Report on Electric Production Forecast : A Comparison of Different Time Series Models

Esther Waweru - 171444

2024-07-27

Introduction

Electricity demand forecasting is crucial for planning and expanding facilities in the electricity sector. Accurate forecasts can reduce operating and maintenance costs, enhance the reliability of the power supply and delivery system, and support informed decisions for future development.Time series analysis provides a robust framework for predicting future electricity production based on historical data. This report compares different time series models, including Holt-Winters exponential smoothing, Autoregressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), and Seasonal ARIMA (SARIMA), highlighting their strengths, weaknesses, and suitability for electricity production forecasting.

Methods

- **Data Source**

The dataset under analysis is from kaggle and pertains to the industrial production of electric and gas utilities in the United States, spanning the years 1985 to 2018. This monthly production output data provides a comprehensive overview of the energy sector's performance over more than three decades

- **Data Description**

From the plot above,the data exhibits clear seasonal patterns, a rising trend, and some irregular fluctuations. There is a noticeable upward trend in the production levels indicating a steady growth in production of electric and gas utilities. It smooths out the short term fluctuations and highlights the underlying direction of the series. Consistent cyclic patterns that repeat every year are observed, with peaks and troughs at similar times each year.

- **Check for Stationarity**

Using Augmented Dickey Fuller Test :

The Hypothesis:

H_0 : The time series data is not stationary

H_1 : The time series data is stationary

The p-value is less than 0.05 hence we reject the null Hypothesis and conclude that our series is stationary.

However, while performing Kwiatkowski-Philips-Schmidt-Shin (KPSS) test with hypothesis :

H_0 : The time series data is stationary

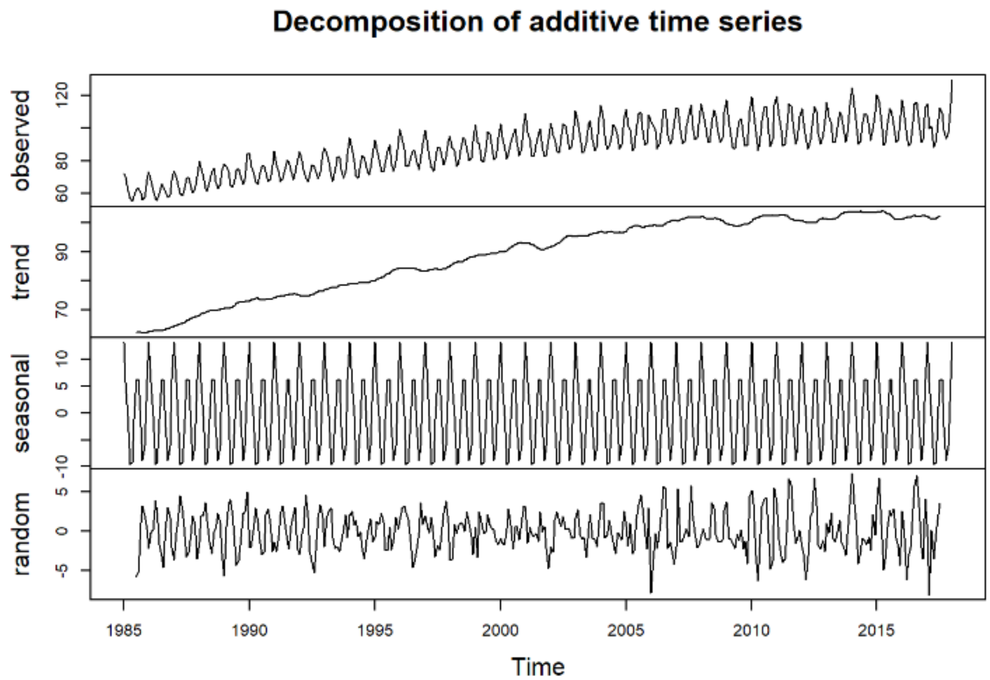


Figure 1: Time Series

H_1 : The time series data is not stationary

The p-value is less than 0.05 hence we reject the null Hypothesis and conclude that our series is non-stationary.

ADF shows that our series is stationary while KPSS-test shows that it is non-stationary. This indicates difference stationary hence we perform differencing to make our series stationary.

