Econometric model for forecasting oil production in OECD member states

Carmen Valentina Rădulescu[1,*], *Dumitru Alexandru* Bodislav[1], *Sorin* Burlacu[1], *Florina* BRAN[1] and *Lyaman Karimova*[1]

[1]Bucharest University of Economic Studies, 010374 Bucharest, Romania

Abstract. In this article we present an econometric model of oil production forecast at OECD member level that will allow decision makers but also other oil product stakeholders to be responsible for oil production in OECD member states. This responsibility can be perceived from several perspectives: economic, social, environmental, political, military etc.

In order to be able to find the ideal formula for our calculation, we went through the specialized literature and brought elements of analysis during the research through several econometric paths traveled by other researchers and who provided us with support for our research. Before proceeding technically, in order to understand the urgency of this approach and of this study, we also discussed how oil and natural gas are explored, exploited and extracted from the underground deposits.

We considered that the proposed model could be improved in the future so as to portray certain geopolitical or economic factors, determinants for oil production, such as embargoes, periods of armed conflict in the main extraction areas or times of financial crisis and the decline of financial markets.

Introduction

The demand for oil began to accelerate by inventing the internal combustion engine at the end of the 19th century, becoming one of the most important merchandise marketed worldwide. Modern civilization, as it is known, depends largely on oil and its by-products. To date, the industrialized nations have completely accepted an uninterrupted supply of cheap hydrocarbons.

One of the main objectives of the oil industry is to offer the modern world the continuous flow of fluids, hydrocarbons, oil and gas with hydrocarbons, while making a profit. Liquids are exhaustible resources; thus, a good forecasting scheme for oil supplies will be crucial for all parties involved in oil business, such as oil companies, financial institutions, public policy

---

[*] Corresponding author: carmen-valetina.radulescu@eam.ase.ro

planners and decision makers, oil exporting and importing countries. Such a model will also contribute to stabilizing and securing the oil market.

Literature review

The specialized literature deals with the production forecast from many perspectives so as to respond to the needs of those who wish to use this data but given the conditions and availability of these data (Angheluta et al., 2019). Some researchers propose new methods for oil price forecasting, based on the support vector machine (SVM) in which the procedure for developing a support model for time series forecasts involves data collection, sample pre-processing, training and learning, and out-of-sample forecasting Xie, Yu, Xu, & Wang (2006). There are also researchers who consider that the problem is not only the precise definition of the oil price variable, but also whether the oil price was expressed in nominal terms or in real terms, what estimation and evaluation periods were chosen or how the accuracy of the forecast is evaluated, etc. (Alquist, Kilian, & Vigfusson, 2011). Thus, we identify more and more studies that are trying to develop forecast models to predict global crude oil supply with better accuracy than existing models, some starting with the Hubbert model but considering that they go beyond the limitations and restrictions associated with the original model by introduction of several cycles, depending on the historical trend of oil production and known oil reserves (Nashawi, Malallah, & Al-Bisharah, 2010).

Methodology

In the past, many researchers have developed different methods for forecasting future oil production using either available final reserve determinations or extrapolation of production history. Traditional statistical and econometric techniques, such as linear regression (LinR), cointegration analysis, GARCH models, Random Walk naive models, vector self-regression models (VAR) and error correction models (ECM), have proliferated in recent decades. forecast oil price.

The model chosen in our research is one of the autoregressive vector types (class of models known in the specialty literature as VAR, popularized in the economic sciences by Sims, 1980). The foundations underlying this decision to choose an autoregressive model were some of a theoretical nature, namely that the global oil production from the present moment, can be influenced by the production from the previous moment, but also by the revenues from the previous year (the principle profit reinvestment).

In our research, for the realization of the econometric model, we used primary data from two public sources: The World Bank database and the OECD database. The analyzed period began with the year 1989 and as close as possible to the present, depending on the availability of the necessary data in specifying the econometric model. Usually, the above models can give good prediction results when the price range studied is linear or almost linear. However, in the real-world crude oil price range, there is a great deal of nonlinearity and irregularities. Numerous experiments have shown that predictive performance could be very poor if we continue to use these traditional statistical and econometric models (Weigend, Gershenfeld, 1994).

