Economic Forecasting Model in Practice Using the Regression Analysis: The Relationship of Price, Domestic Output, Gross National Product, and Trend Variable of Gas or Oil Production

Ashiquer Rahman, Ummey Salma, Afrin Jannat

Open Science Index, Mathematical and Computational Sciences Vol:18, No:11, 2024 publications.waset.org/10013875.pdf

***Abstract*—**Recently, oil has become more influential in almost every economic sector as a key material. As can be seen from the news, when there are some changes in an oil price or Organization of the Petroleum Exporting Countries (OPEC) announces a new strategy, its effect spreads to every part of the economy directly and indirectly. That's a reason why people always observe the oil price and try to forecast the changes of it. The most important factor affecting the price is its supply which is determined by the number of wildcats drilled. Therefore, a study in relation between the number of wellheads and other economic variables may give us some understanding of the mechanism indicated the amount of oil supplies. In this paper, we will consider a relationship between the number of wellheads and three key factors: price of the wellhead, domestic output, and Gross National Product (GNP) constant dollars. We also add trend variables in the models because the consumption of oil varies from time to time. Moreover, this paper will use an econometrics method to estimate parameters in the model, apply some tests to verify the result we acquire, and then conclude the model.

***Keywords*—**Price, domestic output, GNP, trend variable, wildcat activity.

I. Introduction

THE wildcats are wells drilled to find and produce oil and/or gas in an improved area or to find a new reservoir in a field formally discovered to be gas or oil producing, or to increase the size of a known gas or oil reservoir [1]. Moreover, the number of wildcats drilled depends on many factors such as demand for oil, energy's price and OPEC policy etc. If demand for oil is high, the oil production and its supply increase. In general, wildcat drilling has decreased over the past ten years, especially in developed nations like the USA, Canada, and UK. Activities have been hurt, specifically, by the decline in oil prices and competition from unconventional oil and gas wells, such shale, that are less expensive to drill.

There are several connections between domestic output, the price at the wellhead, and the GNP. A number of factors, including supply and demand, production cost, market dynamics, and geopolitical events, affect this price. Oil production nations may benefit financially from high wellhead price but on the other hand low wellhead prices can have an effect on investment and profitability in the oil and/or gas sector. A nation's employment and economic growth are positively impacted by rising domestic output. More output has the potential to boost economic activity and revenue. It comprises net revenue from outside the nation (such as earnings made by individuals or businesses operating there) and domestic output (production carried out within the nation). Wellhead prices and GNP have a substantial but indirect link. Rising wellhead prices result in higher revenue for energy-producing nations, which boosts their GNP. On the other hand, changes in wellhead prices have the potential to affect a nation's GNP growth and overall economic performance. The Global events, policy changes, technology breakthroughs, and economic cycles all have an impact. For instance, when wellhead prices are high, domestic output might rise as a result of energy companies' investments in production and exploration. This in turn has a favorable impact on GNP. On the other hand, domestic output and GNP can suffer during recessions or drops in wellhead prices. Indeed, the price at the wellhead affects domestic output, which raises the GNP of a nation.

In the paper, we focus on an a priori rationale, estimate the parameters of the model, their standard errors, and obtain R^2 and $\bar{R}^2$; four main factors: price at the wellhead, domestic output, GNP, time trends; and specification and expectation of the model.

II. Literature Review

The relationship of the number of wildcats drilled, price at the wellhead, domestic output and GNP are very complex. We could learn more about how variations in price, domestic output, and GNP affect the quantity of wildcats' activities by using a regression model. There is a dearth of empirical statistical research in the published literature that could provide a solution. The finding of past exploration drilling offers a chance to evaluate the precision of geoscience interpretations and technology by comparing predrill predictions to postdrill outcome [2]. The rate at which wildcat wells were drilled was strongly correlated with the discovery expectation of the exploration operators and small additional variations in the wildcat drilling rate were explained by the price/cost ratio and target-depth variables [2]. Drilling challenging wells requires a

Dr. Ashiquer Rahman is Assistant Professor and Ummey Salma and Afrin Jannat are Lecturers with the School of Economics, ZNRF University of Management Sciences (ZUMS), Bangladesh (e-mail: ashiq_firoza@gmail.com, ummey22salma@gmail.com, afrinjannat.economics@gmail.com).

combination of drilling analytics and comprehensive simulation to prevent poor drilling performance and avoid drilling issues for the upcoming drilling campaign [3]. Although there was a modest decline in investment and drilling effort, the success rate of exploration wells, particularly deep water exploration wells rose dramatically, and the newly found reserves showed a slight rise over the previous year [4]. In the last decade, offshore oil and gas discoveries, especially those made in deep and ultra-deep waters, have become the main growth pole of global oil and gas resources [5]. Although during the COVID-19 there was a modest decline in investment and drilling workload, the success rate of exploration wells—particularly deep-water exploration wells—rose dramatically, and the newly discovered reserves showed a slight increase over the previous year. The identification and the detection of the drilling vibration are feasible, and an early manual intervention allows vibration mitigation [6].

