ISSN: 0976-3104
ISSUE: Innovations and Economics

ARTICLE

APPLICATION OF THE ECONOMETRIC MODEL AS A MECHANISM OF MANAGEMENT OF SOCIO-ECONOMIC SYSTEMS

Lyudmila Valentinovna Bolshakova[1]*, Alexander Nikolaevich Litvinenko[1], Evgeniya Vladimirovna Baturina[1], Inna Kazemirovna Sidenko[2], Alexander Nikolaevich Ivanov[3], Farid Abdulalievich Dali[3], Grigori Leonidovich Shidlovsky[3]

[1]St. Petersburg University of the Ministry of Internal Affairs of Russia, St. Petersburg, RUSSIA
[2]Russian State Hydro meteorological University, St. Petersburg, RUSSIA
[3]St. Petersburg University of the State Fire Service EMERCOM of Russia, St. Petersburg, RUSSIA

ABSTRACT

The problem of the adequacy test criteria for the econometric model obtained from the sample has been considered. The adequacy testing methods have been shown. Particular attention has been paid to the adequacy testing method for the model using the value of the average approximation error, and significant drawbacks of this method which can lead to erroneous results in the analysis and forecasting have been revealed. In modern conditions, when management decisions are made based on the analysis of statistical, incomplete information, the use of econometric modeling and analysis methods is not only justified but also necessary. Econometric models are applied both at the level of organizations' activities and at the level of planning and analysis of aspects of the economic activity of the region and the country as a whole. The study is aimed at building econometric models to obtain an effective tool for forecasting, analysis and decision making. In the analysis of specific statistical data, the methods of correlation-regression analysis and forecasting are used. The paper highlights and describes the characteristic features of forecasting based on econometric models. Econometric methods and models allow assessing the impact of changes in the internal and external environment on the resulting indicator, analyzing the cause-and-effect relationships between indicators, and performing forecasting. The quality assessment of the model can be obtained not only by the value of the average approximation error but also by the value of the determination coefficient, as well as by verifying the significance of the regression equation by the Fisher's criterion. The simultaneous manifestation of an erroneous assessment of the quality of the regression model for all the listed values of the coefficients and criteria is highly doubtful. Therefore, in order to obtain more objective assessment of the adequacy of the regression model under consideration, it is worth calculating all the coefficients and testing the corresponding statistical hypotheses. Verifying the adequacy of the model can also include checking the feasibility of the assumptions by the least-squares method, which is used to find estimates of the coefficients of the regression equation. If any of the prerequisites is impracticable, the estimates of the coefficients may significantly differ from their actual values.

KEY WORDS
average approximation error, coefficient of determination, coefficient of correlation, Fisher's F-test

INTRODUCTION

Econometric methods and models are integral parts of any modern support system for making economic and managerial decisions. Today, econometric methods are used for diagnostics of the state of the enterprise, when solving problems of management of corporate finance and risks, assessing the efficiency of investment and innovation activities, the value of assets and business, analyzing the dynamics of prices and living standards, as well as assessing the parameters of economic and mathematical models of logistics.

An econometric model, represented by an equation or a system of equations and inequalities, is a mathematical analogue of an object, taking into account all the most important aspects and features of the object functioning, according to which the best option for the development of this object can be found. Obviously, the more detailed the essence and content of the object, the relationship of its elements and their influence on the final result of the activity or the functioning of the object are considered, the more accurate and acceptable for application and implementation in practice will be the solution.

Received: 4 Nov 2020
Accepted: 6 Dec 2020
Published: 11 Dec 2020

Econometric methods allow answering two main questions: what can happen in the future (forecast, foreseeing the development of the economic situation) and how a change in one value can affect another - the task of analysis for managing economic processes.

The construction of an econometric model that explains the relationship between various factors is one of the most important tasks in conducting research both at the micro- and macroeconomic levels. The main element of the model – the sample regression equation – is further used to analyze and forecast the possible values of attributes (factors) describing a particular economic process. However, since the elements of not the entire general population, but only of its part – the sample – are used in the construction of the model, it leads to a question of how the constructed model corresponds to reality. As such, the problem of assessing the quality of the resulting regression equation arises after the model construction. This problem is the most important one in econometric analysis, because the conclusions and forecasts obtained from a low-quality model will be far from reality [1, 2].