Findings

The time series identified as useful for highlighting a relationship between oil rents, the level of production and the amount of energy from renewable sources were as follows (Bodislav et al., 2019):

a) Revenue from oil production as a percentage of gross domestic product for OECD Member States, provided by the World Bank database. This indicator is constructed as a difference between the value of oil production at regional prices and the total costs of production. The estimates are based on the methodology described by Lange et al. (2018). The evolution of the chronological series is shown by the following graph:
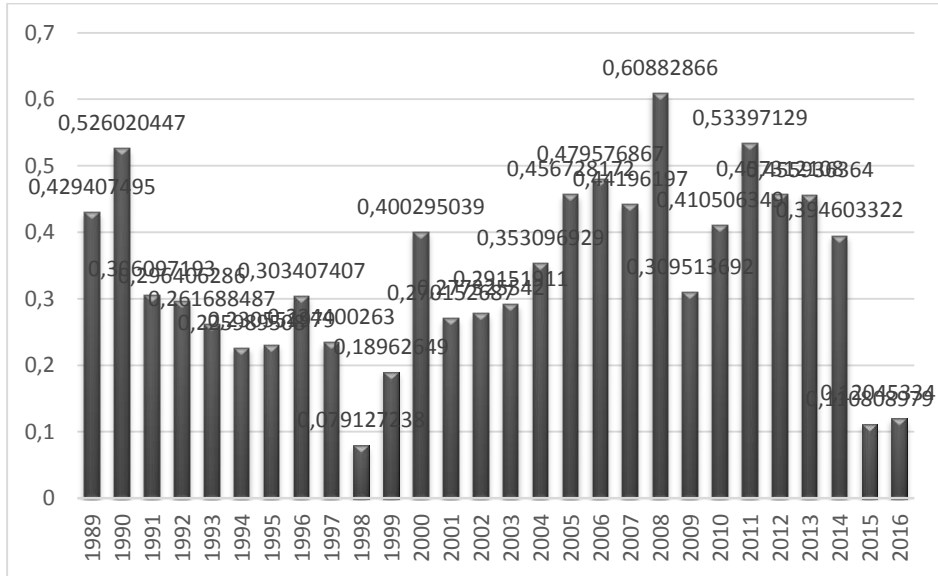

Fig. 1: Average oil revenues in OECD member states (% of GDP - 1989 - 2016).
Source of processed data: World Bank

b) Oil production from OECD Member States quantified in tonnes of oil equivalent (TEP), taken from the OECD database. Oil production is defined by the OECD as the quantity of oil extracted from the soil after removal of inert matter and impurities. At the level of the analyzed group of states, the evolution of oil production is graphically expressed by:
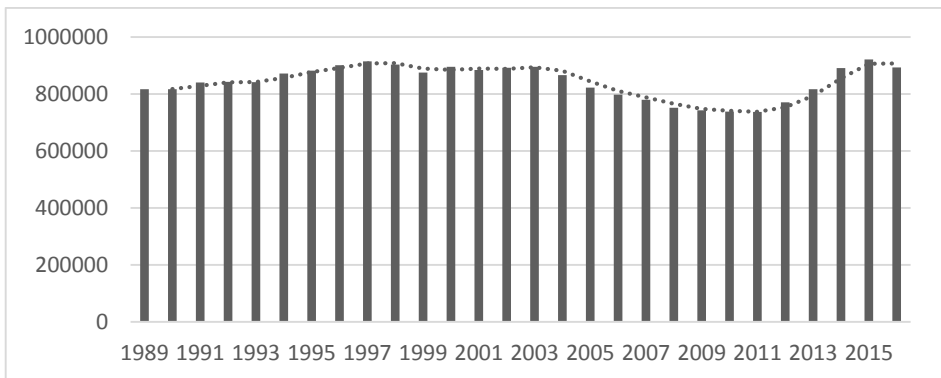

Fig. 2: Average of oil production in OECD member states (1989 - 2016).
Source of processed data: OECD