III. Frameworks

A. Research Aim

The research aims to define the single equation regression model, and identify the relationship of Price, Domestic Output, GNP, and Trend Variable perspective of Wildcat activity. The aim also is to demonstrate the offer an a priori rationale and estimate the parameters of the model; their standard errors, and obtain R^2 and $\bar{R}^2$ using the EVIEWS program to the time-series data. The paper addresses the question of what is the relationship of Price, Domestic Output, GNP, and Trend Variable perspective of wildcat activity. The question is related to the adoption of blanching economic system and fulfillment the economic disparity

B. Data Source and Methodology

We used an annual time-series data of oil production from 1948 to 1978 using EVIEWS Programs and Single Equation Regression Model. We used the secondary data collected from [7, p.236].

C. Formulation of General Model

The simple single equation model is:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 \ln X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$$

where, Y = the number of wildcats drilled; X_2 = price at the wellhead in the previous period (in constant dollar, 1972 = 100); X_3 = domestic output; X_4 = GNP constant dollars (1972 = 100); X_S = trend variable, 1948 = 1, 1949 = 2... 1978 = 31.

According to the model, there are four exogenous variables in the equation: oil price, domestic output, GNP and trend variable. Thus, we can predict directions of the results before estimating the model. Firstly, if the oil prices rise, it can be inferred that there is an increase in demand for oil. As a result, manufactures have to adapt their oil production to response rising demand, that is, coefficient of X_2 is expected to be positive since change in Y moves in the same direction as X_2 change.

TABLE I
ANNUAL DATA WITH VARIABLES [7]

Thousands of wildcats, (*Y*)	Per barrel price constant \$ (X_2)	Domestic output (millions of barrels per day), (X_3)	GNP, Constant \$ billions, (X_4)	TIME (X_5)
8.01	4.89	5.52	487.67	1948 = 1
9.06	4.83	5.05	490.59	1949 = 2
10.31	4.68	5.41	533.55	1950 = 3
11.76	4.42	6.16	576.57	1951 = 4
12.43	4.36	6.26	598.62	1952 = 5
13.31	4.55	6.34	621.77	1953 = 6
13.10	4.66	6.81	613.67	1954 = 7
14.94	4.54	7.15	654.80	1955 = 8
16.17	4.44	7.17	668.84	1956 = 9
14.71	4.75	6.71	681.02	1957 = 10
13.20	4.56	7.05	679.53	1958 = 11
13.19	4.29	7.04	720.53	1959 = 12
11.70	4.19	7.18	736.86	1960 = 13
10.99	4.17	7.33	755.34	1961 = 14
10.80	4.11	7.54	799.15	1962 = 15
10.66	4.04	7.61	830.70	1963 = 16
10.75	3.96	7.80	874.29	1964 = 17
9.47	3.85	8.30	925.86	1965 = 18
10.31	3.75	8.81	980.98	1966 = 19
8.88	3.69	8.66	1,007.72	1967 = 20
8.88	3.56	8.78	1,051.83	1968 = 21
9.70	3.56	9.18	1,078.76	1969 = 22
7.69	3.48	9.03	1,075.31	1970 = 23
6.92	3.53	9.00	1,107.48	1971 = 24
7.54	3.39	8.78	1,171.10	1972 = 25
7.47	3.68	8.38	1,234.97	1973 = 26
8.63	5.92	8.01	1,217.81	1974 = 27
9.21	6.03	7.78	1,202.36	1975 = 28
9.23	6.12	7.88	1,271.01	1976 = 29
9.96	6.05	7.88	1,332.67	1977 = 30
10.78	5.89	8.67	1,385.10	1978 = 31

