A Comparative Study of Forecasting Accuracy for Sectoral IIP using Simple Exponential Smoothing and ARIMAX with Macroeconomic Factors

-Basisth Saha

(UG-2, Dept. of statistics, Presidency University, Kolkata)

Abstract:

Industrial production is crucial for the Indian economy, and the Index of Industrial Production (IIP) is a key indicator. This paper aims to study the trends in sectoral IIP and establish a comparative study between the forecasting accuracy of these two models: Simple Exponential Smoothing (SES) and Autoregressive Integrated Moving Average with Exogenous factors (ARIMAX). The forecasting accuracy of these two models will be compared to determine their effectiveness in predicting sectoral industrial output, which is crucial for policy makers. The models will be constructed using the R programming language.

1. Introduction

- **1.1 Motivation**

The Index of Industrial Production (IIP) is a crucial indicator of economic activity in India, reflecting the growth of various industrial sectors. This study aims to study the trends in sector wise IIP and compare forecasting efficiency of SES and ARIMAX for sectoral IIP. Studying the trends and accurate forecasting of IIP is essential for policymakers, businesses, and investors to make informed decisions.

- **1.2 Objectives of the study**

1. To analyze the historical trends of sectoral IIP in India.
2. To develop SES and ARIMAX models for forecasting sectoral IIP taking year wise GDP(Gross Domestic Product) and CPI(Consumer Price Index) as the covariates.
3. To evaluate and compare the forecasting performance of the SES and ARIMAX models.

- **1.3 Scope of the study**

This study focuses on studying the trends and the forecasting of sector wise IIP of four key sectors: primary goods, capital goods, manufacturing, and electricity taking GDP and CPI as covariates. The analysis is based on time-series data from 2012-2023.

- **1.4 Research Questions**

The paper aims to answer the following research questions:

1. What are the trends and patterns of sector wise IIP in India from 2012-2024?
2. How do SES and ARIMAX models perform in forecasting sectoral IIP?
3. Which model provides more accurate forecasts for different industrial sectors?

2. Literature review

- **2.1 Overview of Index of Industrial Production (IIP)**

The IIP is a quantum index, the production of items being expressed in physical terms. It is a composite indicator that measures the short-term changes in the volume of production of a basket of industrial products during a given period with respect to that in a chosen base year. The Central Statistical Organisation (CSO) is responsible for the compilation and publication of the Index of Industrial Production (IIP) since 1950. The IIP provides a snapshot of the industrial activity in the economy, covering sectors like primary goods, capital goods, manufacturing, mining, intermediate goods, consumer durables and consumer non durables.

- **2.2 Importance of IIP for the Indian Economy**

1. Economic Policymaking: It provides crucial data for formulating monetary and fiscal policies.
2. National Income Estimation: IIP data is used in the estimation of Gross Domestic Product (GDP).
3. Financial Market Analysis: Investors track IIP growth to assess the health of the industrial sector and make investment choices.

- **2.3 Studies on Forecasting IIP**

Several studies have attempted to forecast IIP using various econometric and statistical techniques. Some studies have employed ARIMA models to capture the time-series properties of IIP, while others have used regression models with leading indicators. More recent studies have explored the use of machine learning or rather deep learning techniques for improved forecasting accuracy.

- **2.4 Applications of SES and ARIMAX Models in Economic Forecasting**

1. SES Models: Simple Exponential smoothing (SES) models are used to capture the trends in the past and base the predictions on historical data, it is simple to implement and understand, earning its wide popularity.
2. ARIMAX Models: Autoregressive Integrated Moving Average with Exogenous Regressors (ARIMAX) models are employed to forecast a time series variable by incorporating the influence of external factors. These models are suitable for capturing the impact of macroeconomic variables on industrial production.

3. Methodologies used

- **3.1 Data Sources**

The study utilizes monthly time-series data on sectoral IIP, year wise GDP and CPI of India.

1. The data for sectoral IIP is obtained from the official website of MoSPI (Ministry of Statistics and Programme Implementation).
2. The data for year wise GDP and CPI of India is obtained from official website for the World Bank's open data.

the year wise IIP for the sectors of primary goods, capital goods, manufacturing, and electricity are used.

- **3.2 Plotting**

for plotting the graphs in this paper the ggplot2 package of R is used.

- **3.4 Model Specifications**

1) SES Model Specification: Exponential smoothing is a popular time series forecasting method known for its simplicity and accuracy in predicting future trends based on historical data. It assumes that future patterns will be similar to recent past data and focuses on learning the average demand level over time. Simple smoothing is a method of forecasting time series using univariate data without a trend or seasonality. One must have a single parameter or smoothing factor so as to check how much the impact of past observations should be minimized.the weight to be given to the current data as well as the mean estimate of the past depends on the smoothing parameter. The formula for simple smoothing is as follows;-

$$s_t = \alpha x_t + (1-\alpha)s_{t-1} = s_{t-1} + \alpha(x_t - s_{t-1})$$

Here,

- s_t: Simple weighted average of current observation x_t(smoothed statistic)
- s_{t-1} : Previous smoothed statistic.
- α : Smoothing factor of data. $\in (0, 1)$
- t : time period