- **Forecasting**

1. Holt-Winters exponential smoothing

Holt-Winters exponential smoothing is a time series forecasting method that extends simple exponential smoothing to capture trends and seasonality in the data. It is particularly useful for data with clear seasonal patterns and trends, making it a versatile tool for a wide range of forecasting applications. Given the nature of this data, which exhibits both a trend and seasonality, Holt-Winters exponential smoothing is an appropriate method for forecasting.

2. Autoregressive (AR) model

An Autoregressive (AR) model is a type of time series model that predicts future values based on past values of the series. In an AR model, the current value of the time series is expressed as a linear combination of its previous values plus a random error term. This model is particularly useful for time series data that exhibits temporal dependencies.AR model can be applied to capture the temporal dependencies in the data. Given the monthly frequency of the data, an appropriate lag order p - the number of lagged observations included in the model- might be necessary to capture the dependencies across months.

3. Moving Average

A Moving Average (MA) model is another fundamental type of time series model used for forecasting. Unlike the Autoregressive (AR) model, which uses past values of the series for predictions, the MA model uses past forecast errors to make predictions. This model captures the shock or noise in the data series and is particularly effective in modeling time series with short-term dependencies.MA model can be applied to capture the short-term dependencies and noise in the data. Given the monthly frequency, the appropriate lag order q - the number of lagged forecast errors included - can be determined using the Autocorrelation Function (ACF) plot.

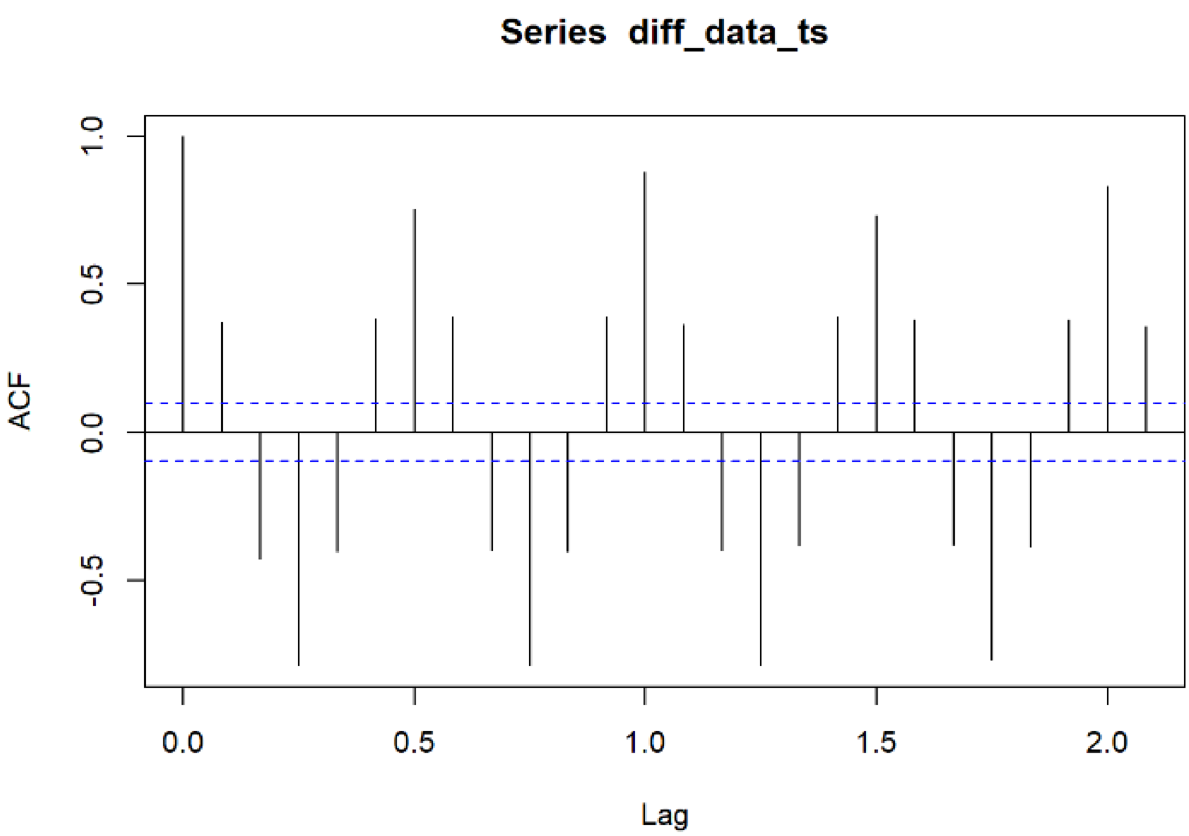
4. Autoregressive Integrated Moving Average (ARIMA) Model

