Forecasting Electric Production

Francisco Gomes 20221810, Maria Henriques 20221952, Marta Almendra 20221878.

Bachelor in Data Science, Nova IMS

Introduction

Electric production is a critical factor in maintaining economic stability and supporting industrial activities. Accurate forecasting of electric production can help in planning and optimizing resources, thereby ensuring a reliable supply of electricity to meet demand. This project aims to develop a **time series model** to forecast electric production with the use of **Box-Jenkins Methodology**.

The United States electric production industry is crucial to the stability and growth of the economy. It encompasses various sectors including power generation, distribution, and management. This industry is significantly impacted by seasonal variations, which reflect changes in energy consumption patterns due to weather conditions, economic activities, and other factors throughout the year. By analyzing these dynamics, stakeholders can gain valuable insights into energy demand, enabling them to plan and optimize resource management strategies effectively.

Tentative Identification

Monthly data on the average electric production was sourced from the Kaggle platform. The data was then turned into a tsibble object, and, to gain a better understanding, an initial graph was created to visualize the patterns and trends present. The data was clearly non-stationary, which was later addressed, with a **rising trend** and **seasonality**, as seen in Figure 1.

Furthermore,the data was split into two sets: a training set (data from January 1985 to December 2015) for model estimation, and a test set (data from January 2016 to January 2018) for evaluating future projections.

.

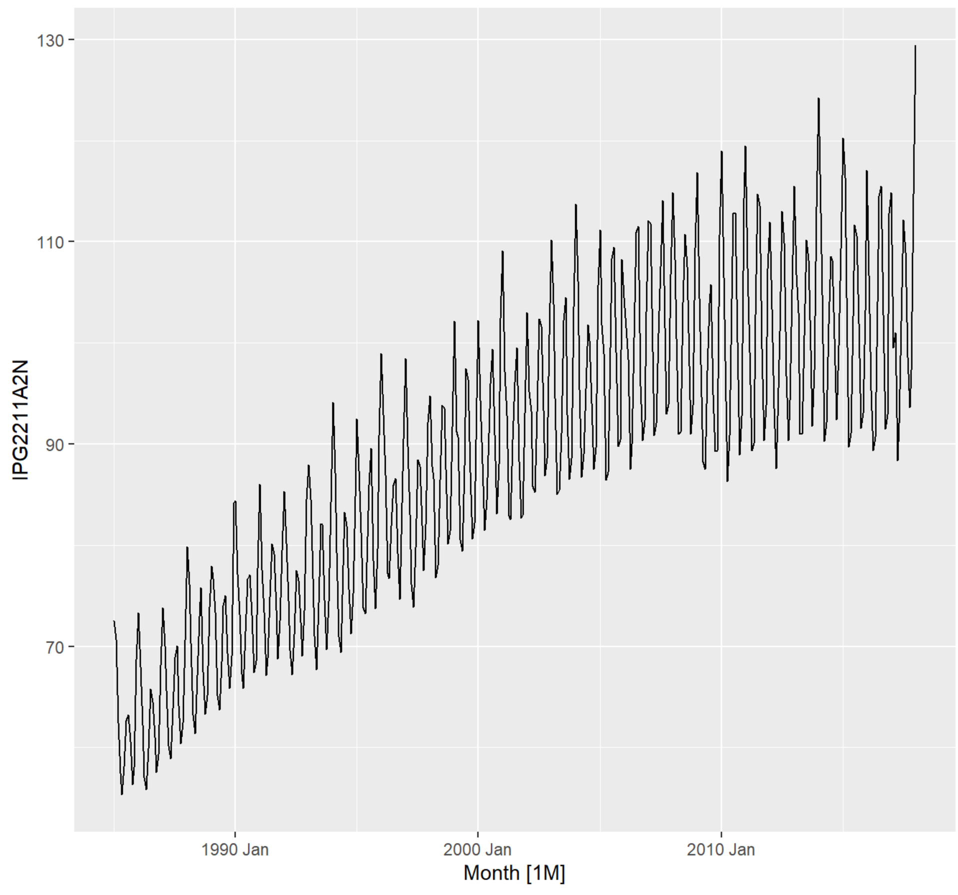

Figure 1: Eletric prodution throughout the years

Since the variance throughout the dataset was not constant, a logarithm transformation was applied.