***Corresponding Author**
Email:
bolshakoval.v@bk.ru

Various coefficients and statistical hypotheses are used to find out how well the obtained model describes the real situation, and conclusions are drawn about the adequacy of the obtained model, based on their values and results.

IIOAB JOURNAL ISSN: 0976-3104

METHODS

The quality of the regression model is most often assessed by the values of the coefficients of correlation and determination, by the average approximation error, as well as by the results of testing statistical hypotheses about the significance of the entire regression equation by the Fisher's F-test and the significance of individual coefficients by the Student's t-test.

These criteria can be briefly described for a paired linear regression.

Let us assume that a paired linear relationship between two characteristics is examined: the resultant Y and the factor X. A sample of volume n is taken from the general population:

Table 1: Samples

Element number	1	2	3	…	*n*
Value of attribute *X*	x_1	x_2	x_3	…	x_n
Value of attribute *Y*	y_1	y_2	y_3	…	y_n

The sample equation of the paired linear regression $\hat{y} = b_0 + b_1 x$ was found based on the obtained sample, using the Excel application [3]. The quality of the resulting equation can be tested using the following criteria.

A preliminary and lax conclusion about the dependence can be drawn from the correlation field. If the points of the correlation field are located not far from the regression line, then the existence of a linear relationship between the attributes can be assumed, which corresponds to the assumption of the adequacy of the model.

Further, the coefficients of correlation and determination and the average approximation error are also found using the Excel applications [3], and their characteristic properties necessary to test the adequacy of the model are described.

The sample linear coefficient of correlation rs describes the existence and strength of a linear relationship between the attributes. The closer is the value of this coefficient to ± 1, the stronger is the dependence and the more reasons are there to consider the model to be adequate. However, the closer rs is to zero, the weaker the linear dependence is, and the quality of the model significantly decreases.

The sample coefficient of determination R, expressed as a percentage, shows which percentage of the changes in the resulting attribute Y occurs due to the influence of the factor attribute X. It is obvious that if R is less than 50%, then the model cannot be considered adequate.

The average approximation error is one of the simple coefficients that allows to briefly describe how much the simulated process corresponds to the real one.

The average approximation error is understood as either the average absolute approximation error, determined by the formula

$$A_{aa} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\%,$$

or the mean-square approximation error, determined by the formula

$$A_{ms} = \frac{1}{\bar{y}}\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \cdot 100\%,$$

where y_i is the actual values of attribute Y;

$\bar{y}$ is the average value of the observations;

$\hat{y}_i$ is the theoretical values of attribute Y, found by the regression equation; and
n is the sample size.

It is obvious that the values of coefficients Aaa and Ams for the same problem differ. It is generally accepted that if any of these errors does not exceed 14%, the regression model is considered to be of good quality, i.e. is adequate to reality, and real forecasts can be built based on it.

The assumption about the adequacy of the model can be confirmed or refuted by testing the following statistical hypotheses.

The statistical hypothesis about testing the significance of the regression equation generally involves testing the main hypothesis that the resulting regression equation is insignificant (inadequate, of poor quality).

This hypothesis is tested according to the general scheme:

- The main hypothesis is formulated:
  H0: the regression equation is significant.
- 2. The level of significance α is selected.
- 3. The criterion is determined – a random variable distributed according to the Fisher's law (F distribution).
- 4. The sample value of the criterion is found using the following formula:

$$F_s = \frac{r_s^2}{1 - r_s^2} \cdot (n - 2).$$

5. The right-sided critical range is determined, and the critical point Fcr is found according to the Fisher's distribution table with parameters v1 = 1 and v2 = n – 2.
6. A statistical conclusion is made:
If Fs < Fcr, then there is no reason to reject the main hypothesis and recognize the equation as significant. In this case, the poor quality of the model is confirmed.
If Fs > Fcr, then the main hypothesis should be rejected as contradicting the sample data, and the regression equation should be considered significant. In this case, the good quality of the model is confirmed

