Establishment and comparison of predictive models for oil and petroleum products, electricity and gas: A Cross-Border Analysis

Aglaia Papakyrillou[1,2], Anna Mpoukouvala[1,2], George Tsaklidis[2]

[1]ELSTAT, Greece

[2]AUTH, Greece

Abstract

This study investigates and compares five different predictive models for three critical parameters: oil and petroleum product deliveries, electricity availability and inland gas consumption. The research covers six consecutive months, from October 2023 to March 2024, providing a comprehensive overview of predictions during periods of both low and high energy consumption for five countries of EU, each differing in various parameters.
Various model types were developed, capturing a broad range of approaches and reflecting methodological diversity. Specifically, AutoRegressive Moving Average models with exogenous variables were applied to stationary time series. Two indicators were utilized for the first and second models, while more than two indicators were employed for the third model. Corresponding indicators for all combinations of countries and parameters were used in the implementation of the first and second models, in contrast to the third model, where indicators were selected considering the specificities of each case. The fourth model was based on a Recurrent Neural Network (RNN), specially designed for analysing time series data. In the last predictive approach, linear models were employed, fitted separately for each combination, taking advantage of their flexibility to adapt.
In conclusion, both the performance and the advantages of the selected models are presented, highlighting significant differences and similarities that are considered crucial for the comparison of the provided approaches.

Keywords: Energy, Oil, Electricity, Gas, Forecasting

1. Introduction

This study, carried out in cooperation with the Hellenic Statistical Authority (ELSTAT), focuses on energy. In modern society, energy is emerging as a critical factor that shapes and influences every aspect of human existence and economic activity. From the operation of industries to our daily activities and needs, energy is the foundation of progress and human comfort. Beyond its primary recognition as an essential resource for the operation of the machines and appliances used every day, energy is becoming vital to promoting sustainability, economic growth and ensuring social well-being. The efficient use of energy and efforts to develop and implement sustainable technologies are priorities that will ensure sustainable development and environmental protection for future generations. Because of all these factors, it is necessary to anticipate the evolution of energy sources, in particular oil, electricity and gas.

1.1 Brief description of the study

The current project focuses on three critical parameters, namely oil and petroleum product deliveries, electricity availability and inland gas consumption. More specifically, this thesis covers a six-month period, from October 2023 to March 2024, with the aim of forecasting the aforementioned parameters. This forecast is carried out using five different forecasting models. The aim was to select the best model by comparing them. In particular, the first three follow the methodology of AutoRegressive Moving Average (ARMAX) models, the fourth is based on Recursive Neural Networks (RNN), while the fifth uses linear models. For each of the three cases, five European Union countries are selected. For oil, the chosen countries are Spain, Ireland, Luxembourg, Portugal and Slovakia. For electricity, the included countries are Denmark, Greece, Ireland, France and Latvia, while for gas, the focus shifts to Spain, Finland, Ireland, Lithuania and Luxembourg.

1.2 Data

Following the analysis and a prudent understanding of the central framework of the study, it becomes imperative to provide an extensive description of the datasets used to carry out this research project. The databases belong to the official Eurostat website, which is widely accessible. The data on oil and petroleum product deliveries were extracted from the category entitled ≪Supply and transformation of oil and petroleum products - monthly data≫ and code [NRG_CB_OILM]. In addition, data on electricity availability are derived from the category ≪Supply, transformation and consumption of electricity - monthly data≫ with code [NRG_CB_EM], while data on inland gas consumption are extracted from the section called ≪Supply, transformation and consumption of gas - monthly data≫ with code [NRG_CB_GASM].

Finally, three powerful tools - software Matlab 2023b, Python and R 4.4.2 were used to achieve the following reliable results and efficient data processing and analysis.