2) ARIMAX Model specification: An ARIMAX model, which stands for AutoRegressive Integrated Moving Average with eXogenous inputs, is an advanced version of the ARIMA (AutoRegressive Integrated Moving Average) model. The ARIMAX model extends the ARIMA framework by integrating exogenous variables, which are external factors that can influence the time series being studied. This integration allows the model to leverage additional information that can significantly enhance forecasting accuracy. The ARIMAX model can be mathematically represented as:

$$Y_t = c + \sum_{i=1}^{p} \Theta_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{k=1}^{m} \beta_k X_{t-k} + \epsilon$$

Here,

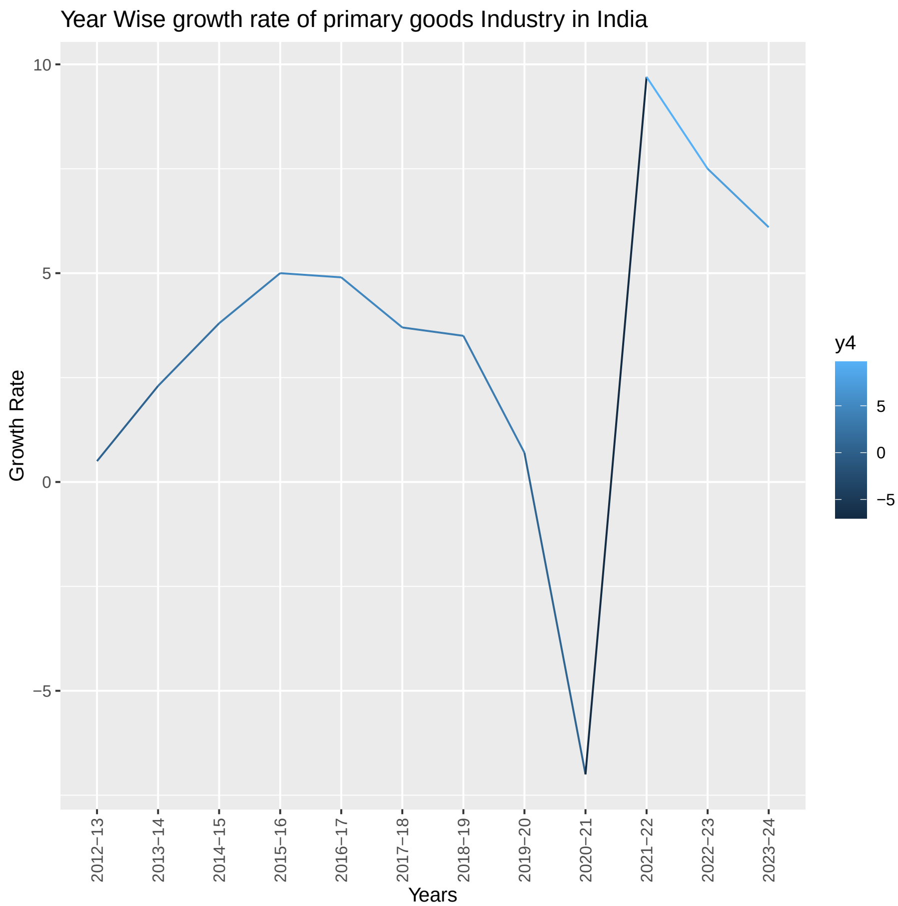
- Y_t : Represents the value of the time series at time t.
- c : Represents the constant intercept.
- Θ_i : Represent the coefficients of autoregressive terms.
- Y_{t-i} : Represent the lagged values of the time series.
- θ_j : Represent the coefficients for the moving average terms.
- ϵ_t : Represent the error term at time t.
- β_k : Represent the coefficients for the exogenous variables.
- X_{t-k} : Represent the lagged values of the exogenous variables.

The variables are split into train and test values with the data for 8 years being the train values used to train the models. Then we will use the models to forecast IIP for the remaining 4 years and compare with the real values for that four years.