Moreover, to address non-stationarity, one **seasonal difference transformation** was applied. Since it became stationary after this step, we proceeded to examine the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to identify potential candidate models.

Strategies to choose the best model passed through minimizing information criteria (such as AIC or BIC), analyzing the residuals of the models, and conducting the Ljung-Box test to check for autocorrelation in the residuals. Finally, the accuracy of the more promising models was analyzed, and a final model was selected based on the evaluation criteria and the overall performance of the model in forecasting the test set.

Results

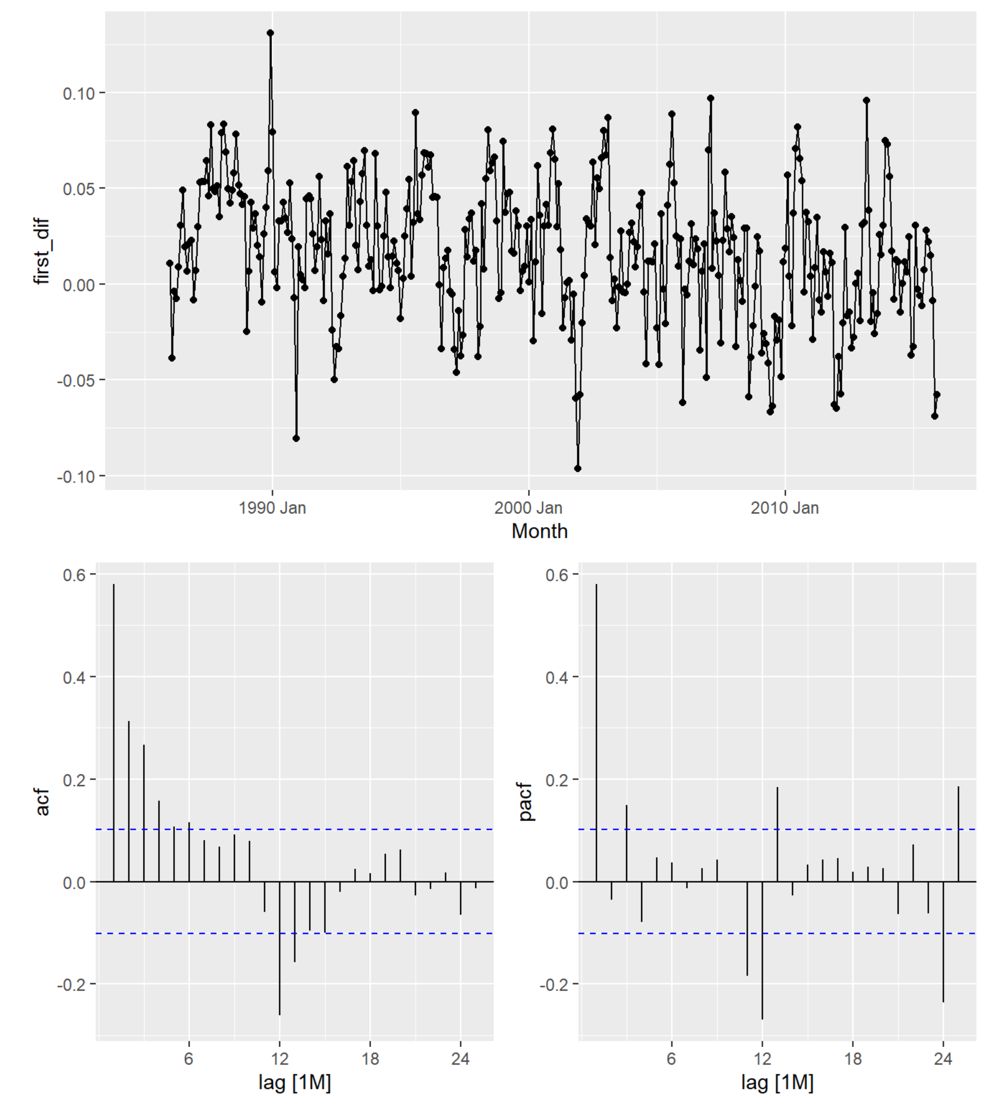

Figure 2: ACF and PACF of the first seasonal difference

As previously mentioned, after achieving stationarity, the ACF and PACF were examined (Figure 2). The ACF displayed significant spikes at lag 1 and 2, which suggested the presence of an MA component of order 1 or 2. Regarding the PACF, it showed significant spikes also at lag 1 and 2, pointing towards an AR component of order 1 or 2. After this examination some promising models were analysed.