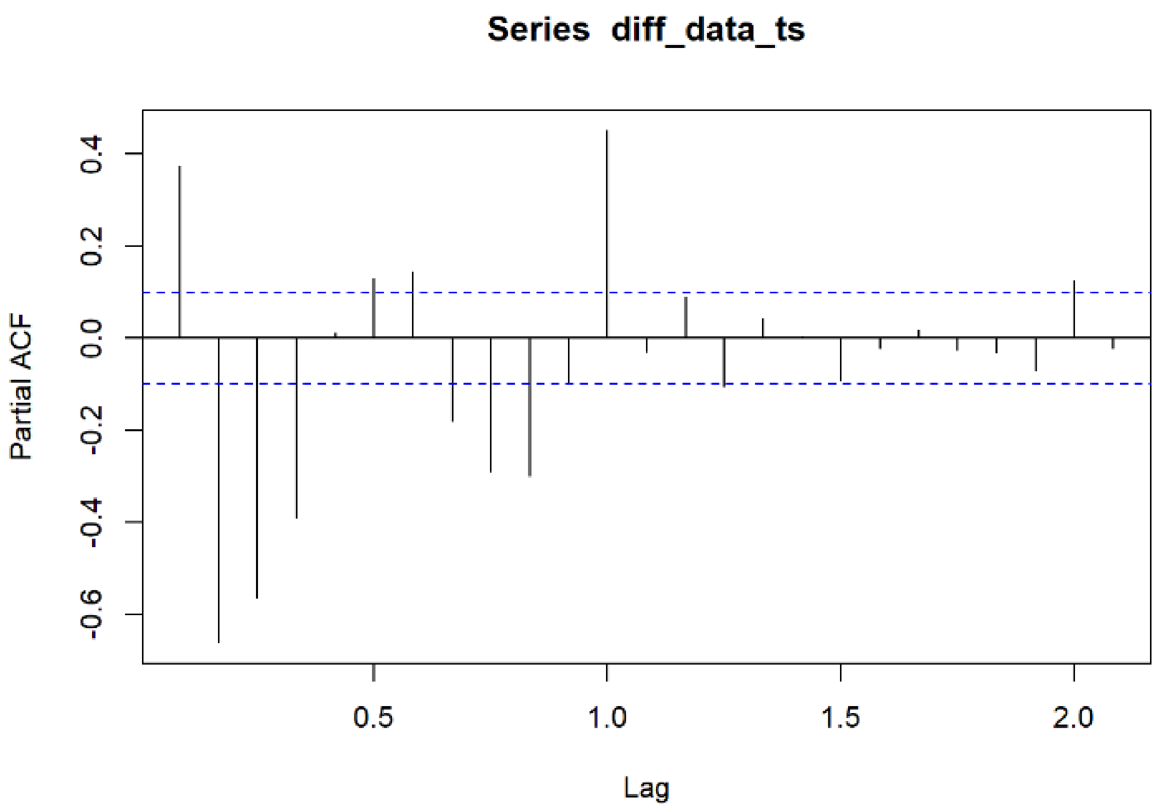
The Autoregressive Integrated Moving Average (ARIMA) model is a popular and versatile time series forecasting method that combines the principles of autoregression (AR), differencing (to achieve stationarity), and moving averages (MA). The ARIMA model is particularly useful for time series data that exhibit non-stationarity and can capture both long-term trends and short-term dependencies.

5. Seasona ARIMA model

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model extends the ARIMA model to handle data with seasonality. It incorporates both non-seasonal and seasonal components, making it well-suited for time series data that exhibit regular, periodic patterns over time.