4. Trends in the IIP of the 4 sectors

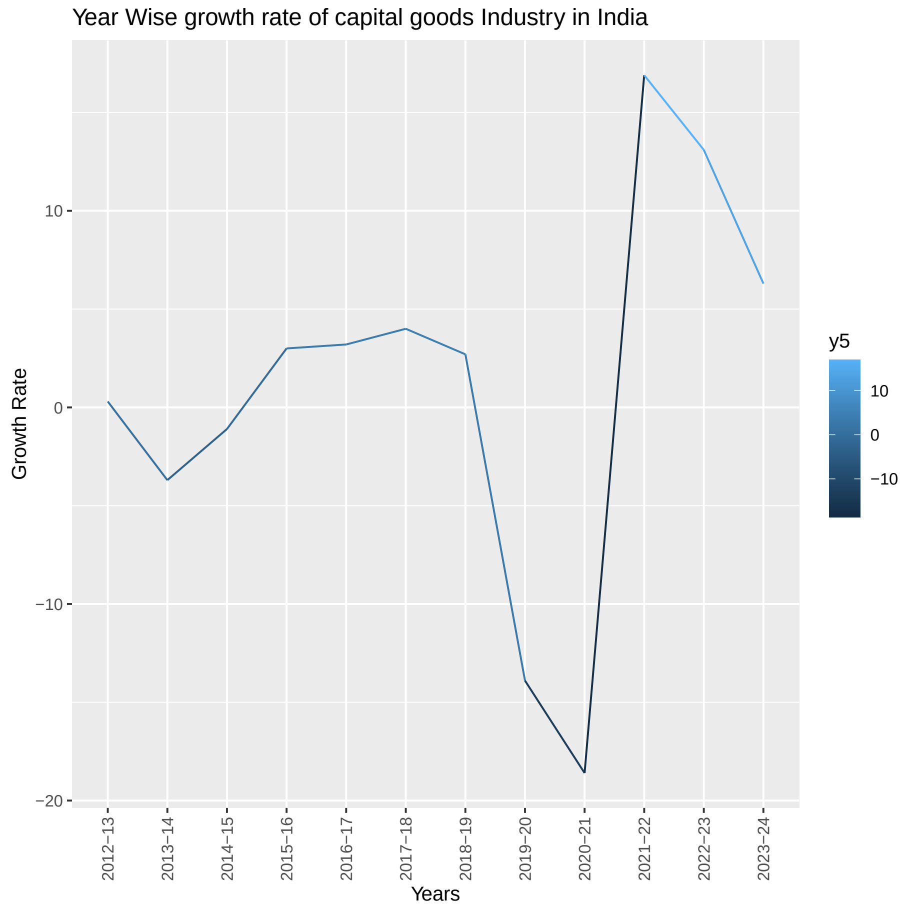
4.1 Primary goods

```
## Warning:  package 'ggplot2' was built under R version 4.4.2
```



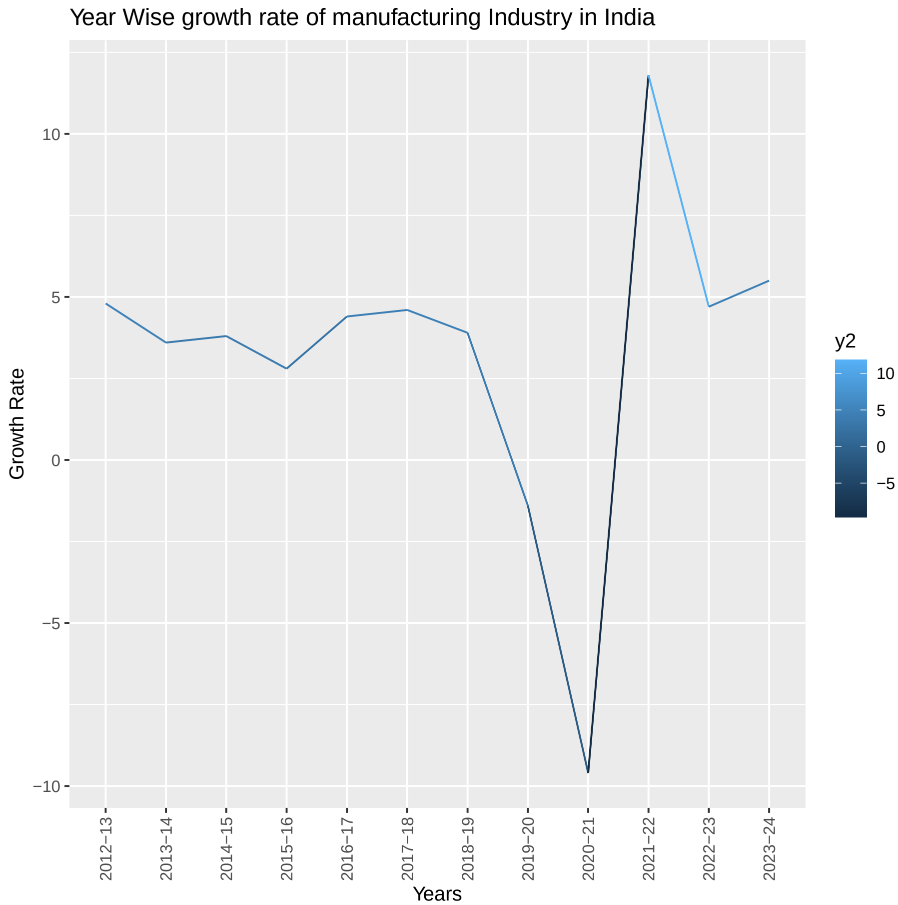
2014-2019: Noticeable fluctuations during this period influenced by numerous policy changes like demonitization, GST implementation, etc.
2020-2021: The sharp decline in this period reflects the impact of the COVID-19 pandemic. Lockdowns, supply chain disruptions, and a decrease in demand would have significantly affected the primary goods industry.
2021-2022: The strong recovery during this period could be attributed to the easing of pandemic-related restrictions, the resumption of economic activities, and pent-up demand.
2022 onwards: Moderation in the growth rate can be seen due to various factors, such as the stabilizing of the post-pandemic recovery, inflationary pressures, tightening monetary policies, and evolving global economic conditions.

