

Homework 2: ARIMA

Instructions: You may discuss this assignment with other students in the class and use AI or other on-line support, but you must submit your own answers to the questions below. Indicate who you consulted. Include an honor pledge with your submission. Submit on-line in a jupyter notebook, a link to your jupyter notebook, or a pdf of your jupyter notebook. This homework is worth 50 points and the point totals for each question are shown in parentheses.

For this assignment you obtain a real time series data set that is an area of interest to you and for which forecasting is a goal. Provide one of the following:

- A link to the data set,
- A reference to the data set that describes how it can be obtained, or
- Include the data set with your submission.

You will apply ARIMA, SARIMA, or SARIMAX to the data set to obtain and evaluate forecasts for your data set.

1. (5) Show a data description.
 - (a) Provide a summary of the data set to include descriptive statistics.
 - (b) Plot the time series.
 - (c) (optional) Provide any additional plots that may show important characteristics of your data.
2. (45) Apply the Box-Jenkins methodology using ARIMA, SARIMA, or SARIMAX depending on your data to obtain and evaluate forecasts.
 - (a) (9) **Identification:** Perform all the necessary steps for identification and include automatic and manual identification. Summarize the results from your identification process.
 - (b) (9) **Estimation:** Perform all necessary steps for identification and summarize the results of your estimation process.
 - (c) (9) **Diagnostic Checking:** Perform all necessary steps for diagnostic checking and summarize the results of your diagnostic checking process.
 - (d) (9) **Forecasting:** Compare the performance of the forecasts of your selected models using the appropriate methods on a test data set. Show graphics of the forecast performance to include confidence intervals. Describe your results and conclusions about the usefulness of the models you evaluated to forecast in the application domain of the data set.