RESULTS AND DISCUSSION

As noted above, the quality of the model can be judged by the results of testing statistical hypotheses about the significance (difference from zero) of various coefficients – in particular, the coefficient of correlation rg and the coefficient of regression β1. The values rg and β1 are characteristics of the entire general population, while their estimates rs and b1, respectively, are found from the sample data. Such hypotheses can be tested similarly to the hypothesis about the significance of the equation in general, with the corresponding changes in the test criteria. The Student's t-test is used instead of the Fisher's F-test in these cases. The general scheme will be considered only to test the hypothesis about the significance of the coefficient of correlation for the general population, because the results of testing the hypothesis about the significance of the general coefficient of regression completely coincide with the results of testing the above hypothesis about the significance of the equation in general.

- The main and alternative hypotheses are formulated
  H0: rg = 0, H1: rg ≠ 0.
- The level of significance α is selected.
- The criterion is determined – a random variable distributed according to the Student's law (T distribution).
- The sample value of the criterion is found using the following formula:

$$T_s = \frac{r_s}{\sqrt{1 - r_s^2}} \cdot \sqrt{n - 2}.$$

5. The two-sided critical range is determined, and the critical point tcr is found according to the Student's distribution table with a parameter (number of degrees of freedom) v = n – 2.
6. A statistical conclusion is made:

If $|T_s| < |t_{cr.}|$, then there is no reason to reject the main hypothesis and recognize the equation as significant, i.e. different from zero. In this case, the poor quality of the model is confirmed.

If $|T_s| > |t_{cr.}|$, then the main hypothesis should be rejected as contradicting the sample data, and the correlation coefficient is significantly different from zero. In this case, the good quality of the model is confirmed.

Solution

Each of the above methods of testing the adequacy has a number of drawbacks that can lead to both a significant underestimation of the quality of a "good" model and an overestimation of the quality of a "bad" model. Most often, this drawback manifests itself when checking the adequacy of the model using the value of the average approximation error.

IIOAB JOURNAL
ISSN: 0976-3104

Possible erroneous situations when using the average approximation error can be examined using the example of a paired linear regression.

Of course, erroneous situations do not always arise. There are many examples of high-quality models for which the average approximation errors do not exceed 14% and, conversely, there are examples of poor-quality models with approximation errors significantly more than 14%.

However, in practice, there are situations in which the average error values provide an incorrect idea of the quality and properties of the resulting models. Consequently, the conclusions or predictions made using such models will be far from reality. Below are all possible situations confirming the above.

Let us consider three samples of size n = 5. Let us note that results similar to those described below can be obtained for large samples as well.

Example 1. A sample with the following values is examined [Table 2].

Table 2: Values

X	5	4	2	7	9
Y	70	80	93	95	100

For clarity, a correlation field is built first [Fig. 1], using Excel tools [3]:

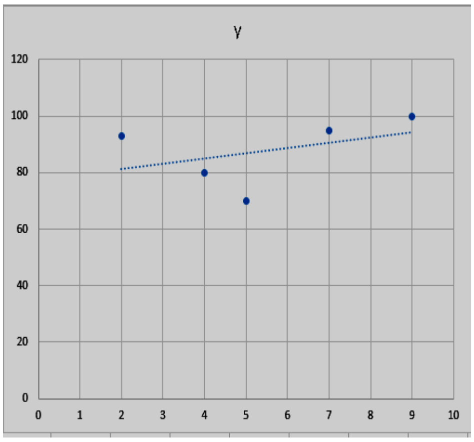

Fig. 1: Correlation field

It can be seen in the correlation field that the linear relationship between the attributes is not strong (the scatter of points around the regression line is too large). Therefore, according to the preliminary conclusion, the paired linear regression model can significantly distort the real picture. Let us find the sample regression equation using the least square method:

$$\hat{y} = 77.28 + 1.91x.$$

The sample coefficient of determination will be equal to R = 17.62, from which it follows that only about 18% of the change in attribute Y is determined by the influence of attribute X. The value of the coefficient of determination confirmed the preliminary conclusion about the poor quality of the model.