4.2 Capital goods



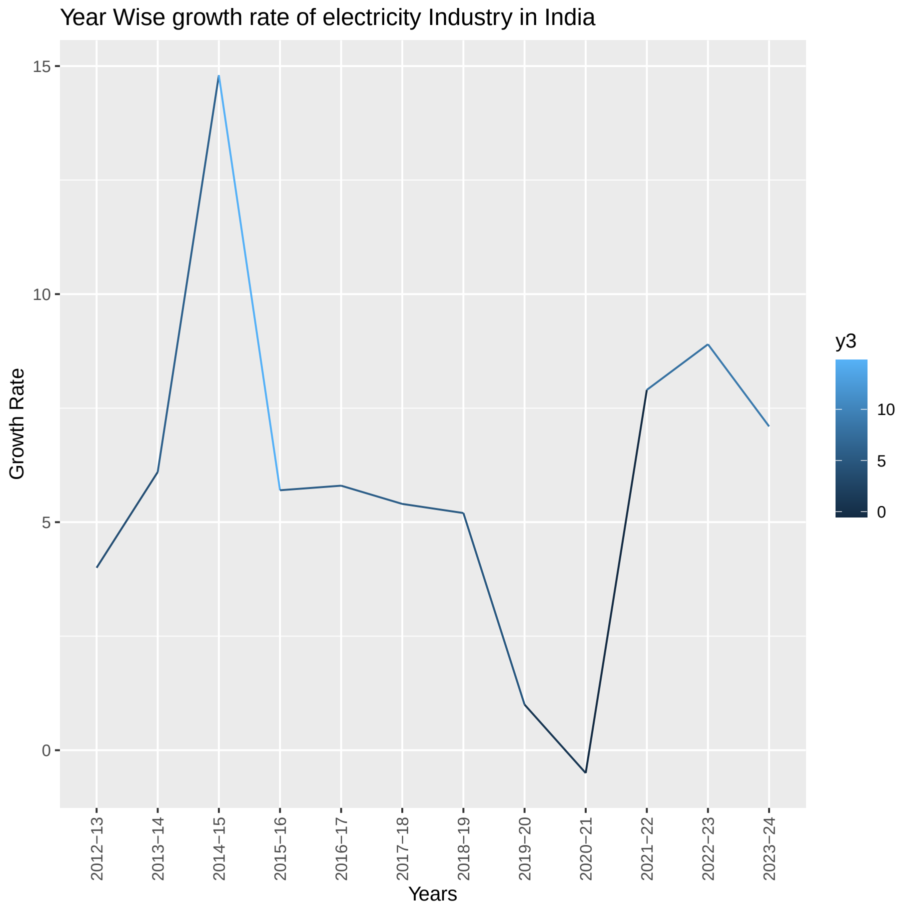
2012-2019: Noticeable fluctuations in this period could be attributed to changes in government policies related to industrial development, infrastructure projects, and investment.
Economic Slowdown
Late 2010s: Slight decline towards the end of this period may be attributed to a broader economic slowdown in India, which would have reduced demand for capital goods.
2020-2021: The sharp contraction is directly to economic slowdown caused by the COVID-19 pandemic.
2021-2022: The very high growth rate in 2021-22 likely reflects the economy's recovery post pandemic.
2022-2024: The moderation or rather slowdown in recent years could be due to factors like global supply chain issues, inflationary pressures, rising interest rates, etc.

4.3 Manufacturing sector



2012-2019: The fluctuations in the growth rate during this period could be attributed to a variety of economic policies, both domestic and international, that affected India's manufacturing sector, for example a slight slowdown can be seen in 2017-18 which can be attributed to demonetization being introduced the previous year.
2020-2021: The prominent sharp contraction is almost certainly due to the COVID-19 pandemic just like the previous plots.
2021-2022: The strong recovery likely reflects the base effect (a large increase from a very low base) as well as the resumption of economic activity following the easing of pandemic-related restrictions.
2022-2024: The moderation in the growth rate in the later years could be due to a combination of factors, including the stabilization of post pandemic recovery, global supply chain challenges, inflationary pressures, etc.

4.4 Electricity sector



2014-15: The sudden and high spike in the growth of electricity sector's production in this period is due to the inclusion of electricity produced from renewable sources
late 2010s: Slowdown can be attributed to combined effect of broader economic slowdown and different policies.
2020-2021: Sudden dip as a direct effect of economic standstill due to COVID pandemic.
2021 onwards: Fast recovery with the opening up of economy following pandemic and then subsequent stabilization.

