Multiple correlation-regression analysis of the impact of major factors on oil production

*Klara Gabdrakhmanova*¹, *Gulnara Izmailova*^{1,*}, *Lilia Samigullina*¹, and *Muhammad Majidov*²

¹Branch of Ufa State Petroleum Technological University in the City of Oktyabrsky, Russia

²FSAEI HE “Kazan Federal University”, Russia

Abstract. This article discusses the multiple regression analysis techniques to determine the effectiveness of the factors used. The study examines the various relationships between the elements. It is important to identify which factor will be the most important when selecting wells to determine the amount of oil recovery. Nowadays, the most important problem in the fields of Tatarstan and Bashkortostan is the depletion of deposits. To maintain the profitability of mining companies, therefore, the issue of preparing new reserves remains relevant. This process involves high costs and risks. For a more reliable picture, it is crucial to determine the most relevant factors. The use of the triad of studies proposed by the authors makes it possible to more reliably determine the effectiveness of oil companies. The initial data are direct measurements and methods of mathematical statistics that allow more accurate predictions. Statistical analysis made it possible to identify the parameters on which the effectiveness of the factors depends. In domestic practice, the assessment of resources and reserves of hydrocarbons is usually made by deterministic methods, while abroad the statistical method is used. When studying the relationships between objects, the analyst should be interested not only in the presence and quantitative assessment of the relations but also in the form and relationship of the effective and factor characteristics, its analytical expression. Correlation and regression analysis help to solve these problems. Correlation analysis aims to measure the tightness of the relationship between the varying variables and to evaluate the factors that have the greatest impact on the resulting trait. Regression analysis is designed to select the form of the relationship, to determine the calculated values of the dependent variable (the effective feature) [1]. For the factor analysis, we used data on the oil industry published in the annual statistical collections of Rosstat, as well as specialized periodicals for ten years.

1 Introduction

The fields of Tatarstan and Bashkortostan are characterized by significant oil reserves depletion nowadays. To conduct exploratory drilling, one needs to know the factors that are fundamental. In Russia, the main method is the deterministic method, which is more

* Corresponding author: gulena-86@mail.ru

expensive both from an economic point of view and in terms of time costs. In foreign practice, preferences are given to statistical forecasts.

The theoretical and practical aspects of the fuel and energy complex sustainable development are widely discussed by the scientific community. The problems of economics and the oil complex management in the country and regions are reflected in the works of such researchers as A.O. Amosov, [2] V.F. Dunaev [3], M.N. Isyanbaev [4], and others. The works of I.V. Akimochkina [5], V.I. Vernadsky, O.M. Prishchepa, V.A. Skorobogatov, G.I. Schmal are devoted to the subsoil hydrocarbon potential of our country development. The works of P.G. Egorin, A.D. Dumnov, A.D. Boriskin, G.N. Romashkina, M.Yu. Gordonov are devoted to the statistical study of modern problems of geological exploration and the evaluation of natural resources.

Despite a sufficient amount of research in this area, the issues of modeling and forecasting using statistical forecasting are not sufficiently covered. The multiple correlation-regression method for calculating oil resources is based on the regression equation, the coefficient of determination, where each parameter involved in the resource calculation formula is considered as a random variable, and the resource values are considered as a function of these random parameters.

The main research method employed in this work is a multiple-regression analysis based on the results of work for 2015-2018 [1,2,3,4,5] of large oil companies of the Russian Federation. In order to optimize the risks, a multiple correlation and regression analysis was performed, using multiple regression, performed with the aim to select the optimal factors. Such methods should include, first of all, the so-called probabilistic-statistical methods of research. This is due to the fact that the amount of oil production is influenced by many random factors. Randomness does not allow us to describe phenomena within the framework of deterministic models, because it manifests itself as insufficient regularity in mass phenomena and, therefore, does not allow us to reliably predict the occurrence of certain events. However, when studying such phenomena, certain patterns are found. The correlation analysis aims to measure the closeness of the relationship between the varying variables and to evaluate the factors that have the greatest impact on the effective feature. In their research, the regression and correlation analyses, which make up the probabilistic-statistical method, are considered.