A weak linear relationship between the attributes is also indicated by the correlation coefficient rs, which is approximately equal to 0.42.

Let us test the statistical hypothesis about the adequacy of the model according to the Fisher's F-test:
1. H0: the regression equation is insignificant.
2. The generally accepted level of significance α = 0.05 is selected.
3. Fisher's F-test (F distribution).
4. The sample value of the criterion:

$$F_s = \frac{r_s^2}{1-r_s^2}\cdot(n-2) = \frac{(0.42)^2}{1-(0.42)^2}\cdot 3 \approx 0.6418.$$

5. The right-sided critical range is determined, and the critical point Fcr is found according to the Fisher's distribution table with parameters v1 = 1 and v2 = 3:
Fcr = 10.1.
6. A statistical conclusion is made:

Fs = 0.6418 < Fcr = 10.1, which means that there is no reason to reject the main hypothesis and recognize the equation as significant. In this case, the poor quality of the model is confirmed.

Finally, the significance of the coefficient of correlation is tested.

1. H0: rg = 0, H1: rg ≠ 0.
2. Level of significance α = 0.05.
3. Student's t-test (T distribution).
4. The sample value of the criterion:

$$T_s = \frac{r_s}{\sqrt{1 - r_s^2}} \cdot \sqrt{n-2} = \frac{0.42}{\sqrt{1-(0.42)^2}} \cdot \sqrt{3} = 0.8016.$$

5. The two-sided critical range is determined, and the critical point tcr is found according to the Student's distribution table with a parameter (number of degrees of freedom) v = 3: tcr = 3.18.
6. A statistical conclusion is made:
$|T_s| = 0.8016 < |t_{cr.}| = 3.18$, which means that there is no reason to reject the main hypothesis and recognize the equation as significant, i.e. different from zero. In this case, the poor quality of the model is confirmed.

As such, all the main methods of testing the adequacy of the model indicated its poor quality.

However, the following results were obtained when calculating the average approximation errors:

Aaa = 10.62; Ams = 11.40,

which indicated the good quality of the model.

Thus, there was a significant underestimation of the average approximation errors as a result of which the model could be recognized as adequate to reality. In this case, the forecasts made using this model would be far from reality.

Example 2. A sample with the following values is examined [Table 3]:

Table 3: Values

X	5	4	2	7	9
Y	0.3	0.2	0.1	0.8	0.9

A study similar to the first example is conducted for this sample. The correlation field is as follows [Fig. 2]:

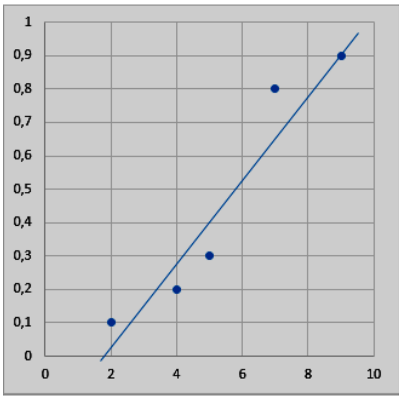

Fig. 2: Correlation field.

It can be seen from the correlation field that the linear relationship between the attributes is quite strong (a scatter of points around the regression line is small). Therefore, according to the preliminary conclusion, the paired linear regression model can be considered adequate to reality. The sample regression equation is found using the least square method:

$$\hat{y} = -0.24 + 0.12x.$$

The sample coefficient of determination will be equal to R = 91.98, from which it follows that only about 92% of the change in attribute Y is determined by the influence of attribute X. The value of the coefficient of determination confirmed the preliminary conclusion about the very good quality of the model.

A strong linear relationship between the attributes is also indicated by the correlation coefficient rs, which is approximately equal to 0.96.