5. Implementing the models

5.1Importing the required packages

```
library(forecast)

## Warning:  package 'forecast' was built under R version 4.4.3
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

library(vars)

## Warning:  package 'vars' was built under R version 4.4.3
## Loading required package:  MASS
## Loading required package:  strucchange
## Warning:  package 'strucchange' was built under R version 4.4.3
## Loading required package:  zoo
## Warning:  package 'zoo' was built under R version 4.4.3
##
## Attaching package:  'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
## Loading required package:  sandwich
## Warning:  package 'sandwich' was built under R version 4.4.3
## Loading required package:  urca
## Warning:  package 'urca' was built under R version 4.4.3
## Loading required package:  lmtest
## Warning:  package 'lmtest' was built under R version 4.4.3

library(tseries)

## Warning:  package 'tseries' was built under R version 4.4.3

library(Metrics)

## Warning:  package 'Metrics' was built under R version 4.4.3
##
## Attaching package:  'Metrics'
## The following object is masked from 'package:forecast':
##
##     accuracy

library(dplyr)

## Warning:  package 'dplyr' was built under R version 4.4.2
##
## Attaching package:  'dplyr'
```

```
## The following object is masked from 'package:MASS':
##
##     select
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

5.2Forecasting for primary goods sector

5.2.1Preparing the data

```
years = c(2012, 2013, 2014, 2015, 2016,
2017, 2018, 2019, 2020, 2021, 2022, 2023)
gdp=c(1.82764E+12,1.85672E+12,2.03913E+12,
2.10359E+12,2.2948E+12,2.65147E+12,2.70293E+12,
2.83561E+12,2.67485E+12,3.16727E+12,3.35347E+12,3.56755E+12)
cpi_data=c(9.478996914,10.01787847,6.665656719,
4.906973441,4.948216341,3.328173375,3.938826467,
3.729505735,6.623436776,5.131407472,6.699034141,5.649143189)

#preparing the data frame
#here we are considering the primary goods data
data = data.frame(years, y4, cpi_data, gdp)
```

5.2.2Splitting the data into train and test set, the data for first 8 years are used as train set and that of the rest 4 years are used as test set

```
# 2. Chronological split (first 8 years = train, last 4 years = test)
train_data = data[1:8, ]
test_data = data[9:12, ]

# Sort by year
train_data = train_data %>% arrange(years)
test_data = test_data %>% arrange(years)
```

5.2.3Training the models

```
# Convert the data to time series for ARIMA
ts_IIP_train = ts(train_data$y4, start = min(train_data$years),
frequency = 1)
xreg_train = as.matrix(train_data[, c("gdp", "cpi_data")])
```

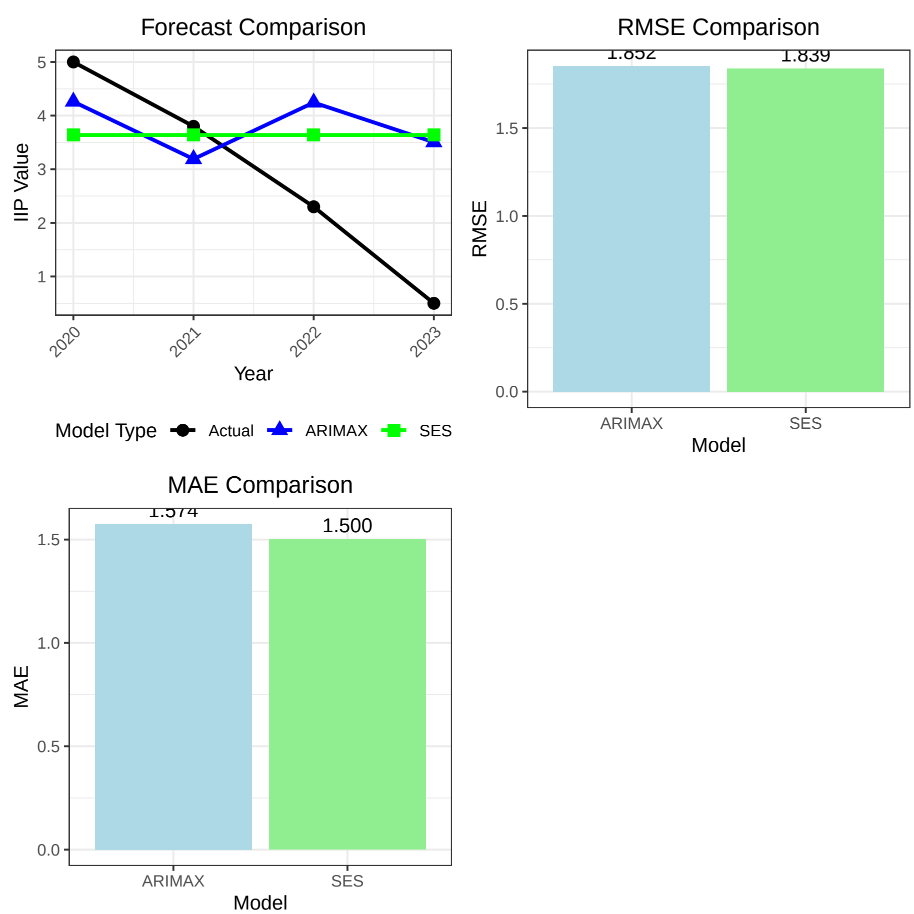
```r
xreg_test = as.matrix(test_data[, c("gdp", "cpi_data")])

