Laboratory Work 2: Time Series Models
ARIMA, SARIMA, ARIMAX, VARIMA

Angsar Shaumen

255782@astanait.edu.kz

Group: AAI-2501M

December 29, 2025

Abstract

This laboratory work explores the modeling and forecasting of time series using ARIMA, SARIMA, ARIMAX, and VAR models. We apply these techniques to both univariate (AirPassengers) and multivariate (Macrodata) datasets. The results demonstrate the superiority of SARIMA for seasonal data and the utility of VAR for analyzing interdependent macroeconomic variables.

Contents

1 Introduction

Time series analysis is a critical component of modern data science, used in finance, economics, and resource planning. This lab focuses on:

- **ARIMA**: AutoRegressive Integrated Moving Average for non-seasonal data.

- **SARIMA**: Extensions for seasonality.

- **ARIMAX**: Incorporating exogenous variables.

- **VAR**: Vector Autoregression for multivariate systems.

2 Methodology and Data

We utilized Python with the `statsmodels` library. Two primary datasets were used:

1. **AirPassengers**: Monthly totals of international airline passengers (1949-1960). Used for ARIMA and SARIMA.

2. **US Macrodata**: Quarterly US macroeconomic data (1959-2009). Used for ARIMAX and VAR.

Stationarity was tested using the Augmented Dickey-Fuller (ADF) test. Model performance was evaluated using RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error).

3 Results

3.1 Part A: ARIMA (Univariate)

The original AirPassengers series was non-stationary (ADF $p > 0.05$). First-order differencing was applied, though the variance remained non-constant.
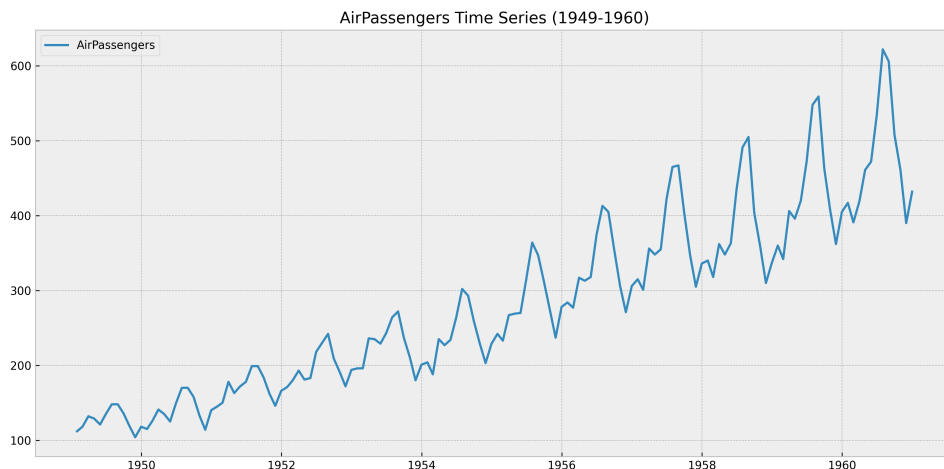

Figure 1: Original AirPassengers Series

We fitted an ARIMA(2,1,2) model.