For the models : AR, MA , ARIMA , SARIMA the values of p,d,q are determined by ACF, PACF plots and differencing.





The first differencing appears to have made the series stationary, hence d = 1.

The PACF plot shows significant spikes at the first few lags, then cuts off. Here, the significant spike at lag 1 suggests p = 1.

The ACF plot shows significant spikes at the first few lags, then gradually dies off. Here, the significant spikes at lag 1 suggest q = 1.

Results and Discussion

- **Box Ljung Test**

The Hypothesis:

H_0 : The residuals are independently distributed (no significant autocorrelation).

H_1 : The residuals are not independently distributed (significant autocorrelation present).

Models	P_value
Holt-Winters exponential smoothing	0.0023650
Autoregressive (AR)	0.0000006
Moving Average	0.4996000
Autoregressive Integrated Moving Average (ARIMA)	0.5807000
Seasonal Arima	0.3233000

For the models : Holt-Winters exponential smoothing,Autoregressive (AR) , the p-value is less than 0.05 hence we reject the null hypothesis and conclude that the residuals of these models are not independently distributed.

For the models : Moving Average , Autoregressive Integrated Moving Average (ARIMA) , Seasonal ARIMA, the p-value is greater than 0.05 hence we fail to reject the null hypothesis and conclude residuals of these models are independently distributed.

- **Performance metrics**

Models	RMSE	MAE	MAPE
Holt-Winters exponential smoothing	2.7179	2.0828	94.3619
Autoregressive (AR)	7.1849	6.0360	203.2847
Moving Average	6.5974	5.4458	212.7252
Autoregressive Integrated Moving Average (ARIMA)	8.4427	6.8440	436.6483
Seasonal Arima	2.5952	1.9300	91.4431

RMSE (Root Mean Squared Error): Measures the square root of the average of squared differences between predicted and observed values. Lower RMSE indicates better fit.

MAE (Mean Absolute Error): Measures the average of absolute differences between predicted and observed values. Lower MAE indicates better fit.

MAPE (Mean Absolute Percentage Error): Measures the average of absolute percentage errors between predicted and observed values. Lower MAPE indicates better fit.

Holt-Winters exponential smoothing performs well with relatively low RMSE and MAE, and a moderate MAPE. The AR model has higher RMSE, MAE, and MAPE compared to other models, indicating it does not perform as well. The MA model performs slightly better than the AR model but still has high error metrics.The ARIMA model has the highest RMSE, MAE, and MAPE, indicating it performs poorly for this dataset.The SARIMA model has the lowest RMSE, MAE, and MAPE, indicating the best performance among all the models.

Conclusion

In our analysis of electricity production forecasting models, the Seasonal ARIMA (SARIMA) model emerged as the most effective, outperforming Holt-Winters exponential smoothing, Autoregressive (AR), Moving Average (MA), and standard ARIMA models. The SARIMA model demonstrated the lowest RMSE, MAE, and MAPE values, indicating superior accuracy in capturing both seasonal patterns and long-term trends in the dataset. Therefore adopting the SARIMA model for electricity production forecasting would be ideal to enhance planning and operational efficiency, ensure regular model evaluation, maintain data quality, and integrate forecasting insights into strategic decision-making processes.