# Fitting ARIMAX
arimax_model = auto.arima(ts_IIP_train, xreg = xreg_train)
arimax_forecast = forecast(arimax_model, xreg = xreg_test,
h = nrow(test_data))
arimax_pred = arimax_forecast$mean

# Simple Exponential Smoothing (SES)
ses_model = ses(ts_IIP_train, h = nrow(test_data))
ses_pred = ses_model$mean
```

5.2.4Comparing their forecasting accuracy

```r
actual = test_data$y4
arimax_rmse = rmse(actual, arimax_pred)
ses_rmse = rmse(actual, ses_pred) # Calculate RMSE for SES
arimax_mae = mae(actual, arimax_pred)
ses_mae = mae(actual, ses_pred)   # Calculate MAE for SES

accuracy_df = data.frame(
  Model = c("ARIMAX", "SES"),
  RMSE = c(arimax_rmse, ses_rmse),
  MAE = c(arimax_mae, ses_mae))


# Printing the comparison
cat("\nAccuracy Comparison (on Test Set):\n")

##
## Accuracy Comparison (on Test Set):

cat(sprintf("ARIMAX RMSE: %.3f | MAE: %.3f\n", arimax_rmse,
arimax_mae))

## ARIMAX RMSE: 1.852 | MAE: 1.574

cat(sprintf("SES    RMSE: %.3f | MAE: %.3f\n", ses_rmse,
ses_mae))

## SES    RMSE: 1.839 | MAE: 1.500
```

5.2.5Visualizing the forecasts and their comparison

```
## Warning:  package 'gridExtra' was built under R version 4.4.3
##
## Attaching package:  'gridExtra'
## The following object is masked from 'package:dplyr':
##
##     combine
## Warning:  package 'tidyr' was built under R version 4.4.2
```




```
##
## Accuracy Comparison (on Test Set):
## ARIMAX RMSE: 1.852 | MAE: 1.574
## SES    RMSE: 1.839 | MAE: 1.500
```

5.3Forecasting for capital goods sector

5.3.1Preparing the data

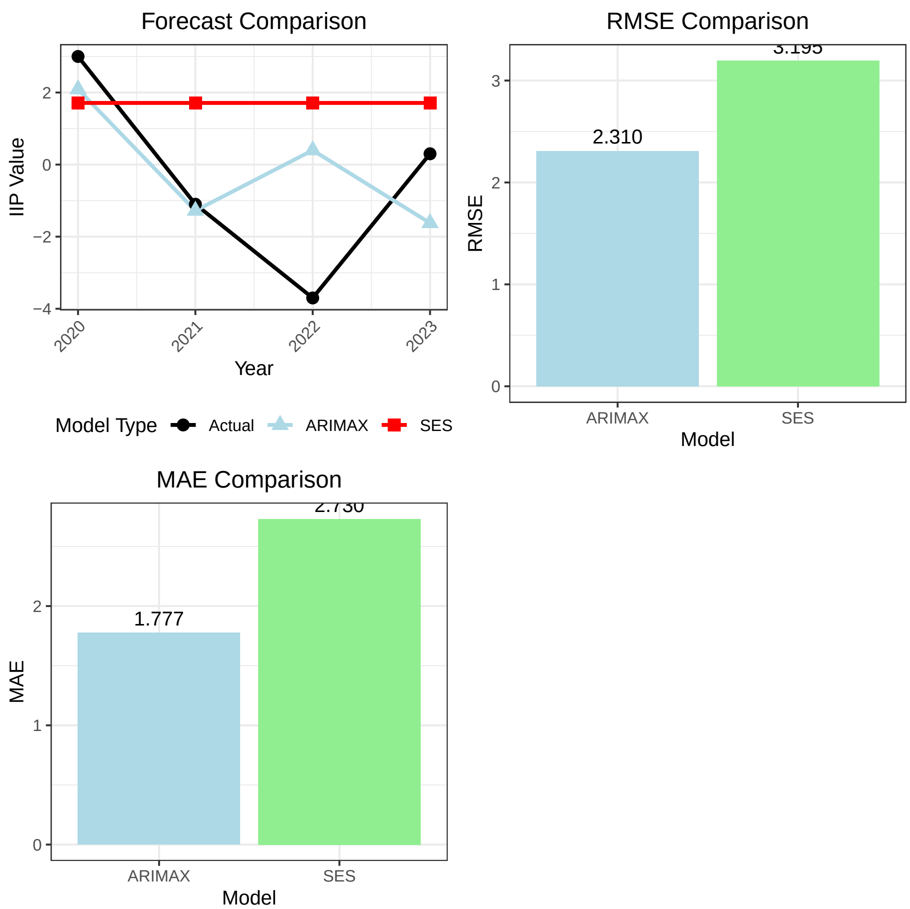
```
data2 = data.frame(years, y5, cpi_data, gdp)
# 2. Chronological split (first 8 years = train, last 4 years = test)
train_data2 = data2[1:8, ]
test_data2 = data2[9:12, ]