Regression analysis is aimed at selecting the form of the relationship, to determine the calculated values of the dependent variable (the effective feature) [6].

By multiple correlation and regression analysis, the authors understand an analysis that allows us to obtain an empirical relationship between the results of observations and independent variables based on a small number of planned experiments in the form of a functional dependence of varying degrees, which takes into account the separate influence of individual parameters, as well as their joint action.

For this analysis, we used data on the oil industry published in the annual statistical collections of Rosstat [2-7], as well as specialized periodicals [8] for ten years.

In the studies the authors, based on multiple correlation and regression analysis, considered different factors, and identified, from their point of view, the most important ones.

The calculations were performed using the "Analysis Package" tool of the "Data Analysis" add-in of the Microsoft Excel PPP. The results of the main statistical characteristics for all data sets are presented in the summary Table. 1. The most important indicators of this Table 1 are the standard deviation and the coefficient of variation, since they indicate the uniformity of the studied information. The standard deviation shows the absolute deviation of individual values from the arithmetic mean, and the coefficient of variation shows the relative measure of the deviation of individual values from the arithmetic mean.

The issues of modeling and forecasting the oil reserves development are not fully covered in the context of limited access to the performance indicators of oil companies. The solution to these issues will greatly contribute to the further improvement of the system of analysis and assessment of factors influencing the oil complex development, and these factors determine the subject of the research.

2 Research methods

The research uses a triad of factors, such as: the dependence of oil production on the introduction of new wells, production, exploration drilling, research, the proposed authors allow us to more reliably determine the effectiveness of oil companies.

The initial data are direct measurements and methods of mathematical statistics, which make it possible to make more accurate forecasts.

Despite quite a lot of research in this area, the issues of modeling and forecasting, using statistical forecasting, are not sufficiently covered. The multiple correlation-regression techniques for calculating oil resources are based on the regression equation, the coefficient of determination, where each parameter involved in the resource calculation formula is considered as a random variable, and the resource values are considered as a function of these random parameters.

To identify the impact of natural resource and production potential indicators on oil production, we use the multiple regression analysis based on the results of large oil companies of the Russian Federation performance in 2015-2018. Let us estimate the signification of the paired and multiple regression equations using Fisher's F-test. The value of the determination coefficient is one of the criteria for assessing the quality of the model.

Based on the considered data [2,3,4,5,9,10,11], the authors calculated and analyzed mathematical models of correlation and regression analysis, calculated the determination coefficient, and compared the Fisher criterion for the practical and theoretical values of the development of the oil and gas complex in the Russian Federation.

To perform the analysis, we constructed regression equations for various factors (see Table 1):

Table 1. regression equations for various factors.

No	Calculation Sage	R ²	F _{fact}	F _{table}	Oil Extraction Variation
1	First	0.720	8.6	4.7	52%
2	Second	0.742	9.8	4.7	55%
3	Third	0.465	2.2	4.7	22%
4	Sum	0.542	3.8	3.4	69,2%

At the first stage, a linear correlation-regression model was built for the volume of oil production from the factor of new wells commissioning, which is represented by the following parameters for 2015:

$$\tilde{y}_x = 9689.3 + 65.2x \quad (1)$$

where $\tilde{y}_x$ is oil production volume, thousand tons;

x is the number of new wells commissioned, units;

$R=0.270$; $F_{fact}=8.6$; $F_{table}=4.7$ at a significance level of 0.05.

From the regression coefficient value of equation (1), it follows that an increase in well commissioning by one unit will lead, on average, to an increase in oil production by 65.2 thousand tons.

Basing on the regression equation, we can conclude that the relationship between oil production and the number of commissioned wells is strong, positive based on Fisher's F-criterion. The value of the determination coefficient shows that the change in the volume of oil production of the analyzed companies depends by 52% on the change in the new wells commissioning.