Appendix

Code

```
## ----setup, include=FALSE------------------------------------------------
knitr::opts_chunk$set(echo = TRUE)


## ---------------------------------------------------------------------
library(TSstudio)
library(dplyr)
library(ggplot2)
```

```r
library(forecast)
library(tseries)
library(urca)
library(zoo)
library(summarytools)


## ----------------------------------------------------------------------------
setwd('C://Users//pc//Documents//Masters_Statistical_Sciences//Module 3//Time Series Analysis//Semester

data <- read.csv('Electric_production.csv')
head(data)


## ----------------------------------------------------------------------------
#Change column names
colnames(data) <- c("date","value")

#Convert data column to date datatype
data$date <- as.Date(data$date, format="%m/%d/%Y")

str(data)


## ----------------------------------------------------------------------------
descr(data)


## ----------------------------------------------------------------------------
#Create Time series object

data_ts = ts(data = data$value,start = c(1985,1),frequency = 12)

glimpse(data_ts)


## ----------------------------------------------------------------------------
start(data_ts)


## ----------------------------------------------------------------------------
end(data_ts)


## ----------------------------------------------------------------------------
frequency(data_ts)


## ----------------------------------------------------------------------------
#check missing values
colSums(is.na(data))
```

```r
## ----------------------------------------------------------------------------
# Plot the time-series graph
plot(data_ts,main = "Electricity Production Over the Years",
        ylab = "Electricy Production')",
        xlab = "Year")

abline(reg=lm(data_ts~time(data_ts)))


## ----------------------------------------------------------------------------
#Decompose the series into its three components: trend,seasonal, residual

data_decomp <- decompose(data_ts)

plot(data_decomp, )


## ----------------------------------------------------------------------------
boxplot(data_ts~cycle(data_ts, xlab="Date", ylab = "Electricity Production", main = "Monthly electricit


## ----------------------------------------------------------------------------
#Augmented Dickey Fuller Test

adf_test <- adf.test(data_ts)
print(adf_test)



## ----------------------------------------------------------------------------
#Kwiatkowski-Philips-Schmidt-Shin (KPSS)

kpss_test <- kpss.test(data_ts)

print(kpss_test)


## ----------------------------------------------------------------------------
#differencing the series

diff_data_ts <- diff(data_ts)

#Retest for staionarity
adf_test_2 <- adf.test(diff_data_ts)
print(adf_test_2)


kpss_test_2 <- kpss.test(diff_data_ts)
print(kpss_test_2)


## ----------------------------------------------------------------------------
decomp <- decompose(diff_data_ts)
```

```r
plot(decomp, )


## ----------------------------------------------------------------------------
#Apply the Holt-Winters Seasonal Model with additive seasonality
hw_model_add <- hw(diff_data_ts, h = 24 , seasonal = "additive")
# Plot the forecasts from the additive model
plot(hw_model_add, main="Holt-Winters Additive Seasonal Model Forecast")



## ----------------------------------------------------------------------------
accuracy(hw_model_add)


## ----------------------------------------------------------------------------
#check for autocorrelation
Box.test(hw_model_add$residuals,type = "Ljung-Box")


## ----------------------------------------------------------------------------
acf(diff_data_ts)


## ----------------------------------------------------------------------------
pacf(diff_data_ts)


## ----------------------------------------------------------------------------
moving_average <- arima(diff_data_ts,order = c(0,0,1))
summary(moving_average)



## ----------------------------------------------------------------------------
#check for autocorrelation
Box.test(moving_average$residuals,type = "Ljung-Box")


## ----------------------------------------------------------------------------
#Forecast using MA model
ma_forecast <- forecast :: forecast(moving_average,h=24)
plot(ma_forecast)


## ----------------------------------------------------------------------------
ar_model <- arima(diff_data_ts, order = c(1,0,0))
summary(ar_model)


## ----------------------------------------------------------------------------
#check for autocorrelation
Box.test(ar_model$residuals,type = "Ljung-Box")
```

```r
## -----------------------------------------------------------------------------
#Forecasting

ar_forecast <- forecast::forecast(ar_model, h = 24)
plot(ar_forecast)


## -----------------------------------------------------------------------------
arima_model <- arima(diff_data_ts, order = c(1,1,1))
summary(arima_model)


## -----------------------------------------------------------------------------
#check for autocorrelation
Box.test(arima_model$residuals,type = "Ljung-Box")


## -----------------------------------------------------------------------------
arima_forecast <- forecast::forecast(arima_model,h=24)
plot(arima_forecast)


## -----------------------------------------------------------------------------
sarima_model <- arima(diff_data_ts, order = c(1, 1, 1), seasonal = list(order = c(1, 1, 1),period = 12))
summary(sarima_model)


## -----------------------------------------------------------------------------
#check for autocorrelation
Box.test(sarima_model$residuals,type = "Ljung-Box")


## -----------------------------------------------------------------------------
sarima_forecast <- forecast::forecast(sarima_model,h=24)
plot(sarima_forecast)
```