Modeling and Forecasting Energy Variables and Financial Volatility Using ARIMA, VAR, and GARCH Models

Time Series Analysis for Finance & Economics

Pablo Alfaro Albert Baito David Carrascosa Camilo Mercado

Problem 1

Data Preparation & Exploratory Analysis

For problem 1 we analyze the data from the Energy Information Administration about Nuclear Energy Production. It contains information about net electricity generation over time, nuclear share of total electricity generation, and different types of nuclear- related measurements in the following series.

We converted the date into datetime format, cleaned our data which passed the quality check. We limit the data to the last 15 years from January 2010 to December 2024. Now some exploratory descriptive statistic metric:

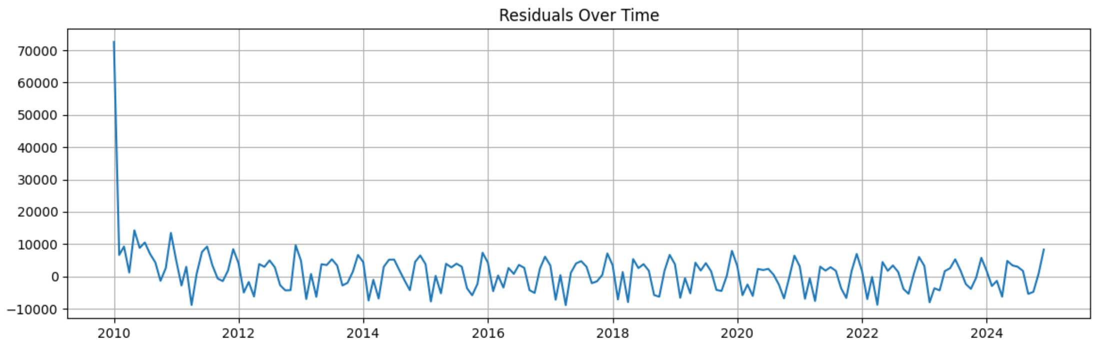
Metric	Value
Count	180
Mean	65,960 million kWh
Standard Deviation	4,774 million kWh
Minimum	54,547 million kWh
Quartile 1	62,714 million kWh
Median	65,721 million kWh
Quartile 3	69,787 million kWh
Maximum	74,649 million kWh

ARIMA Model Selection and Preparation

Before selecting our ARIMA Model we must first check our data and apply transformations and differencing when needed. The data is stable, and no log-transformation is needed. Also, by looking at this behaviour we conclude that including a Constant deterministic component is appropriate. We checked for stationarity, showing no increase or decrease in volatility. This shows an almost stationary series, but the seasonality might cause problems, so we perform the Augmented Dickey-Fuller to test and then difference before fitting ARIMA.

ADF Statistics	-5.96
p-value	0.00…2

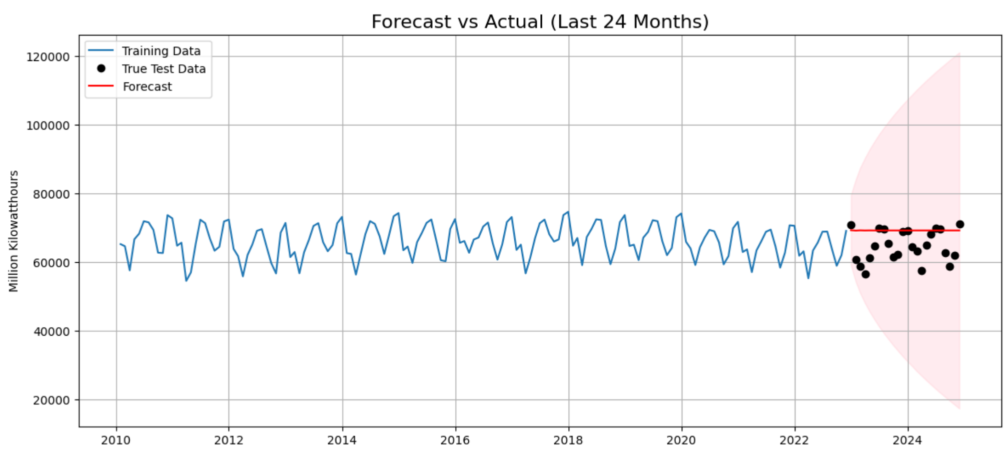
We can now reject the null hypothesis since the series became stationary with our first differencing. Seasonality will most likely be a problem since we are using ARIMA, but the model will be safe to build. To know our MA and AR lags, we plot ACF and PACF:
We see big spikes at 1 and some small spikes later but not too strong, suggesting AR(1) structure. So: p=1 d=1 q=1. Since we are using d = 1. We see that our ARIMA(1,1,1) is very reasonable for this data based on AIC and BIC, but we notice: Jarque-Bera (JB p=0.28) Not significant (Residuals are roughly normal).