Renewable energy production for the group of OECD member countries, expressed in tonnes of oil equivalent (TEP), taken from the OECD database. According to the definition provided by the Organization for Economic Cooperation and Development, the indicator is defined as

the contribution of renewable energy in the total energy supply and includes energy obtained from hydro, geothermal, wind, as well as solar and wave energy. Also, here comes the energy derived from solid biofuels, biogasoline, biodiesel or other liquid biofuels, biogas and the fraction of municipal waste recycled to obtain fuels. It can be seen that the contribution of energy from renewable sources brought to the total energy supply, is noticeable upward trend, starting with 2006, and at the end of the analyzed periods, the value of this indicator was almost double compared to that of 2000.
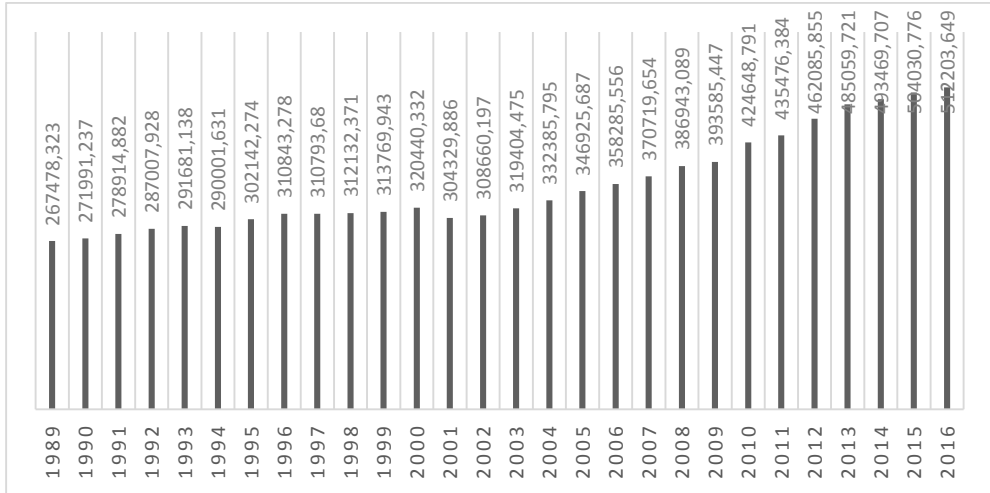

Fig. 3: The average of the energy supply from renewable sources brought to the total energy supply at the level of the OECD member states (1989 - 2016).
Source of processed data: OECD

Discussion - Post-specification tests and impulse-response function
The first test run to test if the model is a stable one and it complies with the assumptions of the classical model, was the one for testing the autocorrelation of errors. The autocorrelation was tested using the Lagrange multiplier. The null hypothesis of this test expresses that there is no autocorrelation in the series of residuals.

Table 1. *Self-correcting error testing at VAR level (1) using Lagrange multiplier.*
Lagrange-multiplier test

lag	chi2	df	Prob > chi2
1	2.3459	4	0.67243
2	2.6559	4	0.61694
3	1.3280	4	0.85660
4	1.6366	4	0.80220

H0: no autocorrelation at lag order
Source: own estimates

According to the values offered by this test, we can accept the null hypothesis, namely that the model does not show autocorrelation in the series of residues until the fourth delay.

In the next stage we ran the error symmetry test for the second equation, considered as statistically relevant for the analysis performed. As a result of the estimates provided by the symmetry test, it can be seen that the series of residues maintains a normal distribution and is mesocurtic.

***Table 2**. Testing the residual series symmetry at the VAR level (1.)*

Jarque-Bera test

```
+------------------------------------------------------+
|      Equation |       chi2   df  Prob > chi2 |
|-----------------+----------------------------------|
|        vp_OR |     10.390   2    0.00554   |
|        vp_prod |     0.665   2    0.71725   |
|          ALL |     11.055   4    0.02595   |
+------------------------------------------------------+
```

Skewness test

```
+------------------------------------------------------+
|      Equation | Skewness   chi2  df  Prob > chi2 |
|-----------------+----------------------------------|
|        vp_OR | 1.1037   5.278   1    0.02159   |
|        vp_prod | .33796   0.495   1    0.48174   |
|          ALL |       5.773   2    0.05576   |
+------------------------------------------------------+
```

Kurtosis test

```
+------------------------------------------------------+
|      Equation | Kurtosis   chi2   df  Prob > chi2 |
|-----------------+----------------------------------|
|        vp_OR | 5.1723   5.112   1    0.02376   |
|        vp_prod | 2.6042   0.170   1    0.68034   |
|          ALL |       5.282   2    0.07130   |
+------------------------------------------------------+
```