Now let's build a model reflecting the oil production dependence on the rate of production drilling. This indicator reflects the production potential of oil companies. Let us consider this model:

$$\tilde{y}_x = 19348.6 + 17.1x \quad (2)$$

where $\tilde{y}_x$ is oil production volume, thousand tons;

x is production drilling, thousand m;

$R=0.742$; $F_{fact}=9.8$; $F_{table}=4.7$ at a significance level of 0.05.

The regression coefficient in the model (2) turned out to be higher than in model (1). The strength of the relationship and the reliability of the model have increased. Analyzing these two models, we conclude that with an increase in production drilling by 1 thousand meters, oil production increases on average by 17.1 thousand tons in 2015 and by 17.9 thousand tons in 2016.

The considered factor was 55% of the variation in oil production in 2016.

At the third stage, a model of the oil production dependence on the level of exploration drilling was built. This factor reflects the natural resources potential of this type of activity:

$$\tilde{y}_x = 28091.5 + 257.8x \quad (3)$$

where $\tilde{y}_x$ is oil production volume, thousand tons;

x is exploration drilling, thousand meters;

$R=0.465$; $F_{fact}=2.2$; $F_{table}=4.7$ at a significance level of 0.05.

In the resulting model (3), the strength of the relationship between the factorial and effective indicators is weaker than in the previous models. The relationship is moderate, positive, the equation is significant. It follows from the equation that with an increase in exploration drilling by 1 thousand meters, oil production increases on average by 257.8 thousand tons. This factor accounts for 22% of the variation in oil production in 2015.

At the next stage, multiple correlation-regression models were built, reflecting the joint influence of all three factors in 2015:

$$\tilde{y}_x = -5380.3_{*(1)} + 232.3x_{1(2,3)} - 28.5x_{2(-1.8)} - 492.8x_{3(-1.2)} \quad (4)$$

where $\tilde{y}_x$ is oil production volume, thousand tons;

x_1 is well commissioning, units,

x_2 is production drilling, thousand meters;

*L is Student's t-test.

x_3 is exploration drilling, thousand meters.

A close relationship was revealed between the effective indicator and factorial signs, the model is stable according to Fisher's F-criterion.

$R=0.592$; $F_{fact}=3.8$; $F_{table}=3.4$ at a significance level of 0.05.

The variation in oil production by 69.2% depends on changes in the factors included in the model. According to the Student's L-criterion, the most significant factor x_1 the commissioning of wells, ($t_{fact}=2.3$ and $t_{table}=2.2$) at a significance level $\alpha=0.05$.

Thus, the factor of well commissioning is the determining factor in the three considered factors.

Application of model (4) makes it possible to determine the reserves for increasing oil production depending on the reserves and the efficiency of using the factors under study. The proposed methodology allows us to establish the factors of increasing the total efficiency of oil production.

To obtain statistically significant estimates, we will construct a regression model of oil production based on the data of large companies for 2015-2018.

$$\tilde{y}_x = -380.3_{(-0.5)} + 157.7x_{1(3.70)} - 12.4x_{2(-1.08)} - 321.9x_{3(-2.98)} \quad (5)$$

$$R^2 = 0.757; F_{fact}=3.54$$

In model (5), the statistical significance of factors x_1 and x_3 of the model as a whole has increased.

By comparing the actual and calculated oil production data for this model (5), the effectiveness of the factors used can be determined.

All the results are included in Table 1 and analyzed.

Let us analyze the results of the obtained equation (see Table 1). An increase in the introduction of the wells into development by one unit will lead to an average increase in oil production by 65.2 thousand tons. According to the coefficient of determination, we can conclude that the relationship between oil production and the number of wells put in is strong, positive based on Fischer's F-test.

Let's turn to the second stage, where the parameters of the study are taken as the oil production dependence on the production drilling indicator. The regression coefficient in the model at the second stage (see Table 1) was higher than at the first stage. The tightness of the connection and the reliability of the model have been improved.

