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 of it, an initial graph was created to visualize patterns and trends that might be present. The data was clearly non-stationary, which was later addressed,it depicted a **rising trend** and showed **seasonality**, as seen in Figure 1.

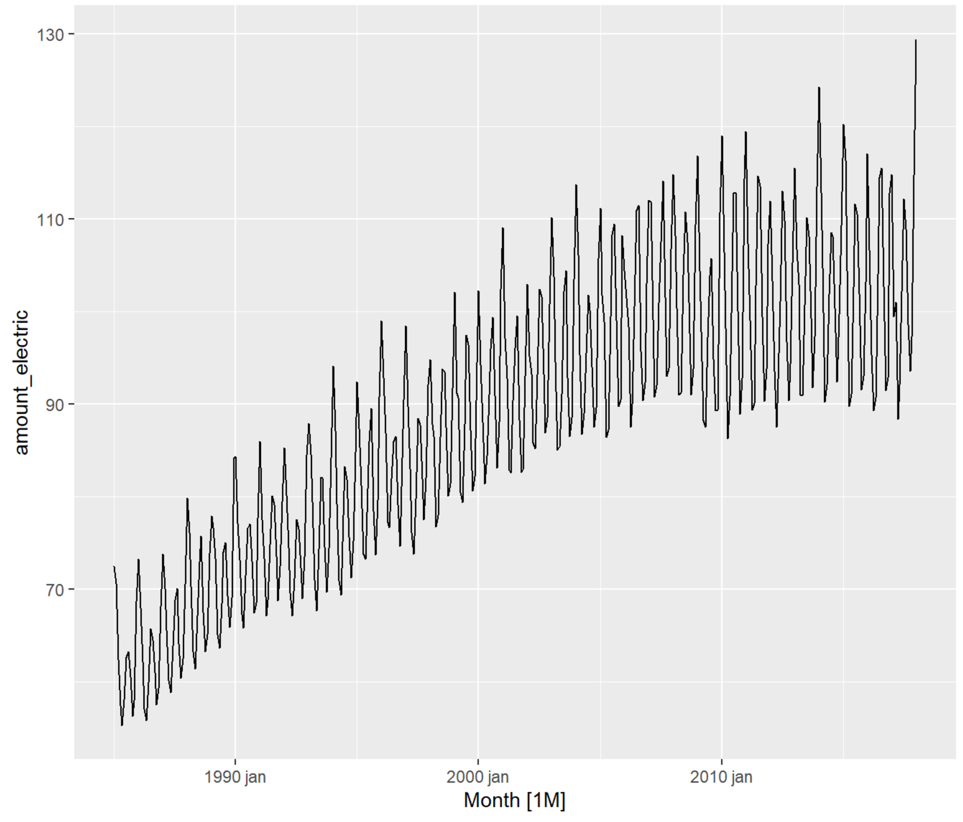

Figure 1: Eletric prodution throughout the years

Furthermore, the data was split into two sets to aid in the next phases: a training set that contained data from January 1985 to December 2015 for model estimation, and a test set with data from January 2016 to January 2018 for evaluating future projections.

By assessing the dataset's variance, we discovered that it was in fact not constant over time, hence a logarithmic adjustment was used to stabilize it.

Moreover, to address non-stationarity, one **seasonal difference transformation** was applied. Since it became stationary after this step as confirmed through the **Augmented Dickey-Fuller** test, 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 from among those tested passed through minimizing information criteria, such as Akaike information criterion (AIC) and Bayesian information criterion (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 determined, and a final model was selected based on its accuracy and overall performance in forecasting the test set.

Results

As previously mentioned, after achieving stationarity, the ACF and PACF were examined (Figure 2), and various ARIMA models were selected and analyzed.