Source: own estimates

The most important of the post-specification tests of the VAR model is the system stability test using the Eigenvalue stability condition. The essential condition of this test for the specified model to be stable is that all the coefficients of the test are strictly less than 1, so that they can be found inside the unit circle. The results provided by this test are shown in the following table, and in addition the unit circle is graphically represented:
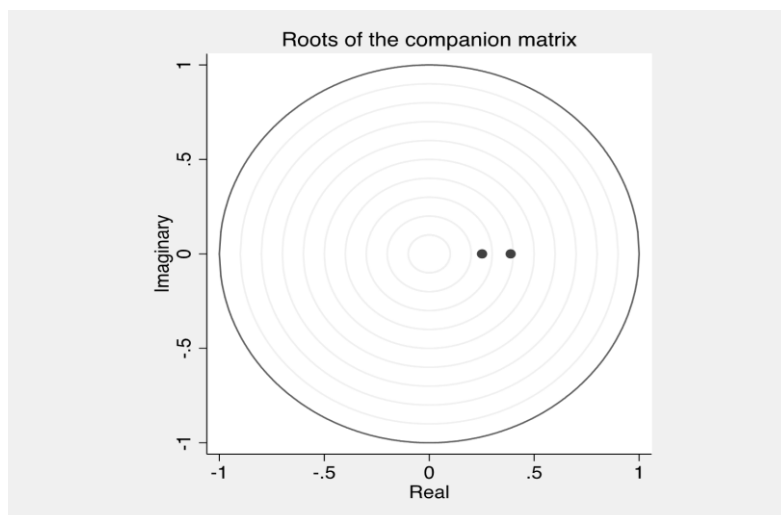
Fig. 4: The unit circle related to the stability condition.
Source: own estimates

In order to establish causality in the Granger sense, we used the Wald test. In a general framework, it can be said about a variable A that will cause a variable B in the Granger sense, if the delays of variable A can improve the subsequent forecast of variable B. The test results for our model are presented below:

Table 3. *Wald test for causality testing in the Granger sense.*

Granger causality Wald tests

```
+----------------------------------------------------------------+
|      Equation        Excluded |  chi2    df Prob > chi2 |
|------------------------------------+---------------------------|
|       vp_OR           vp_prod | 4.7743   1    0.029   |
|       vp_OR               ALL | 4.7743   1    0.029   |
|------------------------------------+---------------------------|
|       vp_prod           vp_OR | 4.7497   1    0.029   |
|       vp_prod             ALL | 4.7497   1    0.029   |
+----------------------------------------------------------------+
```

Source: own estimates

The coefficients of the test, according to the values in the last column of the table, are significantly different from zero, which means that the delays of both endogenous variables will cause in Granger's sense the forecast of the other variable. From the tests presented above, it follows that the specified model is a stable one (autocorrelation, residue symmetry and unit circle), and the estimates can produce robust results (causality in the Granger sense) to make forecasts.

The main component of a VAR model in the post-estimation stage is to establish the **impulse-response function**. This function determines what will happen to one of the system variables (response), when there is a positive shock of one unit at another system variable (impulse). For the previously specified model VAR (1), the generalized impulse-response function will have the following graphical representation:
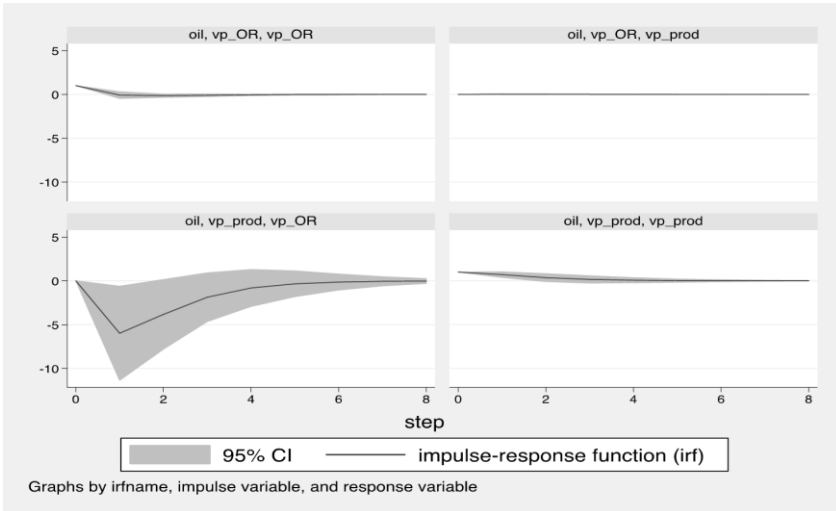
Fig. 5: The impulse-response function of the VAR model (1).
Source: own estimates