# Sort by year
train_data2 = train_data2 %>% arrange(years)
test_data2 = test_data2 %>% arrange(years)
```

5.3.2Training the model, with same code as before just using capital goods sector's data.

5.3.3Comparing their accuracy, again with similar code

```
##
## Accuracy Comparison (on Test Set):
## ARIMAX RMSE: 2.310 | MAE: 1.777
## SES    RMSE: 3.195 | MAE: 2.730
```

5.3.4Plotting the forecasts and the comparison for capital goods sector




```
##
## Accuracy Comparison (on Test Set):
## ARIMAX RMSE: 2.310 | MAE: 1.777
## SES    RMSE: 3.195 | MAE: 2.730
```

5.4Forecasting for manufacturing sector

5.4.1Preparing the data

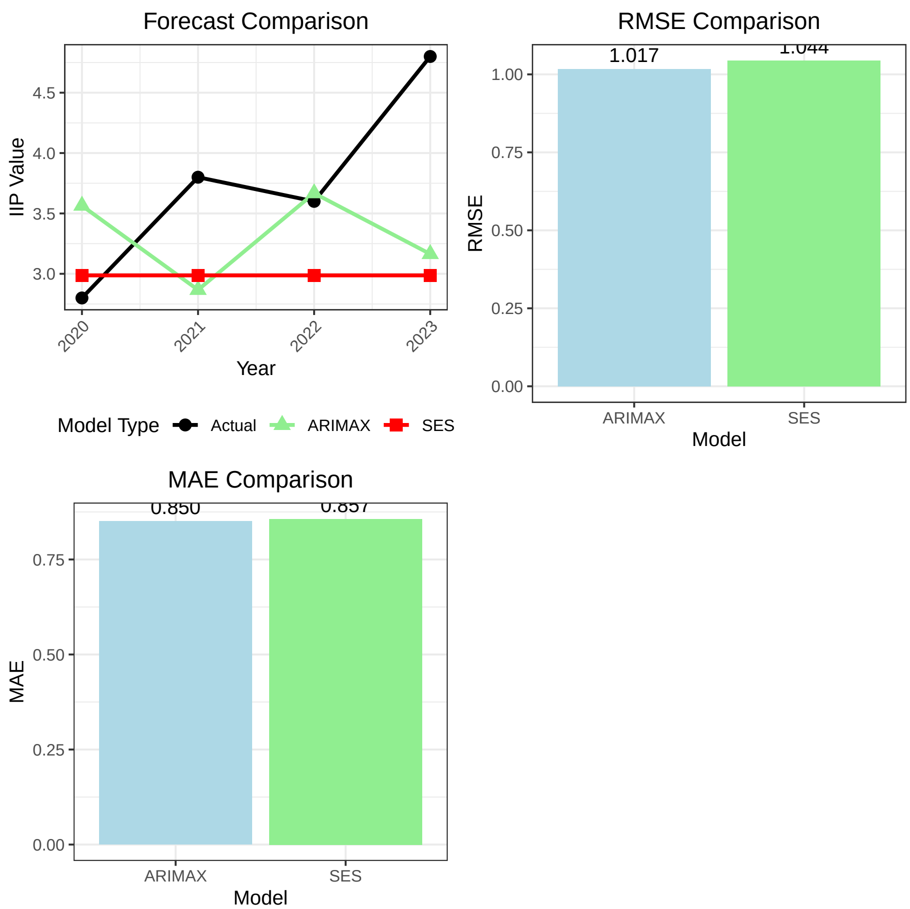
```
data3 = data.frame(years, y2, cpi_data, gdp)
# 2. Chronological split (first 8 years = train, last 4 years = test)
train_data3 = data3[1:8, ]
test_data3 = data3[9:12, ]

# Sort by year
train_data3 = train_data3 %>% arrange(years)
test_data3 = test_data3 %>% arrange(years)
```

5.4.2Training the model with same code as before, just using manufacturing sector's data.

5.4.3Comparing their accuracy again with same code as before.

```
##
## Accuracy Comparison (on Test Set):
## ARIMAX RMSE: 1.017 | MAE: 0.850
## SES    RMSE: 1.044 | MAE: 0.857
```

5.4.4Plotting the forecasts and the comparison for this sector.