IIOAB JOURNAL
ISSN: 0976-3104

Let us test the statistical hypothesis about the adequacy of the model according to the Fisher's F-test:

1. The main hypothesis H0 is formulated: the regression equation is insignificant.
2. The generally accepted level of significance α = 0.05 is selected.
3. The criterion is determined – a random variable distributed according to the Fisher's law (F distribution).
4. The sample value of the criterion is found:

$$F_s = \frac{r_s^2}{1 - r_s^2} \cdot (n - 2) = \frac{(0.96)^2}{1 - (0.96)^2} \cdot 3 \approx 34.4.$$

5. The right-sided critical range is determined, and the critical point Fcr is found according to the Fisher's distribution table with parameters v1 = 1 and v2 = 3:
Fcr = 10.1.
6. A statistical conclusion is made:

Fs = 34.4 > Fcr = 10.1, which means that the main hypothesis should be rejected as contradicting the sample data, and the regression equation should be considered significant. In this case, the good quality of the model is confirmed.

It can also be made sure that the significance of the coefficient of correlation is confirmed.

As such, all the main methods of testing the adequacy of the model indicated its adequacy to reality.

However, the following results were obtained when calculating the average approximation errors:

Aaa = 35.02; Ams = 20.08,

which indicated the poor quality of the model – therefore, it could not be applied for further analysis. As such, there was a significant overestimation of the values of the average approximation errors, as a result of which the model adequate to reality was rejected.

Example 3. A sample with the following values is examined [Table 3]:

Table 3: Values

X	5	4	2	7	9
Y	0.3	0.2	0.1	0.8	0.9

A study similar to the first and second examples is conducted for this sample.
The correlation field is as follows (Figure 3):

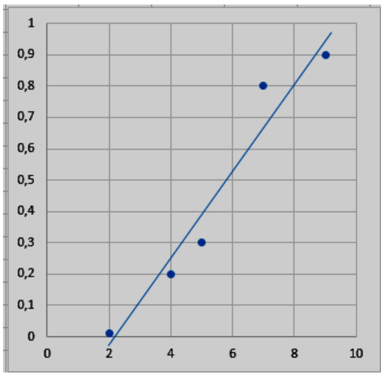

Fig. 3: Correlation field.

It can be seen from the correlation field that the linear relationship between the attributes is quite strong (a scatter of points around the regression line is small). Therefore, according to the preliminary conclusion, the paired linear regression model can be considered adequate to reality. The sample regression equation is found using the least square method:

$$\hat{y} = -0.31 + 0.14x.$$

The sample coefficient of determination will be equal to R = 94.78, from which it follows that only about 95% of the change in attribute Y is determined by the influence of attribute X. The value of the coefficient of determination confirmed the preliminary conclusion about the very good quality of the model.

A strong linear relationship between the attributes is also indicated by the correlation coefficient rs, which is approximately equal to 0.97, and the hypothesis about the significance of the general coefficient of correlation will also be confirmed.

Let us test the statistical hypothesis about the adequacy of the model according to the Fisher's F-test:

1. The main hypothesis H0 is formulated: the regression equation is insignificant.
2. The generally accepted level of significance α = 0.05 is selected.
3. The criterion is determined – a random variable distributed according to the Fisher's law (F distribution).
4. The sample value of the criterion is found:

$$F_s = \frac{r_s^2}{1 - r_s^2} \cdot (n-2) = \frac{(0.97)^2}{1-(0.97)^2} \cdot 3 \approx 54.42.$$

5. The right-sided critical range is determined, and the critical point Fcr is found according to the Fisher's distribution table with parameters v1 = 1 and v2 = 3:
Fcr = 10.1.
6. A statistical conclusion is made:

Fs = 54.42 > Fcr = 10.1, which means that the main hypothesis should be rejected as contradicting the sample data, and the regression equation should be considered significant. In this case, the good quality of the model is confirmed.

As such, all the main methods of testing the adequacy of the model indicated its adequacy to reality.

However, the following results were obtained when calculating the average approximation errors:

Aaa = 102.3; Ams = 17.96,

which indicated the poor quality of the model – therefore, it could not be applied for further analysis. As such, there was a significant overestimation of the values of the average approximation errors, as a result of which the model adequate to reality was rejected.