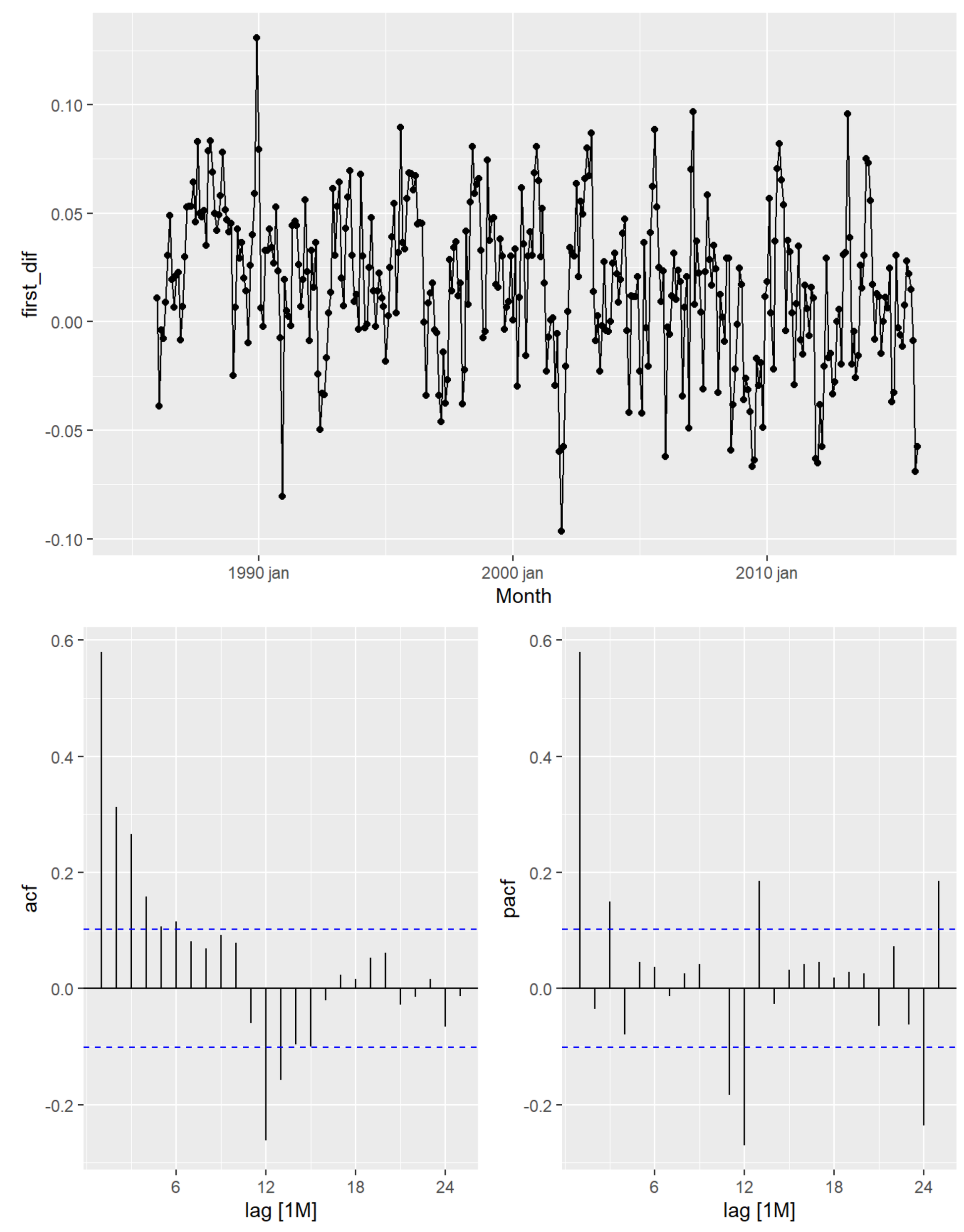

Figure 2: ACF and PACF of the first seasonal difference

Based on **information criteria**, we identified the best models in the previous set, which were SARIMA(1,0,1)(1,1,1), SARIMA(2,0,1)(1,1,1), and SARIMA(1,0,1)(0,1,1), since they presented the lowest values for AIC and BIC.

Considering these were the best models, all three were submitted to **Ljung-Box test** and their residuals analyzed; nevertheless, only one was proven to be a white noise process and pass the Ljung-Box test, while the others showed statistical evidence of autocorrelation in the residuals.

Forecasts were then executed for the best models and compared to a benchmark method, with the respective results shown in Figure 3.