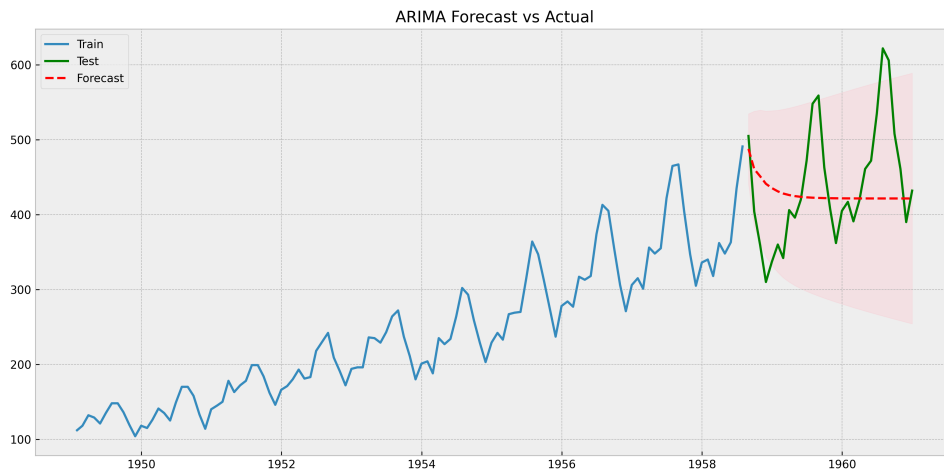- **RMSE**: 82.51

- **MAPE**: 14.22%

- **AIC**: 1057.51

Figure 2: ARIMA Forecast

3.2 Part B: SARIMA (Seasonal)

To address the clear annual seasonality, we fitted a SARIMA(1,1,1)(1,1,1,12) model on log-transformed data.

- **RMSE**: 15.02

- **MAPE**: 2.80%

- **AIC**: -368.06

The SARIMA model significantly outperformed the standard ARIMA model, capturing the seasonal peaks effectively.
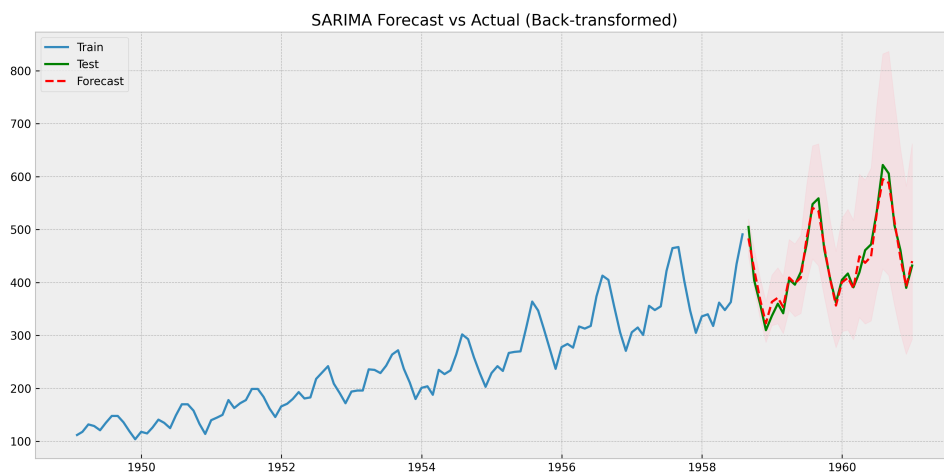

Figure 3: SARIMA Forecast

3.3 Part C: ARIMAX (Exogenous Regressors)

We modeled Real GDP Growth (`realgdp`) using Real Consumption (`realcons`) as an exogenous variable.

- **Model**: ARIMAX(1,0,1)

- **Exogenous Coef**: 1.185 ($P < 0.001$), indicating a strong positive relationship.

- **MAPE**: 143.82% (High variance in growth rates leads to high percentage errors).

3.4 Part D and E: Multivariate (VAR and VARMAX)

We modeled the interaction between `realgdp`, `realcons`, and `realinv`.

- **VAR(1)**: Selected by AIC (-27.86). The lag order of 1 suggests immediate past values drive the current economic state.

- **VARMAX(1,1)**: As an extension (Exercise 4), we added a Moving Average component.

  - **RMSE (GDP)**: 0.0091
  - While VARMAX offers theoretical flexibility, the computational cost and identification difficulties (many insignificant p-values) often make standard VAR preferred for macro data.