```
##
## Accuracy Comparison (on Test Set):
## ARIMAX RMSE: 1.017 | MAE: 0.850
## SES    RMSE: 1.044 | MAE: 0.857
```

5.5Forecasting for electricity sector.

5.5.1Preparing the data

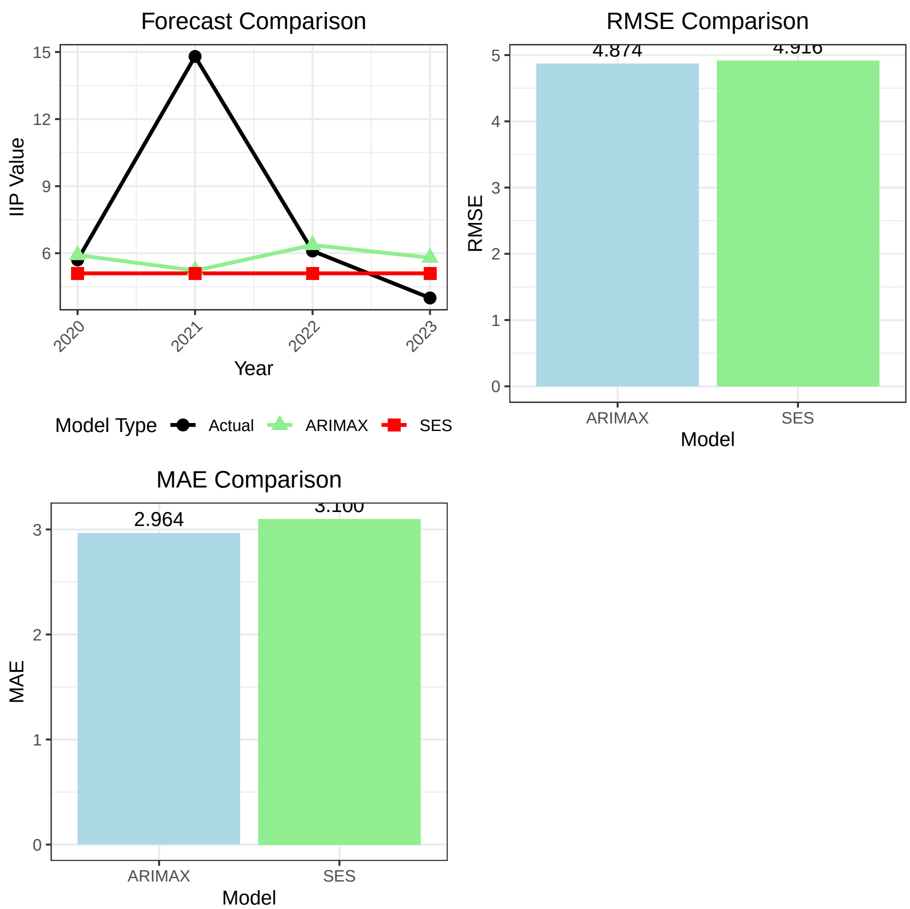
```
data4 = data.frame(years, y3, cpi_data, gdp)
# 2. Chronological split (first 8 years = train, last 4 years = test)
train_data4 = data4[1:8, ]
test_data4 = data4[9:12, ]

# Sort by year
train_data4 = train_data4 %>% arrange(years)
test_data4 = test_data4 %>% arrange(years)
```

5.5.2Training the models for this sector again with the same code just using the data for electricity sector.

5.5.3Comparing the forecasting for this sector with same code as before.

```
##
## Accuracy Comparison (on Test Set):
## ARIMAX RMSE: 4.874 | MAE: 2.964
## SES    RMSE: 4.916 | MAE: 3.100
```

5.5.4Plotting the forecasts and the comparison.




```
##
## Accuracy Comparison (on Test Set):
## ARIMAX RMSE: 4.874 | MAE: 2.964
## SES    RMSE: 4.916 | MAE: 3.100
```

6.Conclusion

- The analysis of the trends of IIP over the years suggests that in general the starting of this decade displayed fluctuations, moving towards the end of the decade every sector displayed significant dip during 2020-21 owing to COVID pandemic, which is followed by recovery growth and stabilization.

- This study aimed to analyze trends and compare the forecasting accuracy of Simple Exponential Smoothing (SES) and Autoregressive Integrated Moving Average with Exogenous factors (ARIMAX) models for sectoral Industrial Production in India, utilizing year-wise GDP and CPI as exogenous variables for ARIMAX.

- The overall findings, based on the limited dataset analyzed from each sector, suggest that both SES and ARIMAX models exhibit comparable forecasting performance for the specific sectoral IIP considered. However, better accuracy metrics (lower RMSE and MAE) that can be observed for ARIMAX in general (except for primary goods sector, that too by a slight difference), it can be considered a better alternative to the SES model, especially when dealing with short time series data, that is in this case.

Now, It is important to acknowledge the limitations of this study, primarily stemming from the limited number of data points available for model training and evaluation. Research with a more extensive dataset, encompassing longer time periods and potentially different sectoral breakdowns, would be beneficial to further validate these findings and explore the robustness of both models under varying conditions. Additionally, investigating the impact of other relevant macroeconomic indicators as exogenous variables in the ARIMAX model could potentially enhance its forecasting capabilities. Ultimately, the choice between SES and ARIMAX for forecasting sectoral IIP may depend on the specific requirements of the forecasting task, the availability of data, and the desired level of model complexity.

7.References

- https://www.geeksforgeeks.org/exponential-smoothing-for-time-series-forecasting/
- https://www.geeksforgeeks.org/what-is-an-arimax-model/
- https://esankhyiki.mospi.gov.in/
- https://data.worldbank.org/country/india
- https://archive.indianstatistics.org/datadocuments/mcsinghi_iip_asi.pdf
- https://mospi.gov.in/54-index-industrial-production