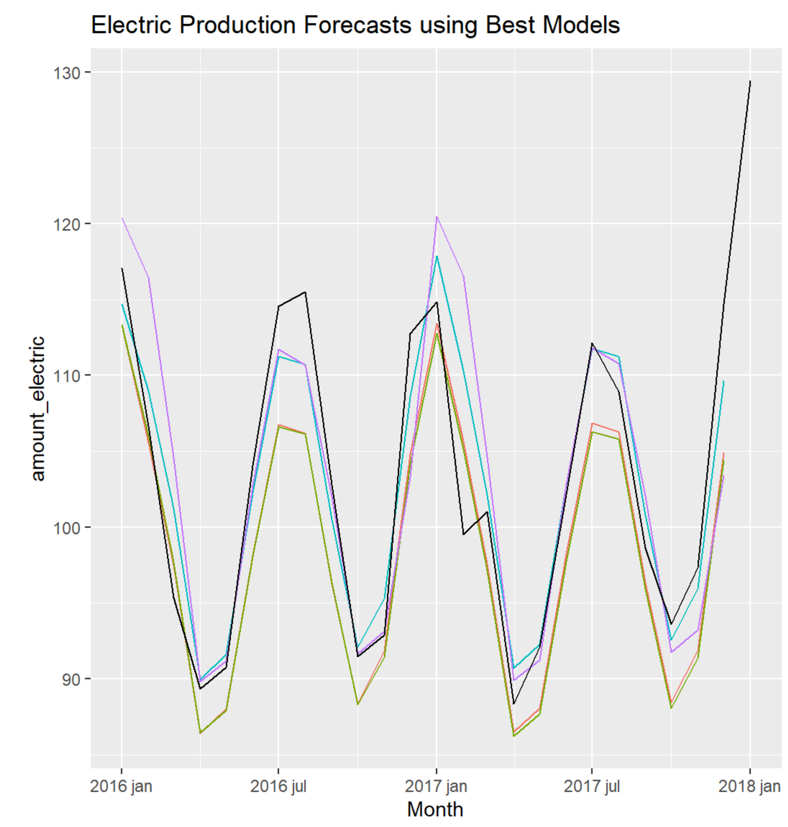

Figure 3: Forecast with best models

In terms of prediction accuracy, the findings suggest that SARIMA(2,0,1)(1,1,1) provided the best results, with smaller RMSE and MAE errors and a ME close to zero, indicating minimum bias. This was not surprising considering that it was the only model with no autocorrelation in its residuals (passed the Ljung-Box test). Therefore, **SARIMA(2,0,1)(1,1,1)** was the **final proposed model**.

Table 1: Model Error Measures on the Testing Sample

.model	ME	RMSE	MAE	MPE	MAPE
sarima101011	3.7204904	5.067816	4.413700	3.5319831	4.236504
sarima101111	3.9849500	5.281482	4.648862	3.7889586	4.464705
sarima201111	−0.3629319	3.439818	2.576074	−0.4854087	2.483093
snaive	−0.8164647	5.841415	3.962346	−0.9096964	3.778021

Conclusion

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

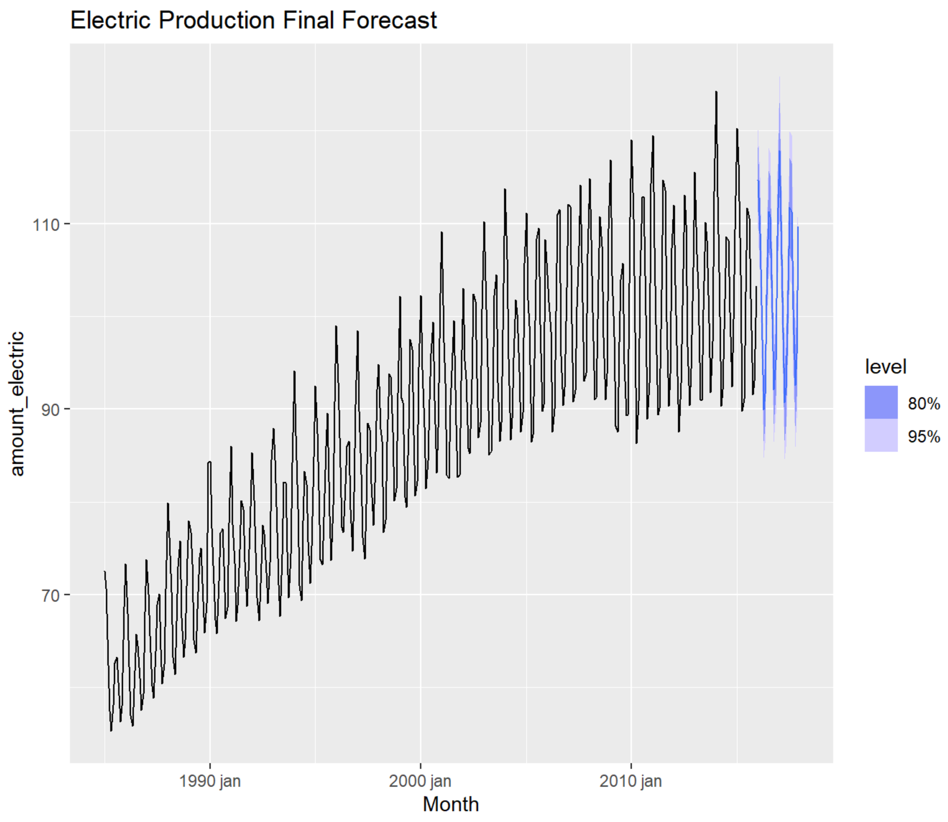

Figure 4: Forecast with the SARIMA Model (2,0,1)(1,1,1)

This model provides useful insights for electric firms, assisting with resource planning and optimization by forecasting shifts in electrical demand. Our forecasts suggest 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 this information to better satisfy consumer expectations and adjust to market changes.

References

1. Time Series Basics and Getting Started with R
2. Forecaster's Toolbox
3. Introduction to Time Series. ARMA models
4. Box - Jenkins methodology. Forecasting with ARIMA models. Introduction to Seasonal ARIMA models
5. Industrial Production: Utilities: Electric and Gas Utilities (NAICS = 2211,2). (2024b, May 16) https://fred.stlouisfed.org/series/IPUTIL
6. Brentthorne. (n.d.). posterdown_html. GitHub. https://github.com/brentthorne/posterdown/wiki

Data acquired from:

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