Table 1: Description of monthly European data

Timeseries	Oil	Electricity	Gas
Basic	[O4671] Gas oil and diesel oil [GID_CAL] Gross inland deliveries – calculated [THS_T] Thousand tons	[E7000] Electricity [AIM] Available to internal market [GWH] Gigawatt-hour	[G3000] Natural gas [IC_CAL_MG] Inland consumption of gas calculated as defined in MOS GAS (Monthly Statistics GAS) [TJ_GCV] Terajoule (gross calorific value - GCV)
Indicator type 1	Electricity (Basic)	Oil (Basic)	Oil (Basic)
Indicator type 1	Gas (Basic)	Gas (Basic)	Electricity (Basic)
Indicator type 2	[O4671] Gas oil and diesel oil [EXP] Exports [THS_T] Thousand tons	[E7000] Electricity [IMP] Imports [GWH] Gigawatt-hour	[G3000] Natural gas [IMP] Imports [TJ_GCV] Terajoule (gross calorific value - GCV)
Indicator type 2	[O4671] Gas oil and diesel oil [INTMARB] International maritime bunkers [THS_T] Thousand tons	[E7000] Electricity [EXP] Exports [GWH] Gigawatt-hour	
Indicator type 2	[O4671] Gas oil and diesel oil [STK_CHG] Change in stock [THS_T] Thousand tons		

In the following paragraphs, a detailed description of the theoretical approach and the application of the methods will be presented. Each method will be explained in detail, providing an in-depth analysis of the individual steps involved. In addition, the results will be presented further through the evaluation of the relevant errors and the conclusions drawn will also be summarised.

2 Body of paper

2.1 Theoretical framework – Methods

Oil and petroleum product deliveries, electricity availability and inland gas consumption were estimated for six consecutive months for five different European countries. As previously mentioned, five different forecasting models were developed to approximate the requested values, the theoretical description of which follows.

2.1.1 First model - ARMAX-1: 2 mutual exogenous variables

The existence of trend and seasonality in the respective basic time series was examined using the graphs of the time series and its lags. The time series was then converted to a stationary one, eliminating the trend and seasonality, if detected. Then, appropriate parameters p, q were selected based on the values of the Akaike information criterion (AIC) to fit ARMA model to the basic time series.

Subsequently, indicators type 1 were processed. Their lengths were adjusted to the length of the corresponding basic time series, either by truncation or by estimating the values for the more recent months. These indicators were then converted into stationary time series according to the procedure described for the basic ones.

Then, the ARMAX model was applied to the stationary basic time series, using the stationary time series indicators as exogenous variables. The result of the application corresponded to the stationary basic time series, therefore for the calculation of the final forecast the trend and seasonality were added - since they were removed in the previous processing stages. In particular, the trend and seasonality were calculated as the average of the trend and seasonality respectively observed in the last two years for the month in question.

2.1.2 Second model - ARMAX-2: 2 mutual exogenous variables

The treatment of the basic and indicator time series of type 1 is identical to that applied in the first model. The difference between the two models lies in the computation of the final forecast. In particular, for the calculation of the final forecast, trend and seasonality were added - since they were removed in the previous processing steps. The seasonality was obtained as the average of the seasonality observed over the last two years for the month in interest, while the trend was considered equal to that observed in the most recent months for which information was available.

2.1.3 Third model - ARMAX-3: individualized exogenous variables

The treatment of the basic and indicator time series of type 1 and 2 (the combination of indicators differs by country) is identical to that followed in the first model. The formation of the final forecast was carried out according to both first and second model. However, only one of the two predicted values was picked, according to the value of Normalized Root Mean Square Error (NRMSE) in the previous months for which information was available.

2.1.4 Fourth model - RNN: training based on 12 previous months

Recurrent Neural Network (RNN) was applied on non-stationary time series to derive the estimates. Firstly, the size of the time series was defined as the length of the window - in this case, it was specified as 12 months. The input and output data were normalized to be within the range [0,1] to improve the stability of the training and the convergence of the model. The data were then partitioned into training and test sets, with 80% of the data used for training and the remaining 20% for evaluation. Using the trained model, predictions of the target value were made after returning to the original scale via inverse transformation. Using the aforementioned method, five forecasts were generated for each country and for all months. The final forecast was determined as the average of the five forecasts for each country for the corresponding month.

2.1.5 Fifth model - Linear Model: training based on 3 previous months

A Linear Model (LM) was fitted to non-stationary time series to derive the estimates. In detail, quadruplets of months were formed, such that the month of interest and three consecutive months for which all values of the basic time series are available were included. For example, of the twelve months of the year, suppose that the target was the forecast for month D of the year 2024. The available information for month D spanned from 2008 to 2023. The three consecutive months for which the most recent information was available, namely months A, B, C, were selected to complete the quadruplets. The linear model was trained from the filled quartiles and was instructed to fill in the last quartet for which the values for months A, B, C were known, while the value for month D needed to be predicted. The prediction was given according to the formula: $D = d + aA + bB + cC + e$.