Now while analyzing these two stages, we conclude that with an increase in production drilling by 1 thousand meters, oil production increases on average by 17.1 thousand tons in 2015 and by 17.9 thousand tons in 2016.

At the third stage, the dependence of oil production volumes on the exploration drilling level was considered. From the resulting equation, we conclude that the tightness of the relationship between factorial and resultant features is weaker than in previous models. The relationship is moderate, positive, and the equation is significant. It follows from the equation that with an increase in exploration drilling by 1 thousand meters, oil production increases on average by 257.8 thousand tons. This factor accounts for 22% of the variation in oil production in 2015.

And now let us consider the impact of all three factors on the increase in oil production. According to Table 1 and Figure. 1, it can be concluded that the factor of well commissioning is the determining factor in the factors considered by us, such as the dependence of oil production on the new wells commissioning, production drilling, as well as on exploration drilling on production drilling.

Fig. 1. The relationships between the stages in the conducted studies.

3 Discussions

The considered models make it possible to determine the effectiveness of the factors used and are one of the methods of reliable forecasting. When considering the relationship between the new wells commissioning and the amount of oil produced, as well as production drilling and production, and exploration drilling and the amount of production, the regression equations show a clear advantage of the relationship between the commissioning of new wells and the amount of oil produced. When planning the development of new fields, the use of multiple regression analysis has undoubtedly advantages. The results of multivariate correlation and regression analysis have important scientific and practical value. This is manifested in the fact that factor analysis is significantly deepened, the place and role of each factor in the formation of the level of the indicator under study are established, plans, forecasts, and management decisions are more reliably justified, and the results of the activities of enterprises, industries, and regions are more accurately evaluated.

4 Results

1. Based on the regression equation we can conclude that the relationship between oil production and the number of wells commissioned is strong, positive based on Fisher's F-criterion.
2. The value of the determination coefficient shows that the change in the volume of oil production by the analyzed companies by 52% depends on the change in the commissioning of new wells.
3. Using a triad of factors allows you to more reliably predict the effectiveness of the factors used.

References

1. T.N. Agapova, *Methods for the statistical study of the complex system's structure and its changes* (1996)

2. S.A. Ayvazkhyan, V.S. Mkhitarian, *Applied statistics and foundations of econometrics* (1998)
3. V.N. Afanasyev, M.M. Yuzbashev, *Time series analysis and forecasting: a textbook for students of higher educational institutions studying in the specialty 080601 "Statistics" and other economic specialties* (2010)
4. I.V. Burenina, V.V. Biryukova, S.A. Zats, *Increasing the economic efficiency of an oil and gas production enterprise (theory and methodology)* (2010)
5. A.O. Amosov, *Economist* **6**, 10-16 (2010)
6. V.F. Dunaev, N.P. Epifanova, V.A. Shpakov, *The economics of oil and gas industry enterprises* (2004)
7. K.F. Gabdrakhmanova, G.R. Izmaylova, L.Z. Samigullina, *IOP Conference Series: Materials Science and Engineering* **952(1)**, 012045 (2020)
8. K.F. Gabdrakhmanova, G.R. Izmaylova, L.Z. Samigullina, *Information technology and nanotechnology (ITNT-2020)*, 282-286 (2020)
9. K. F. Gabdrakhmanova, G.R. Izmaylova, L.Z. Samigullina, *IOP Conference Series: Materials Science and Engineering*, **860(1)**, 012008 (2020)
10. K. F. Gabdrakhmanova, G.R. Izmaylova, L.Z. Samigullina, *International Scientific Conference on Philosophy of Education, Law, and Science in the Era of Globalization, PELSEG 2020*, 136-139 (2020)
11. K.F. Gabdrakhmanova, G.R. Izmaylova, L.Z. Samigullina, *E&ES* **459(4)**, p 042080 (2020)