The analysis of the third sample indicates that the average approximation error can be greater than 100% for a sufficiently good quality model.

The above regression equations for all three samples were built using the least square method, which gave the best estimates of the coefficients of regression. However, examples of samples can also be provided for which the model of the regression equation built using the least square method gives larger average approximation error than the model of the equation with randomly selected coefficients, which clearly contradicts the theory [8-10].

Let us indicate some general cases for which the values of the average approximation errors can be overestimated or underestimated.

It is easy to see that the mean-square approximation error Ams depends significantly on the average value of observations of attribute Y. If the average value turns out to be close enough to zero, then the value of Ams can increase significantly, which will lead to an incorrect assessment of the quality of the constructed model.

The values of the average absolute approximation error Aaa to a large extent depend on the sample values of attribute Y. If any of these values turn out to be close to zero, then Aaa will be overestimated, and the quality of the model will also be assessed incorrectly.

A significant underestimation of both errors can occur in the case when the values of attribute Y are large enough, for example.

CONCLUSION

An incorrect assessment of the quality of the model can result not only from the value of the average approximation error but also from the value of the coefficient of determination, as well as when testing the significance of the regression equation using the Fisher's F-test. However, the latter is extremely rare. Moreover, the simultaneous manifestation of an erroneous assessment of the quality of the regression model for all the listed values of the coefficients and criteria is unlikely. Therefore, it is worth estimating all the coefficients and testing the corresponding statistical hypotheses in order to obtain a more objective assessment of the adequacy of the considered regression model. Finally, it must also be stressed that testing the adequacy of the model can also include testing the feasibility of the assumptions of the least square method, which is used to find estimates of the coefficients of the regression equation. If any of the prerequisites is unsatisfied, the estimates of the coefficients may differ significantly from their actual values. However, the discussion of this problem is beyond the scope of this article and will be considered in subsequent works of the authors. Econometric methods are currently one of the tools for solving problems of analysis and forecasting of economic systems. A well-built econometric model based on a reliable analysis of the existing economic data allows predicting and controlling the economic situation, as well as elaborating options for future development.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENTS

None.

FINANCIAL DISCLOSURE

None.

REFERENCES

[1] Dougherty C. [1999] Introduction to Econometrics. Moscow: INFRA-M.

[2] Eliseeva II. [2007] Ekonometrika [Econometrics]: Textbook. Moscow: Finance and Statistics.

[3] Lavrenov SM. [2003] Excel: Sbornik primerov i zadach [Excel: Collection of examples and tasks]. Moscow: Finance and Statistics.

[4] Bolshakova LV, Yakovleva NA. [2018] Mathematical and statistical methods for processing experimental data in scientific research: guidelines, in 3 parts. Part 2. SPb.: Publishing house of the St. Petersburg University of the Ministry of Internal Affairs of Russia.

[5] Xue J, Ji X, Zhao L, Yang Y, Xie Y, Li D, Wang C, Sun W. Cooperative econometric model for regional air pollution control with the additional goal of promoting employment. Journal of Cleaner Production. DOI: 10.1016/j.jclepro.2019.117814

[6] Basu R, Ferreira J. [2020] Understanding household vehicle ownership in Singapore through a comparison of econometric and machine learning models. Transportation Research Procedia, 48: 1674-1693.

[7] Yang Y, Schmidt P. [2020] An econometric approach to the estimation of multi-level models. Journal of Econometrics, DOI: 10.1016/j.jeconom.2020.04.012

[8] Amusan LM, Charles AK, Adeyemi E, Joshua O, Raphael OA. [2018] Data on expert system-econometric entropy informatics model for adjudicating residential building project costs, Data in Brief. 20: 1721-1729.

[9] Chernyshov KR, Jharko EPh. [2016] Entropy Criteria in the Econometric Model Identification. IFAC-PapersOnLine. 49(12): 827-832.

[10] Morland C, Schier F, Janzen N, Weimar H. [2018] Supply and demand functions for global wood markets: Specification and plausibility testing of econometric models within the global forest sector. Forest Policy and Economics. 92: 92-105.