ with table at significance level $\alpha=0.05$ with $(n-m-1)=(19-3-1)=15$ freedom degrees $t_{tabl}=2.13$.

As can be seen, all the coefficients turned out to be statistically significant. But not all of them are consistent with the current economic situation under study. Therefore, in order to definitively establish the factors to be left in the model and to be excluded from it, it is necessary to study the multicollinearity of factors in more detail.

Let us compose the matrix of inter-factor correlation and calculate its determinant (Table 2).

|       | $X_1$ | $X_3$ | $X_4$ |
|-------|-------|-------|-------|
| $X_1$ | 1     | 0.83  | 0.97  |
| $X_3$ |       | 1     |       |
| $X_4$ |       |       | 1     |

The determinant of this matrix is: $det r_X=0.018448$, that is, close to zero. This means that the multicollinearity problem is great, and its consequences greatly distort the results.

To solve this problem, let us carry out the following analytical considerations. On the one hand, the factor $X_1$ has the greatest relationship with factors $X_3$ and $X_4$ (correlation coefficients are 0.83 and 0.97, respectively) and thereby creates a multicollinearity problem, which is stronger than the others, therefore, the factor $X_1$ shall be excluded from the study. But on the other hand, the results of factor $X_4$ are worse in the sense that its coefficient is not consistent with the economic sense, and its statistical significance is lower than the same in other factors. Therefore, to make a final decision on a set of factors, it is necessary to test two two-factor models: one - with factors $X_3$ and $X_4$, and another one – with factors $X_1$ and $X_3$.

3.3. Two-factor model

3.3.1 Two-factor model with factors $X_3$ and $X_4$. Let us compose the inter-factor correlation matrix with factors $X_3$ and $X_4$ (Table 3), and calculate its determinant.

|       | $X_3$ | $X_4$ |
|-------|-------|-------|
| $X_3$ | 1     |       |
| $X_4$ | 0.78  | 1     |

The determinant of this matrix is: $det r_X=0.3916$. Its value is higher than in the three-factor model, and approaches 0.5. This means that although the multicollinearity problem is present, it has decreased. This level can be considered acceptable. It will not greatly distort the results.

Let us calculate the linear multiple regression equation with two factors $X_3$ and $X_4$: 

$$y(x_3, x_4) = \text{coef}_{x_3} \cdot x_3 + \text{coef}_{x_4} \cdot x_4$$
When comparing with the table value of the Student's t-test at the significance level \( \alpha = 0.05 \) with \((n-m-1)=(19-2-1)=16\) freedom degrees \( t_{tabl} = 2.12 \), only one regression coefficient with \( X_3 \) seems to be statistically significant. Such a model is unsuitable.

### 3.3.2 Two-factor model with factors \( X_1 \) and \( X_3 \)

Let us compose a new inter-factor correlation matrix with factors \( X_1 \) and \( X_3 \) (Table 4), and calculate its determinant.

| \( X_1 \) | \( X_3 \) |
|----------|----------|
| \( X_1 \) | 1        |
| \( X_3 \) | 0.83     | 1        |

The determinant of this matrix is: \( \text{det} r = 0.3111 \). Its value is slightly less than in the case of factors \( X_3 \) and \( X_4 \) (Table 3), but the multicollinearity problem is at the same acceptable level.

Let us calculate the linear multiple regression equation with two factors \( X_1 \) and \( X_3 \):  

\[
\hat{y}(x_3, x_4) = 462,613 + 0.852 \cdot x_1 + 0.490 \cdot x_3
\]

When comparing with the table value of the Student's t-test at the significance level \( \alpha = 0.05 \) with 16 freedom degrees \( t_{tabl} = 2.12 \), the regression coefficients for both factors seem to be statistically significant. Regarding free coefficient, the calculations showed that it can be considered statistically significant with a decrease in the requirements for confidence probability, i.e., at a significance level \( \alpha \), more than 0.17.

### 4. Conclusion

The quality of the created two-factor linear multiple regression model (6) is also confirmed by the high Fisher’s F-test - 109.9.

High determination coefficient \( R^2 = 0.924 \) indicates that changes in \( Y \) for 92.4% are subject to cumulative change in factors \( X_1 \) and \( X_3 \).

Acceptable quality and forecasting accuracy is confirmed by a satisfactory level of relative approximation error \( V = 22\% \).

Therefore, it is obtained the linear multiple regression equation (6) of rather good quality, which adequately reflects the existing relationship between the studied economic indicators.

Thus, this model can be considered suitable for further use. It can be used by banks, financial institutions and other interested players in the financial and economic sector as one of the tools for analyzing and forecasting the amount of cash in circulation in the country.

### References

[1] Zavgorodnaya T, Zavadskaya V 2016 Development of monetary circulation in the Russian Federation J. Siberian. Trade & Economic Journal **4(25)** 11-16

[2] Savradym V, Gorbatko Ju 2019 The structural and dynamic analysis of cash circulation in the Russian federation J. Siberian Financial School **1(132)** 95-98

[3] William H. Greene 2018 *Econometric Analysis* (New York: Pearson)

[4] Eliseeva I *et al.* 2014 *Econometrics Textbook for masters*, ed. Eliseeva I (Moscow: Uraite) chapter 2 pp 39-123

[5] Eliseeva I, Egorova I, Curisheva S 2015 *Statistics* (Moscow: Prospect).