After plotting the residuals we see that at the very beginning there is a huge spike (outlier). Since the outlier is a clear anomaly and not representative of our data, we decide to drop the first row and refit the model: AIC: 3559.868, BIC: 3569.414, Ljung-Box p=0.97: No autocorrelation in residuals, Jarque-Bera p=0.10: Residuals approximately normal.

<u>Check Forecasting Performance on the last 24 months</u>

Now we create a test set containing the last 24 months of our data, and a training set containing the other 155 months. We fit the ARIMA to the training set, then forecast on the test set.

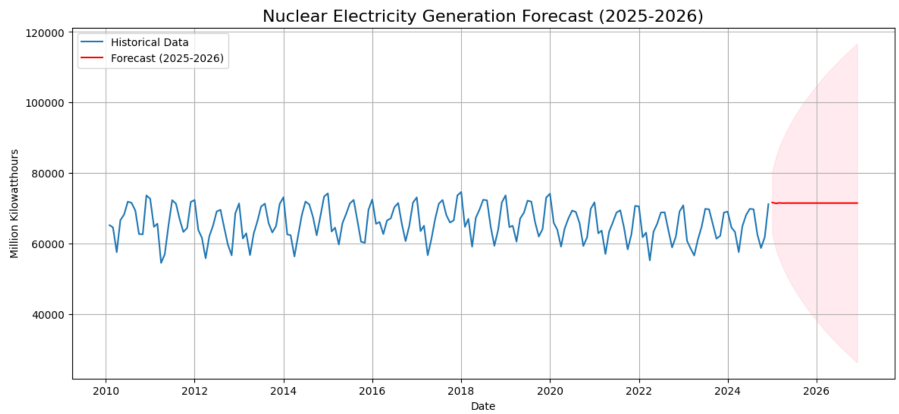


MAE: 5020.92, RMSE: 6353.64

Our model is reasonable but not perfect. We are already starting to see the problems of using ARIMA due to seasonality. Our true observed test values are more volatile than the model predicted, but never move out of our confidence interval. On average our forecast misses around 5,021 million kWh, but given that the typical monthly value sits around 60,000 - 70,000 million, our errors are less than 10%, which is acceptable.

Forecast 24 months out of sample

Now we refit our ARIMA to the full data (February 2010 - December 2024) to forecast ahead:

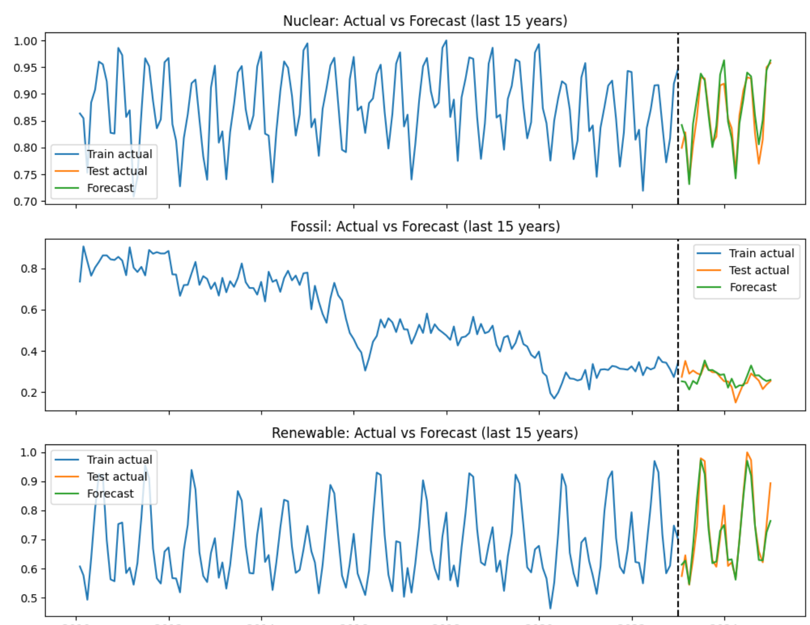


Here we finally confirm the issue of using ARIMA with data that goes in seasonal cycles. Even though our ARIMA model is completely appropriate, we run into the issue of simplicity. We will compare these results against the results of an appropriate VAR model in the last part of Problem 1.