Granger causality tests (implied by the significant coefficients in the VAR summary) suggest strong interdependencies.
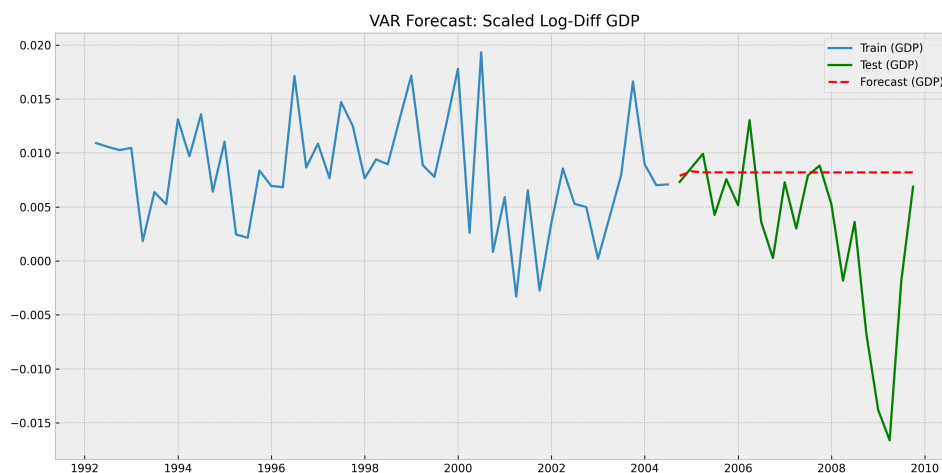
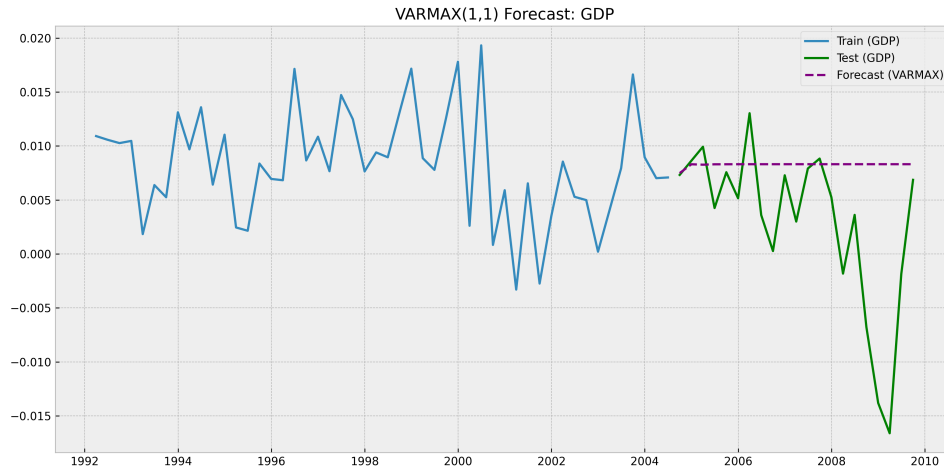

Figure 4: VAR Forecast for Real GDP

Figure 5: VARMAX(1,1) Forecast for Real GDP

4 Control Questions

1. **Role of d in ARIMA**: The parameter d represents the degree of differencing required to make the time series stationary. Over-differencing increases the variance of the series and introduces unnecessary MA correlations.

2. **SARIMA Parameters**: (P, D, Q, s) refer to the Seasonal AR, Seasonal Difference, Seasonal MA, and Seasonality period (e.g., 12 for months). They function similarly to non-seasonal parameters but operate on lag s.

3. **Exogenous Variables**: These are external inputs (X) that help predict the target (Y), but are not influenced by Y. For forecasting, future values of X must be known or forecastable.

4. **VAR vs VARMA**: VAR uses only lagged values of the variables themselves. VARMA adds Moving Average (error) terms. VECM is appropriate when variables are non-stationary but cointegrated.

5. **Diagnostics**: We check residuals for white noise (Ljung-Box test), normality (Jarque-Bera), and heteroskedasticity.

6. **Reverting Forecasts**: If differenced $d = 1$, we calculate the cumulative sum of the forecast and add the last observed value ($Y_t = Y_{t-1} + \Delta Y_t$).

5 Conclusion

This lab demonstrated that accounting for seasonality (SARIMA) drastically improves forecasting for periodic data like air travel. For economic indicators, adding exogenous variables (ARIMAX) or modeling variables jointly (VAR) provides richer insights into structural relationships.

References

[1] Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day.

[2] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer.

[3] Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference.*