TABLE II
DEFINITIONS OF VARIABLES

Variables	Definitions	Units of measurement
Y	The number of wildcats drilled	Thousands of wildcats
X_2	Price at the wellhead in the previous period	Per barrel price, Constant \$ billion
X_3	Domestic output	Millions of barrels per day
X_4	GNP constant dollars	Constant \$ billion
X_5	Trend variable	-

Secondly, the more domestic outputs are, the more amount of oil is used to produce those outputs. The more domestic outputs are, the more amount of oil is used to produce those outputs. Thus, coefficient of X_3 should be positive as well. Thirdly, when national income or GNP increases, it is a sign that people have more purchasing power. Hence, demand for oil will grow directly via consumption of oil and indirectly via consumption of other goods which use oil as a raw material. The relation is predicted to be positive as well.

Finally, the expected coefficient of this trend variable is positive because from time to time, there are new machine created everyday so the usage of oil as a source of energy is more and more.

There is a total of five variables of our model: Y, X_2, X_3, X_4

and X_5. Table II is the table of the definitions of variables.

Descriptive Statistics of Each Variable

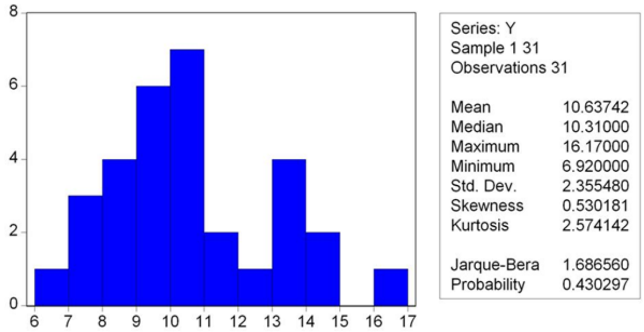

Fig. 1 Descriptive Statistics of "Y" Variable

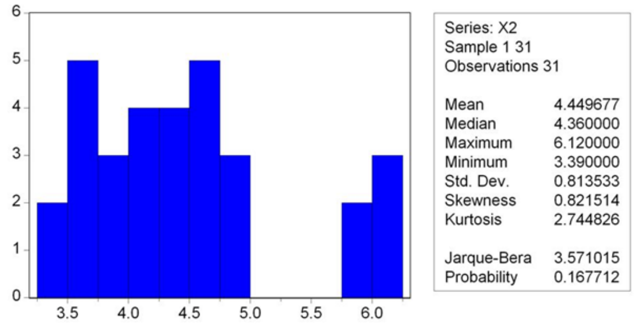

Fig. 2 Descriptive Statistics of "X_2" Variable

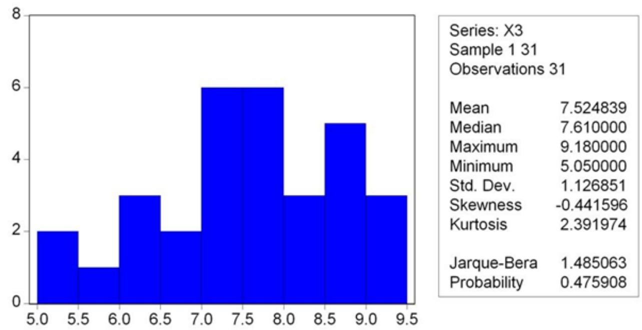

Fig. 3 Descriptive Statistics of "X_3" Variable

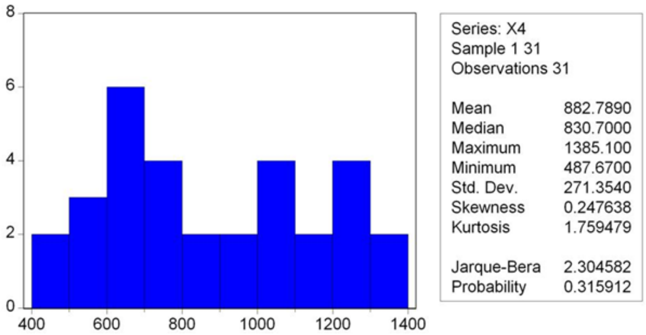

Fig. 4 Descriptive Statistics of "X_4" Variable

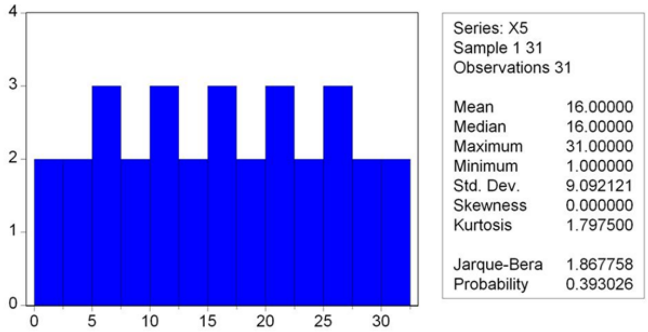

Fig. 5 Descriptive Statistics of "X_5" Variable

D. Model Estimation and Hypothesis Testing

Parameter Estimation: We apply the ordinary least square method and the output is shown in Table III.