2.2 Results

The results of the implementation of the previous models are presented in detail below. In particular, the relative errors (%) are summarised for all countries examined and for the months October 2023 to March 2024. Note that in the error tables, for each country and month combination, 5 error values are displayed for models 1 to 5 respectively. In addition, these time series graphs were designed for the selected countries, starting from January 2023 up to the most recent month for which information is available.

	10 2023	11 2023	12 2023	01 2024	02 2024	03 2024
Spain	7.1 1.2 1.2 2.5 0.9	3.0 7.7 4.3 1.6 5.2	18.7 5.9 5.7 3.8 7.6	3.7 4.5 4.5 9.1 1.6	...	...
Ireland	8.1 14.3 6.5 5.4 1.7	5.6 3.4 4.8 2.9 3.8	4.1 8.3 47.0 3.8 1.8	14.8 6.2 1.5 5.5 3.7	...	...
Portugal	6.0 5.2 6.3 0.1 4.4	0.7 2.5 0.3 4.5 6.7	3.1 1.5 1.7 1.6 2.1	0.0 1.6 0.6 3.6 4.9	...	...
Slovakia	3.3 0.0 3.4 5.8 3.1	2.0 1.0 0.7 2.3 1.0	1.8 16.9 17.8 13.3 5.8	4.6 8.6 9.1 0.0 7.3	...	...
Luxemburg	11.4 3.1 3.4 14.2 10.8	18.5 11.2 3.5 12.4 1.6	13.3 5.7 6.4 26.4 13.1	7.5 8.4 6.9 6.4 12.9	...	...

Figure 1: Table of relative errors (%) in oil and petroleum product deliveries forecasts

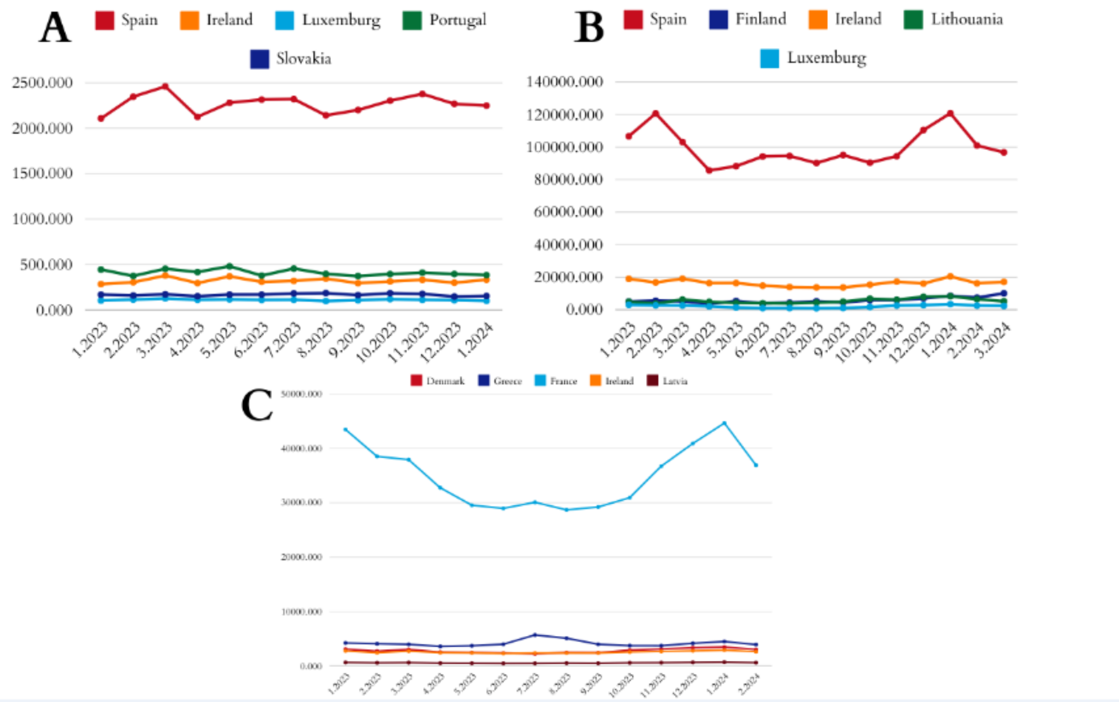
	10 2023	11 2023	12 2023	01 2024	02 2024	03 2024
Denmark	5.1 16.4 3.2 7.7 13.7	4.4 17.8 4.7 8.3 9.1	4.5 16.7 4.6 7.0 7.0	6.8 8.9 6.8 7.2 0.0	10.4 3.2 10.4 3.6 1.1	...
Ireland	3.3 8.2 2.5 4.1 1.4	1.4 4.2 2.6 4.2 0	2.8 3.8 1.5 2.3 1.8	6.2 6.8 6.2 7.7 5.7	5.1 6.7 5.1 0.3 4.4	...
Latvia	1.3 8.8 2.1 3.1 5.5	0.6 11,3 0 3.5 0.4	3.8 12.7 12.7 0.3 2.2	1.5 6.7 1.5 3.7 0.9	5.9 2.4 6.2 10.9 0.2	...
Greece	2.4 31.3 2.1 14.8 8.4	1.5 11.6 2.6 11.7 1.6	6.7 4.0 6.5 12.7 4.3	6.3 8.3 6.4 12.6 2.4	0.6 1.2 0.5 16.7 6.3	...
France	3.4 6.0 5.0 9.9	4.5 0.7 0.6 0.6	8.1 8.6 8.5 3.5	7.1 12.9 7.3 3.2	13.1 21.6 12.3 18.3	...