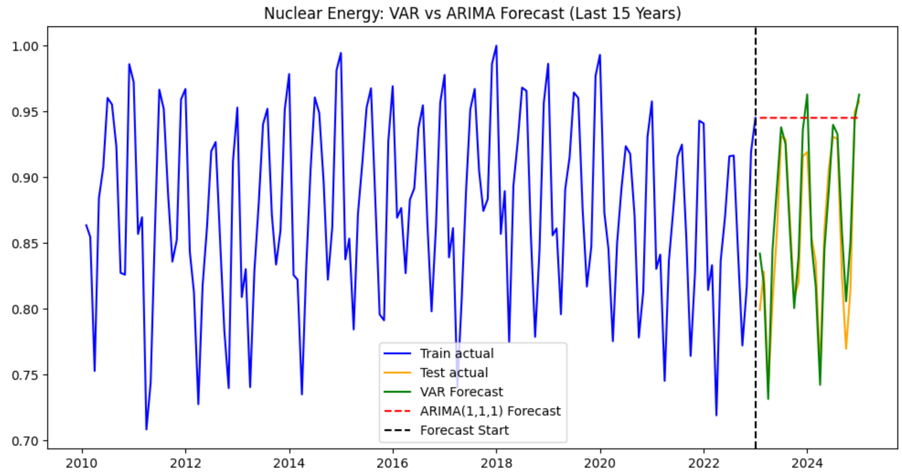
Build & Analyze a VAR model

Now we create a new table containing our Nuclear data, and also data about Fossil and Renewable electricity indicators, ranging from 1973 to 2025. We cleaned and made the data stationary, so we could proceed on creating an appropriate VAR model.

Using BIC and HQIC criteria we find that the appropriate number of lags for our model is 13:



Here we see some slight overfitting, but overall a very strong forecast. Lets compare with our ARIMA(1,1,1):

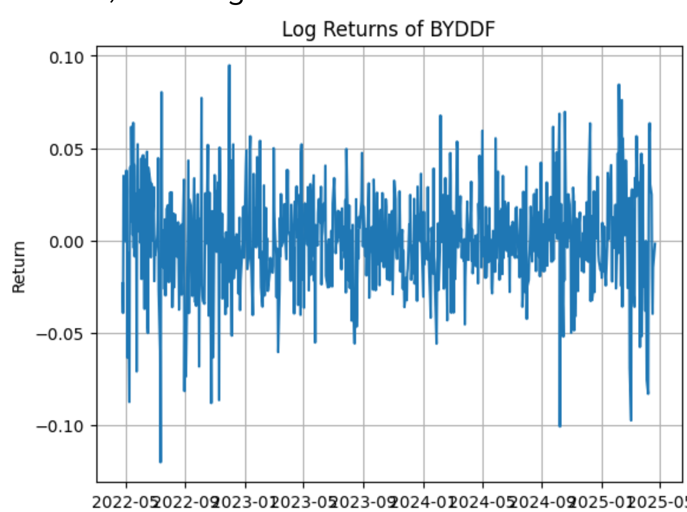


Clearly we see a very important improvement when forecasting with our VAR(13).

PROBLEM 2

For this analysis, data was pulled from Yahoo Finance using the finance package. For this specific project, we chose to use the stock ticker BYDDF which belongs to a chinese manufacturing company, and extracted a total of 3 years of data. After performing the checks to ensure the data pulled was not empty, we proceed with the modelling.

Firstly the log returns were computed using the numpy library, specifically the np.log function applied to the prices and then using the .diff funcion, getting the difference between the cell and its previous value, meaning it returns the return for each day compared to the previous one..



As it can be seen from the log returns for this specific stock,we can see that between 2023 and 2024 the volatility was lower when compared to 2022 and 2025, probably due to normalization after covid and before the economical turbulence that rose during this year. Although this, no clear trend can be inferred from this graph directly, outside of the return averaging around 0 from a glance.

Then the GARCH model was built. It is worth mentioning that the GARCH model can render useful for this type of use cases, since it is used to model volatility over time, which can shed some light on stock markets which may appear unpredictable. In this project, a GARCH(1,1) was used, which means that it reacts to the latest “shock” (squared return) from the system and the volatility from the previous day. After fitting the model, we get the following result

Mean Model

	Coef.	Std. err.	t	P>|t|	95% conf. int.
omega	0.0466	0.0901	0.517	0.605	[-0.130, 0.223]

Volatility Model

	Coef.	Std. err.	t	P>|t|	95% conf. int.
omega	0.1412	0.308	0.459	0.646	[-0.46, 0.74]
alpha	0.0564	0.073	0.776	0.438	[-0.086, 0.199]
beta	0.9251	0.111	8.346	7.05 e-17	[0.708, 1.142]

From the mean model part, it can be interpreted that the model estimates a very small return, of around 0.47. This, however, cannot be used since having a p-value bigger than 0.05, means it does not hold strong statistical significance.
For the volatility model,that omega and alpha have both a big p-value, notedly bigger than 0.05. This means for omega that volatility is more closely related to recent values than to a predetermined baseline, and for alpha it means that daily shocks do not have a strong impact on the volatility.
However, in the case of beta the p-value is miniscule, which indicates that there is a very high impact on the current volatility given by the one in the previous day. Also performing the check of alpha+beta, which is close to 1, it can be inferred that the volatility is persistent but is bound to stabilize, which is usual in the case of financial time series

Finally, after assessing the model, the prediction for the volatility of the next day can be done, giving a forecasted volatility of 3.6539% for the BYDDF stock.

Appendix:
Data Preparation and Exploratory Analysis: