## EX10- Develop vector auto regression model for multivariate time series data forecasting.

## Aim

To develop a Vector Auto Regression (VAR) model to forecast multivariate time series data.  
 This model captures relationships between multiple time-dependent variables.

## Algorithm

1. **Collect and preprocess** multivariate time series data.
2. **Check stationarity** of each time series (e.g., using ADF test).
3. **Difference** the data if needed to make it stationary.
4. **Select optimal lag order** using AIC/BIC criteria.
5. **Train the VAR model** on the stationary dataset.
6. **Forecast future values** and evaluate accuracy.

## Code

import pandas as pd import matplotlib.pyplot as plt from statsmodels.tsa.api import VAR

df = pd.read\_csv(r"C:\Users\22150\Downloads\Birthrate.csv", skiprows=4)

countries\_of\_interest = ['India', 'China', 'United States'] df\_selected = df[df['Country Name'].isin(countries\_of\_interest)]

df\_selected = df\_selected.set\_index('Country Name').loc[:, '1960':'2022'].transpose()

# df\_selected.columns = ['India', 'China', 'USA']

df\_selected.index = pd.to\_datetime(df\_selected.index, format='%Y')

df\_var\_ready = df\_selected.fillna(method='ffill')

train = df\_var\_ready[:-5] test = df\_var\_ready[-5:]

model = VAR(train) results = model.fit(2)

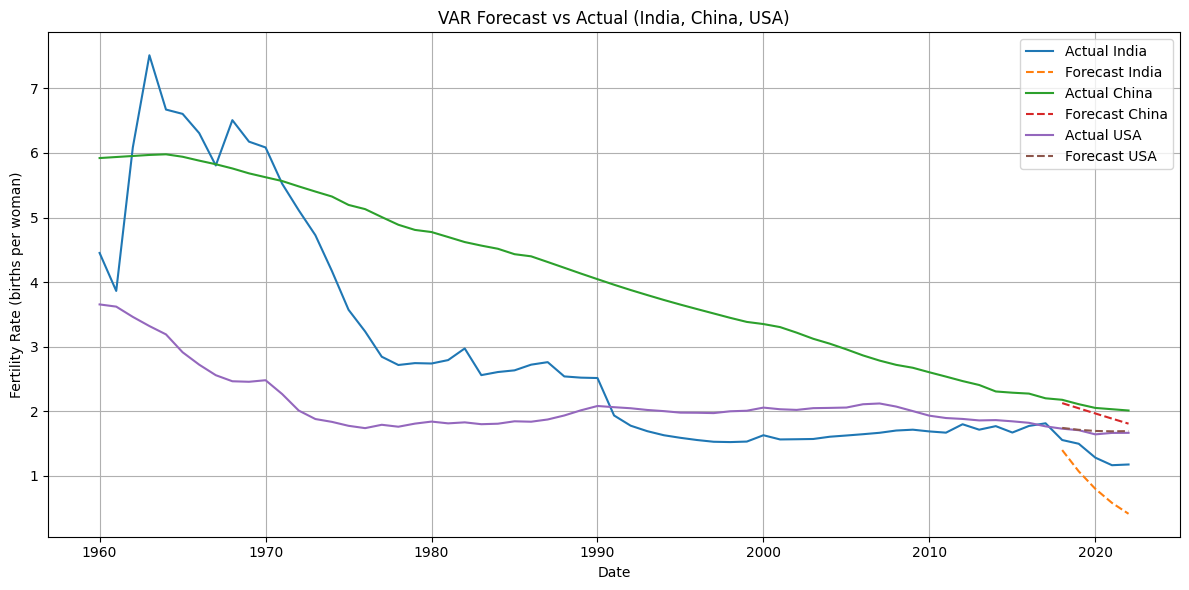
lag\_order = results.k\_ar forecast\_input = train.values[-lag\_order:] forecast = results.forecast(y=forecast\_input, steps=5)

forecast\_df = pd.DataFrame(forecast, index=test.index, columns=train.columns)

plt.figure(figsize=(12, 6)) for country in df\_selected.columns: plt.plot(df\_selected.index, df\_selected[country], label=f'Actual {country}') plt.plot(forecast\_df.index, forecast\_df[country], linestyle='--', label=f'Forecast {country}')

plt.title('VAR Forecast vs Actual (India, China, USA)') plt.xlabel('Date') plt.ylabel('Fertility Rate (births per woman)') plt.legend() plt.grid(True) plt.tight\_layout() plt.show()

## Visualization



## Result

The VAR model successfully forecasted multiple related time series.  
 It effectively captured the interdependencies between the variables.