Figure 2: Table of relative errors (%) in electricity availability forecasts

	10 2023	11 2023	12 2023	01 2024	02 2024	03 2024
Spain	13.4 4.2 4.2 11.1 33.3	25.6 0.6 0.6 0.4 0.6	7.4 11.3 11.3 6.3 3.0	12.5 2.3 7.1 10.4 5.0	28.3 24.5 23.8 17.6 10.1	14.7 10.3 6.8 6.4 2.2
Ireland	5.2 12.2 5.1 2.9 6.8	0 11.3 10.9 4.4 3.5	22.4 9.2 22.4 11.2 3.9	5.8 17.3 5.8 17.9 13.6	10.4 23.0 10.4 18.2 5.2	1.6 2.6 3.0 1.9 6.1
Lithouania	31.2 37.2 35.8 13.5 37.6	9.2 2.9 18.4 18.5 26.8	0.6 21.9 4.2 2.8 0.6	33.2 25.4 36.0 24.1 18.9	4.2 51.9 18.5 43.4 17.2	29.8 31.3 28.8 24.9 37.9
Finland	63.5 56.4 62.4 55.0 38.0	61.3 50.2 46.4 46.3 27.7	18.8 20.7 34.2 19.6 89.2	23.9 17.6 15.9 30.9 13.0	11.2 14.0 11.6 1.7 9.4	39.7 25.4 64.5 30.2 28.1
Luxemburg	12.7 33.5 12.1 5.4 66.8	9.5 27.6 27.8 16.4 3.4	17.1 5.8 17.1 1.9 4.8	4.4 10.8 4.4 13.3 5.5	24.5 33.9 24.5 25.1 11.1	11.1 14.8 11.1 1.2 12.1

Figure 3: Table of relative errors (%) in inland gas consumption forecasts

Figure 4: A: Time series graph for oil and petroleum product deliveries (data in Thousand tons) from January 2023 to the most recent month with available information, B: Time series graph for inland gas consumption (data in Terajoule) from January 2023 to the most recent month with available information, C: Time series graph for electricity availability (data in Gigawatt – hour) from January 2023 to the most recent month with available information



2.3 Conclusions

The main aspects and results obtained will be summarised. Through the systematic examination of the data and the analyses carried out, specific conclusions are reached, highlighting the primary challenges, potential solutions and possible directions for future research, providing a comprehensive vision of the topic.