TABLE III
PARAMETER ESTIMATED BY ORDINARY LEAST SQUARE METHOD

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.798930	8.931248	-1.097151	0.2826
X2	2.700179	0.698589	3.865190	0.0007
X3	3.045134	0.941113	3.235673	0.0033
X4	-0.015994	0.008212	-1.947619	0.0623
X5	-0.023347	0.273410	-0.085394	0.9326
R-squared	0.578391	Mean dependent var		10.63742
Adjusted R-squared	0.513529	S.D. dependent var		2.355480
S.E. of regression	1.642889	Akaike info criterion		3.977479
Sum squared residual	70.17616	Schwarz criterion		4.208767
Log likelihood	-56.65093	F-statistic		8.917142
Durbin-Watson stat	0.938545	Prob(F-statistic)		0.000113

Dependent Variable: Y; Method: Least Squares; Sample: 131; Included observations: 31

Estimated equation with t statistic in the parentheses:

$$Y_t = -9.798930 + 2.700179X_{2t} + 3.456X_{3t} - 0.015994X_{4t} - 0.0237X_{ut} \quad (1)$$

$$R^2 = 0.58; SE = 1.636$$

Hypothesis Testing

Three of five coefficients are insignificant at the 5% level (accept H_0: $\beta_i = 0$) because their t statistics are less than 2.052 (from t distribution table, df = 27) in absolute value and the rest are significant. In addition, it is obvious that R^2 value is only 0.58, which means that the explanatory variables in the right-hand side can explain 58% of the movement in Y. Therefore, verification, and adjustment will be needed to improve the equation.

Multicollinearity

1. Test for Multicollinearity

From Table III, it is obvious that although R^2 value is quite moderate, there are only two significant t ratios. Thus, it may have a relationship among explanatory variables in this model. To ensure the existence of multicollinearity, we used a correlation matrix.

As you can be seen from Table IV, several of these pair-wise correlations are quite high. For instance, correlation between X_4 and X_5 is 0.990589, between X_3 and X_4 is 0.827147 and between X_3 and X_5 is 0.848050, respectively. It indicates that there is a collinearity problem in our model.

TABLE IV
CORRELATION MATRIX

	Y	X2	X3	X4	X5
Y	1.000000	0.135193	0.426595	0.557392	-0.529881
X_2	0.135193	1.000000	0.305424	0.182018	0.160882
X_3	0.426595	0.305424	1.000000	0.827147	0.848050
X_4	-0.557392	0.182018	0.827147	1.000000	0.990589
X_5	0.529881	0.160882	0.848050	0.990589	1.000000

2. Correction for Multicollinearity

According to Table IV, there is a strong relationship among X_3, X_4 and X_5 leading to the multicollinearity in our equation. To correct the model, we will drop a variable owing to we cannot find more information to add or poll in the model. We decide to drop X4 because of two reasons:

- X_3 (domestic output) and X_4 (GNP) are quite similar. Hence, using only one of them would be better for the model.
- From the correlation matrix in Table IV, it manifests a strong relationship among X_3, X_4 and X_5. So, if we drop one of them, it may improve out model. Especially, correlation between X_4 and X_5 is close to one so it may be good to drop X_4 or X_5 instead of X_3

In terms of parameter estimation and using the OLS method in the model again, the regression results are as shown in Fig. 6.

The estimated (2) with t statistic in the parentheses is as follows:

$$Y_t = -16.9922 + 2.6565X_{2t} + 3.1870X_{3t} - 0.5103X_{5t} \quad (2)$$

R2 = 0.52; SE = 1.721

From the analysis of multicollinearity, almost all coefficients are significant at the 5% level (reject H_o: $R_t = 0$ because their t are greater than or equal to 2.048 in absolute value), except only the constant term of coefficient. Besides its t statistic is close to 2. So, we can ignore it for its in signification. Equation (2) is considerably better than uncorrected equation (1), and the multicollinearity is already eliminated from our model.