In detail, for the second equation (the one that passes the statistical tests), a positive shock in the rate of growth of oil revenues for OECD member countries by one unit will lead to an increase in the rate of growth of oil production for countries. OECD members with approximately 0.2 - 0.3 percentage points, as suggested by the following figure:
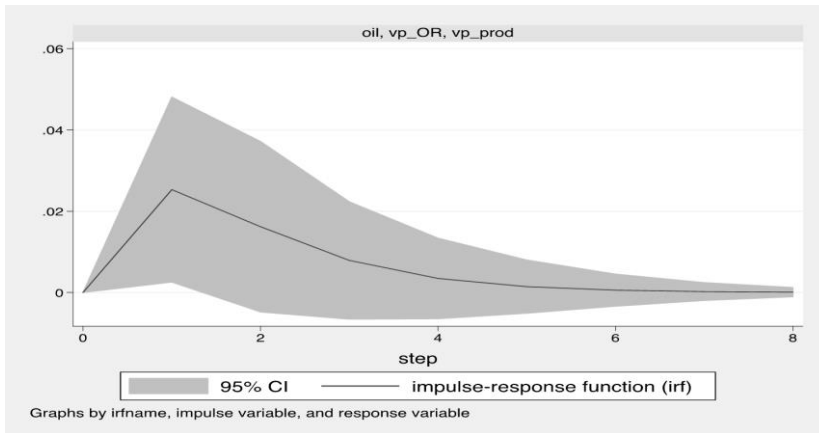

Fig. 6: impulse-response function: impulse vp_OR - response vp_prod.
Source: own estimates

The fact that the presented model is a correctly specified one also results from the evolution of the upper and lower limits of the confidence interval (gray area of the graph). The amount of information with which each variable contributes to the formation of the other variables in the autoregressive process is achieved by the variance decomposition. This determines how much of the variance of forecast error of each variable can be explained by independent shocks occurring at the level of the other variables in the system. Using the same impulse response function, where vp_OR receives the impulse, and the response took place at the level of the variable vp_prod, it was observed that up to step 8, the variance of forecast errors

is explained in a proportion of about 13 percent, by the growth rate dynamics. of oil revenues for the OECD sample.

Table 4. *Variance decomposition for the estimated model.*

```
+--------------------+
|    |   (1)   |
| step |  fevd  |
|--------+-----------|
|0    | 0       |
|1    | .000905  |
|2    | .098543  |
|3    | .124805  |
|4    | .130632  |
|5    | .131759  |
|6    | .131957  |
|7    | .13199   |
|8    | .131995  |
+--------------------+
```

Source: own estimates

"Sample" forecast of the VAR model and conclusions

In the last part of the research, in order to reconfirm the validity of the constructed econometric model, a forecast of the "in sample" type was made, using the estimates provided by the econometric results. The forecast estimates were some of a dynamic nature, and the results are presented in the following figure:
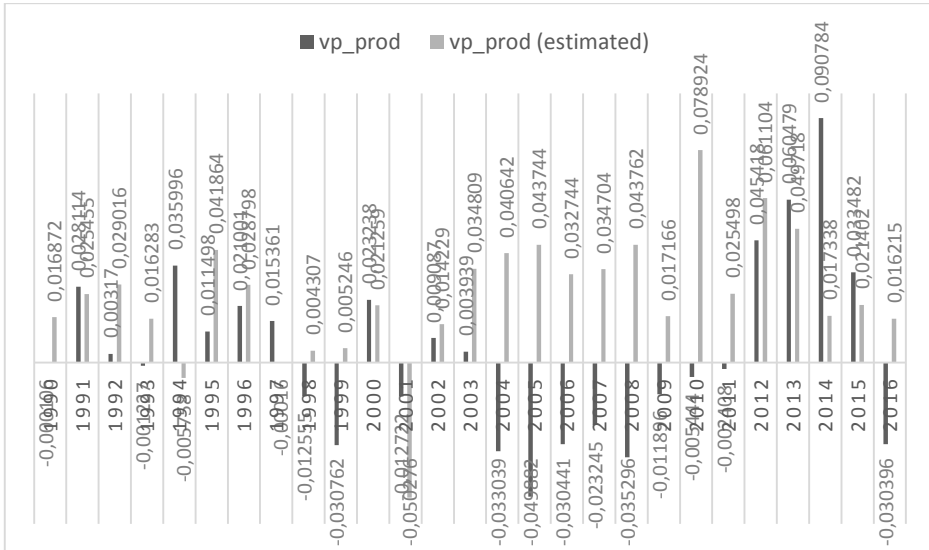