It turns out that the models more accurately approximate the required values in the case of electricity availability, showing an average relative error of all models for all selected countries during the whole study period of 5.15%. Equally well-fitting models were observed when approximating the values of oil and petroleum product deliveries, with an average relative error at 6%. In contrast, in the category of gas consumption the corresponding average exceeded 15%, demonstrating the unsatisfactory fit of the models in this case. Undoubtedly, Russia's invasion of Ukraine and the weaponisation of energy have led to the diversification of energy supply for EU countries.

Diversification is a time-consuming and expensive process, which requires investment in infrastructure, for instance the construction of new pipelines and liquefied natural gas (LNG) terminals. Gas imports from pipelines originating from Russia have fallen sharply, while volumes of LNG imports from other exporting countries is increasing. The share of Russian pipeline gas in EU imports has fallen from over 40% in 2021 to around 8% in 2023. The decline was mainly made possible by a sharp increase in LNG imports and an overall decrease in EU gas consumption. Consequently, gas consumption is highly volatile, which further increases the difficulty of forecasting.

The countries with the best average adjustment in the category of electricity availability and oil and petroleum product deliveries are Ireland and Portugal respectively. It is noted that both countries have a low overall dependence on Russian energy imports, much lower than the EU average of 24.4% (Imports from Russia in gross available energy, EU, 2020). High energy dependence on Russia was identified for Lithuania (over 95%) and Finland (over 40%). These countries were examined in terms of gas consumption and the average relative errors of all models exceeded the 20% barrier, making all models unsuitable for prediction in this case.

In detail, for the approximation of electricity availability, the linear model (fifth model) proved to be the best of the five models with an average relative error of 3.9%. This is closely followed by ARMAX-1 (first model) and ARMAX-3 (third model) with an error of 4.5%. For Denmark, Ireland, Latvia and France, the linear model showed the best fit, while more for Greece more accurate forecasts were observed with the ARMAX-1 and ARMAX-3 models.

Regarding the approach of oil and petroleum product deliveries, the most efficient of the models developed was the linear model (fifth model) with an average relative error of 4.5%. This was followed by the ARMAX-2 (second model) and ARMAX-3 (third model) models with an error of 5.5%. For Spain and Ireland, the linear model provided the best fit, for Portugal and Luxemburg more accurate forecasts were observed with the ARMAX-3, while the best fit for Slovakia was achieved by the ARMAX-1 model.

Diametrically opposite conclusions were drawn regarding the inland gas consumption approach, with the best of the models appearing to be the RNN model (fourth model) with an average relative error of 16%. For Spain, Finland and Luxemburg, RNN model provided the closest fit, for Ireland more accurate predictions were observed with the linear model, while for Lithouania ARMAX-1 was the preferred option.

In conclusion, the analysis highlights the varying degrees of success in modeling and forecasting energy-related parameters across different countries within the EU. Notably, while models exhibit satisfactory performance in approximating electricity availability and oil and petroleum product deliveries, they falter in accurately predicting inland gas consumption. Consequently, it is proposed to use linear models to approximate the values of both electricity availability, oil and petroleum product deliveries, while the RNN model was found to be the most effective of the five models tested for forecasting gas consumption, without however displaying adequate results.

Looking ahead, an important extension of the research would involve applying the models in other EU countries and at different time periods. This would allow an assessment of the generalizability of the models and potentially reveal differences in energy structures and challenges faced by different countries. In addition, testing the models at different time periods may reveal their sensitivity to periods of economic or geopolitical instability. In addition, it is important to consider the use of other indicators that may be more effective in predicting specific energy parameters. Specifically, the extension of the linear model could be examined by including additional variables. This could provide a more complex approach, revealing more intricate relationships.

Future research directions should give priority to improving the accuracy of gas consumption forecasts, especially in regions highly vulnerable to supply disruptions. Additionally, ongoing geopolitical developments and technological advances in renewable energy sources necessitate continuous reassessment of forecasting models and energy policies to ensure – possibly- resilience and sustainability in the face of evolving challenges.

References

Tangirala, A. (2014). Principles of system identification.

Rencher, Alvin C, and G. Bruce Schaalje. Linear Models in Statistics. John Wiley & Sons, 18 Jan. 2008.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of Recurrent Neural networks: LSTM cells and network architectures. Neural Computation, 31(7), 1235–1270.