Dependent Variable: Y
Method: Least Squares

Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-16.90701	8.562803	-1.974471	0.0586
X2	2.655855	0.733446	3.621066	0.0012
X3	3.172020	0.986224	3.216328	0.0034
X5	-0.508888	0.117925	-4.315345	0.0002

R-squared	0.516882	Mean dependent var	10.63742
Adjusted R-squared	0.463202	S.D. dependent var	2.355480
S.E. of regression	1.725778	Akaike info criterion	4.049147
Sum squared residual	80.41438	Schwarz criterion	4.234178
Log likelihood	-58.76178	F-statistic	9.628973
Durbin-Watson stat	0.659223	Prob(F-statistic)	0.000171

Fig. 6 Parameter estimated by OLS method

Autocorrelation

1. Test for Autocorrelation

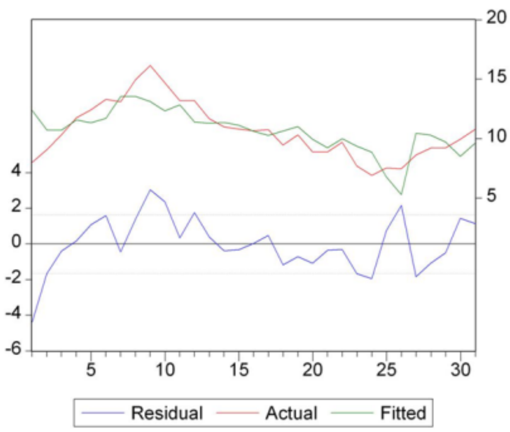

Fig. 7 Residual line

From Fig. 7, the residual line has a pattern indicating that there is a positive autocorrelation. To ensure that this problem exists, we will exert Durbin-Watson test.

According to Fig. 6, Durbin-Watson statistic is equal to 0.653 and from Durbin-Watson d statistic table at 5 percent level: d_L= 1.229 and d_U = 1.650.

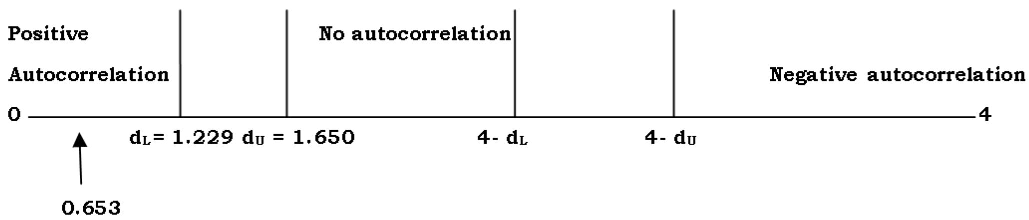

Fig. 8 Durbin-Watson d Statistic

We will find that Durbin-Watson statistic falls in positive autocorrelation region. As a result, we reject null hypothesis (H: P = 0), that is, there is autocorrelation in our model surely.

2. Correction on for Autocorrelation

We re-estimate (2) by using Corchrane-Orcutt procedure and a serial correlation is eliminated. The regression results are:

- Dependent Variable: Y

- Method: Least Squares
- Sample (adjusted): 2 31
- Included observations: 30 after adjusting endpoints
- Convergence achieved after 9 interrelations

$$\mathrm{Y_t} = 5.0952 + 1.373X_{2t} + 3.1870X_{3t} - 0.5103X_{5t} \quad (3)$$

$R^2 = 0.88$; SE = 0.879; DW = 1.823

Now Durbin-Watson statistic is equal to 1.823, so it falls into no autocorrelation region. Therefore, we accept null hypothesis, in other words, there is no statistically significant evidence of autocorrelation, positive or negative. Besides, the equation is greatly better than the prior one because R^2 value rises from 0.52 to 0.88.