Fig. 7: Forecast in sample for VAR model (1).
Source: own estimates

It can be seen from this graph that the adjusted values of the model are represented by the red dashed line, while the initial values are shown by the dashed blue line. Estimates have a similar evolution to the initial values, except for the period beginning in 2005, when the growth rate of oil production has slowed considerably. Subsequently, the forecasted values converge to the initial ones until 2011, when a new break-up at the graphical level

appears. This can be caused by the value below 0.5 of the coefficient R^2 for the equation for which the simulation was performed. However, the value of 0.433 is not a critical one according to Mooi and Sarstedt (2014), who considers a minimum acceptable value in the social sciences close to 0.1, and for the exploratory research an optimal value of this indicator is in the range 0,4 - 0.6. Therefore, this can be considered one of the weaknesses of the model.

Conclusion

The model chosen for highlighting the previously mentioned relationship is one of the autoregressive vector types (class of models known in the literature as VAR). The foundations underlying this decision to choose an autoregressive model were some of a theoretical nature, namely that the global oil production from the present moment, can be influenced by the production from the previous moment, but also by the revenues from the previous year (the principle profit reinvestment).

Unfortunately, for the forecast component, it was not possible to estimate the "out of sample" type (outside the sample), because the VAR model also included an exogenous variable (the growth rate of the energy supply from renewable sources brought to the supply). total energy for OECD member states), and this has caused asymptotic limitations.

The model can be improved in the future by introducing structural breaks for periods when large differences are observed between adjusted values and initial values or by introducing dummy variables that play out certain geopolitical or economic factors, determining for oil production, such as embargoes, periods of armed conflict in the main extraction areas or times of financial crisis and the decline of financial markets.

References

1. R. Alquist, L. Kilian, R. J. Vigfusson. Forecasting the price of oil. (2011)
2. S. P. Angheluta, S. Burlacu, A. Diaconu, C. S. Curea. The Energy from Renewable Sources in the European Union: Achieving the Goals. *European Journal of Sustainable Development, 8(5), 57.* (2019)
3. D. Bodislav, C. V. Rădulescu, D. Moise, S. Burlacu. Environmental Policy in the Romanian Public Sector. *The Bucharest University of Economic Studies Publishing House, 312* (2019)
4. D. Jula. Metode și tehnici performante de testare a ipotezelor în economie – *Workshop, Bucharest*; (2015)
5. G.M. Lange, Q. Wodon, K. Carey. *The Changing Wealth of Nations 2018: Building a Sustainable Future*. Washington, DC: World Bank; (2018)
6. L. Mammadzada, K. Tural, U. Abuzarli. *European Union's Energy Strategy and its Influence on the Global Oil Market*. Quality-Access to Success, 18. (2017)
7. E. Mooi, M. Sarstedt. *A Concise Guide to Market Research*. Springer Verlag-Berlin Heidelberg. E-Book ISBN 978-3-642-12541-6; (2014)
8. S. Nashawi, A. Malallah, M. Al-Bisharah. *Forecasting world crude oil production using multicyclic Hubbert model. Energy & Fuels*, 24(3), 1788-1800. (2010)
9. C.A. Sims, *Macroeconomics and Reality*, Econometrica, 48, 1-48. (1980)
10. W. Xie, L.Yu., S., Xu, S. Wang. A new method for crude oil price forecasting based on support vector machines. In International conference on computational science (pp. 444-451). Springer, Berlin, Heidelberg. (2006)