Based on **information criteria**, we identified the best models in the previous set, which were SARIMA(2,0,1)(2,1,1), SARIMA(1,0,1)(1,1,1), and SARIMA(2,0,0)(2,1,1).

Table 1: Model Error Measures on the Testing Sample

.model	ME	RMSE	MAE	MPE	MAPE
sarima101111	4.4867063	6.140555	5.124062	4.1483198	4.797036
sarima200211	4.4372547	5.765929	4.927195	4.1223170	4.619170
sarima201211	0.2544459	3.679769	2.622908	0.0531509	2.441619

Considering these were the best models, all three were submitted to Ljung-Box and their residuals shown; nevertheless, only some were proven to be a white noise process, and only one model passed the **Ljung-Box** test.

Forecasts were then executed for the best models, and the results are shown in Figure 3. In terms of prediction accuracy, the findings demonstrate that SARIMA(2,0,1)(2,1,1) had the best outcome, as expected given that it was the only model with no autocorrelation in its residuals (passed the Ljung-Box test).

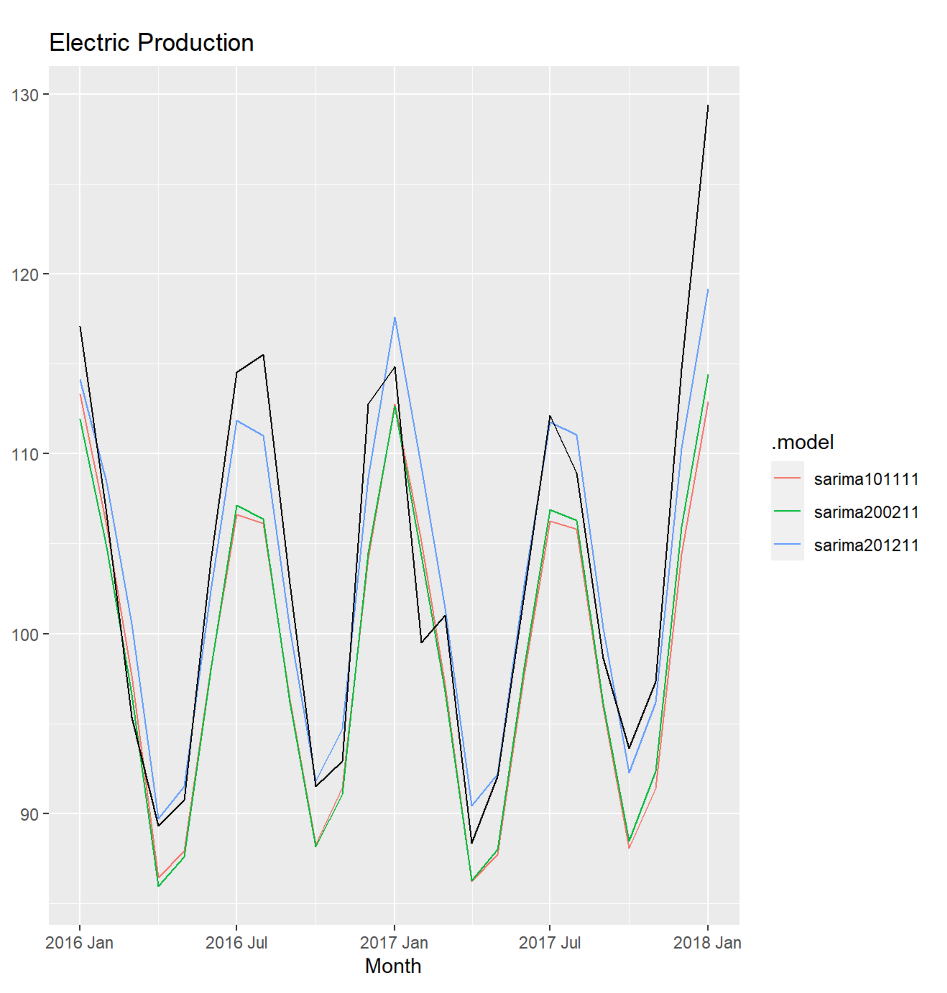

Figure 3: Forecast with best models

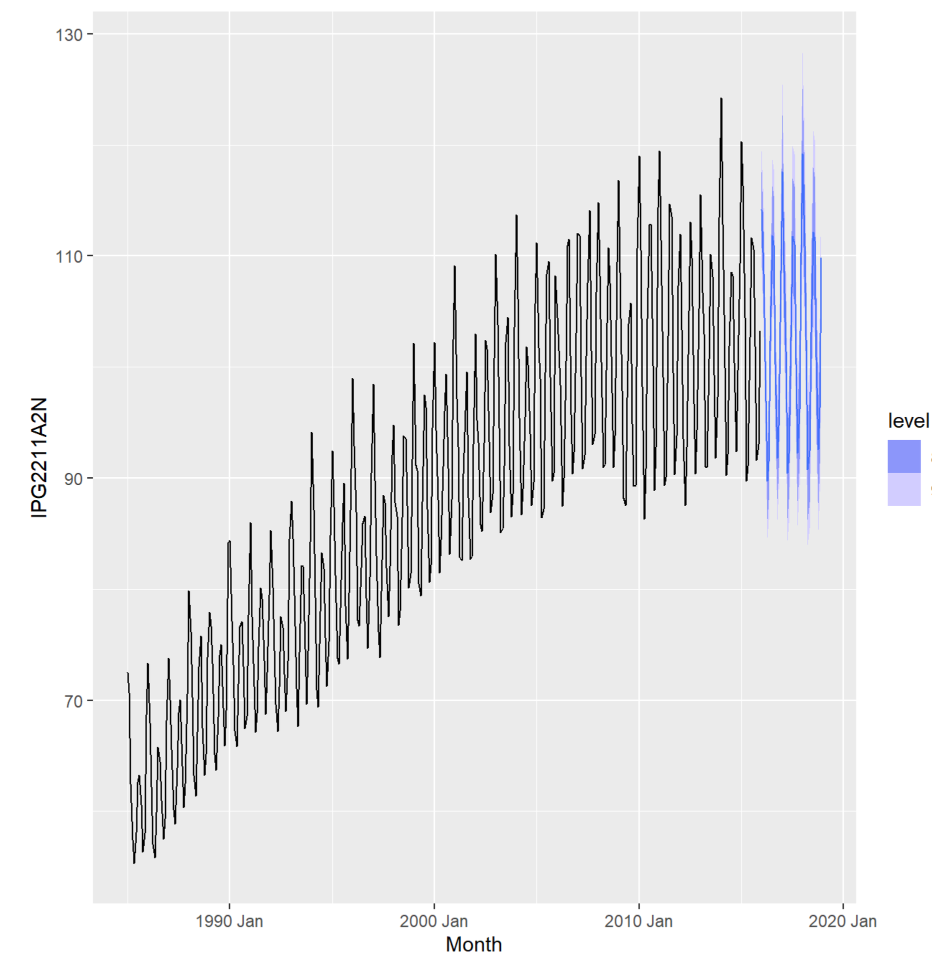

Figure 4: Forecast with the SARIMA Model (2,0,1)(2,1,1)

Conclusion

In conclusion, the study seems to have successfully identified a forecasting model for the usual electric production which is displayed in the obtained results (Figure 4).

This model provides useful insights for electric firms, assisting with resource planning and optimization by forecasting shifts in electrical demand. Our forecasts for the future years suggests that electric production values will show similar behaviors as the previous years. As a result, electric firms may utilize these projections to improve their labor strategy, increase operational efficiency, and maintain a competitive advantage in the volatile electric market.

Furthermore, the forecasting model can assist firms in detecting possible periods of high demand, allowing them to better manage resources and prevent any supply shortages. This proactive strategy not only promotes improved decision-making, but it also helps to ensure the industry's long-term growth and stability. Electric firms may use these information to better satisfy consumer expectations and adjust to market changes.

References

1. Makridakis, S., Hibon, M. ARMA Models and the Box-Jenkins Methodology. Journal of Forecasting. 16(3), 147–163, 1997;

2. https://www.w3schools.com/cssref/css_colors.php

3. https://fred.stlouisfed.org/series/IPG2211A2N

Data acquired from:

https://www.kaggle.com/datasets/mwafia/electric-production