Heteroscedasticity

1. Test for Heteroscedasticity

Value of nR^2 from both tests are less than critical chi-square value at 5% level of significance (df = 3): π^2= 9.815. Thus, we can accept null hypothesis (H_o: (X_I = 0 where OC is a coefficient in auxiliary equation). We can conclude that there is no Heteroscedasticity in our model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.095206	5.260470	0.968584	0.3420
X2	1.137359	0.442307	2.571423	0.0165
X3	0.942234	0.608573	1.548268	0.1341
X5	-0.351872	0.092804	-3.791557	0.0008
AR(1)	0.710037	0.093735	7.574928	0.0000

R-squared	0.878218	Mean dependent var	10.73400
Adjusted R-squared	0.858733	S.D. dependent var	2.339377
S.E. of regression	0.879268	Akaike info criterion	2.731558
Sum squared resid	19.32782	Schwarz criterion	2.965091
Log likelihood	-35.97337	F-statistic	45.07108
Durbin-Watson stat	1.832106	Prob(F-statistic)	0000000
Inverted AR Roots	71		

Fig. 9 Parameter estimated by OLS Method

White Heteroskedasticity Test:

F- statistic	1.257117	Probability	0.315129
Obs`R-squared	7.408680	Probability	0.284699

Fig. 10 White Test with No Cross Term

White Heteroskedasticity Test:

F-statistic	0.955024	Probability	0.502733
Obs'R-squared	9.017470	Probability	0.435663

Fig. 11 White Test with Cross Term

IV. Interpretation of the Results and Conclusion

The final results of the regression (3) show that the explanatory variables on the right hand side can explain 88% of movement in change of the number of wildcats drilled. The remained variables which substantially influence to the dependent variable is 3 variables, we drop X_4 in the correcting process. As we predicted the direction of the results before estimation, we expected the coefficient of X_4 should be positive. But after estimated original data, we found that it is negative. Therefore, it is possible that X_Q or GNP may not a proper variable for this model.

At last, we obtain the final results:

$$\mathrm{Y_t} = 5.0952 + 1.373X_{2t} + 3.1870X_{3t} - 0.5103X_{5t}$$

It shows that domestic output (X_3) has a strong and positive effect on the number of wildcats (Y) as we predicted earlier. The price at the wellhead (X_2) also has the expected positive impact whereas time trend (X_5) has negative effect on the number of wildcats.

V. Limitation and Further Extension

There are some drawbacks in the paper, especially in the part of literature review. There are limited research citations existing in this research arena. Moreover, our model is based on only 31 observations and data-collecting time is out of date which is from time period 1948 to 1978. Therefore, if we apply their results to a current situation, it may not be absolutely correct but very useful. It is recommended that larger and more update observation should be considered. In addition, to improve the model, we ought to observe other variables having effects to the number of wildcats drills such as the decision of OPEC committees about the quantity of world oil. Either added or omitted variables may increase R^2 value of the model as well. We remember that these are complex interactions that are impacted by a number of outside variables. By analyzing them, decision-makers in government, finance, and business organization can make well-informed decision.

REFERENCES

[1] "Basic Econometrics: Student Solutions Manual for Use with Basic Econometrics 0072427922." Ebin.pub, 21 Dec. 2019, ebin.pub/basic-econometrics-student-solutions-manual-for-use-with-basic-econometrics-0072427922.html

[2] Rudolph, Kurt. (2016). Benchmarking Exploration Predictions and Performance Using 20+ Years of Drilling Results: One Company's Experience. AAPG Bulletin. IN PRESS. 10.1306/06281616060.

[3] Koulidis, Alexis & Kelessidis, Vassilios & Shehab, Ahmed. (2021). Exploitation of Field Drilling Data with an Innovative Drilling Simulator: Highly Effective Simulation of Rotating and Sliding Mode. 10.2118/202176-MS.

[4] Lirong Dou, Zhixin Wen, Jianjun Wang, Zhaoming Wang, Zhengjun He, Xiaobing Liu, Ningning Zhang, Analysis of the world oil and gas exploration situation in 2021, Petroleum Exploration and Development, Volume 49, Issue 5, 2022, Pages 1195-1209, ISSN 1876-3804, https://doi.org/10.1016/S1876-3804(22)60343-4. (https://www.sciencedirect.com/science/article/pii/S1876380422603434)

[5] Chen, Xi & Wang, Zuo-qian & Li, Chang-xuan & Gao, Feng & Wei, Qing. (2024). Characteristics of Global Offshore Oil and Gas Development. 10.1007/978-981-97-0475-0_87.

[6] Boukredera, F.S., Hadjadj, A. & Youcefi, M.R. Drilling vibrations diagnostic through drilling data analyses and visualization in real time application. *Earth Sci Inform* 14, 1919–1936 (2021). https://doi.org/10.1007/s12145-021-00649-8

[7] Damodar N. Gujarati, Basic Econometrics, 4th Edition, Tata McGraw Hill.