Case Study #3: Forecasting with AR and ARIMA Models

Consider the quarterly data on Walmart revenues (in $million) from the first quarter of 2005 through the fourth quarter of 2023 (*673_case2.csv*). The goal is to forecast Walmart's quarterly revenue in the fourth quarters (Q1-Q4) of 2024-2025.

As you did in *case study #2*, start this case with the following:

- Create time series data set in R using the *ts()* function.
- Develop data partition with the validation partition of 16 periods and the rest for the training partition.

Questions

1. **Identify time series predictability.**
   a. Using the *AR(1)* model for the historical data, Provide and explain the *AR(1)* model summary in your report. Explain if the Walmart revenue is predictable.
   b. Using the first differencing (lag1) of the historical data and *Acf()* function, provide in the report the autocorrelation plot of the first differencing (lag1) with the maximum of 8 lags and explain if Walmart revenue is predictable.

2. **Apply the two-level forecast with regression model and AR model for residuals.**
   a. For the training data set, use the *tslm()* function to develop a *regression model with linear trend and seasonality*. Forecast Walmart's revenue in the validation period with the *forecast()* function (use the associated R code from case #2). No explanation is required in your report.
   b. Identify the regression model's residuals for the training period and use the *Acf()* function with a maximum of 8 lags to identify autocorrelation for these residuals. Provide the autocorrelation plot in your report and explain if it would be a good idea to add to your forecast an AR model for residuals.
   c. Develop an *AR(1)* model for the regression residuals, present and explain the model and its equation in your report. Use the *Acf()* function for the residuals of the *AR(1)* model (residuals of residuals), present the autocorrelation chart, and explain it in your report.
   d. Create a two-level forecasting model (regression model with *linear trend and seasonality + AR(1)* model for residuals) for the validation period. Show in your report a table with the validation data, regression forecast for the validation data, *AR(1)* forecast for the validation data, and combined forecast for the validation period.
   e. Develop a two-level forecast (regression model with *linear trend and seasonality* and *AR(1)* model for residuals) for the entire data set. Provide in your report the autocorrelation chart for the *AR(1)* model's residuals and explain it. Also, provide a data table with the models' forecasts for Walmart revenue in Q1-Q4 of 2024-2025 (regression model, *AR(1)* for residuals, and two-level combined forecast).

3. **Use ARIMA Model and Compare Various Methods.**
   a. Use *Arima()* function to fit *ARIMA(1,1,1)(1,1,1)* model for the *training data set*. Insert in your report the summary of this ARIMA model, present and briefly explain the *ARIMA* model and its equation in your report. Using this model, forecast revenue for the *validation period* and present it in your report.
   b. Use the *auto.arima()* function to develop an *ARIMA* model using the *training data set*. Insert in your report the summary of this *ARIMA* model, present and explain the *ARIMA* model and its equation in your report. Use this model to forecast revenue in the *validation period* and present this forecast in your report.
   c. Apply the *accuracy()* function to compare performance measures of the two *ARIMA* models in *3a* and *3b*. Present the accuracy measures in your report, compare them and identify, using MAPE and RMSE, the best *ARIMA* model to apply.
   d. Use two *ARIMA* models from *3a* and *3b* for the entire data set. Present models' summaries in your report. Use these *ARIMA* models to forecast Walmart revenue in Q1-Q4 of 2024-2025 and present these forecasts in your report.
   e. Apply the *accuracy()* function to compare performance measures of the following forecasting models for the *entire data set*: (1) regression model with *linear trend and seasonality*; (2) *two-level* model (with *AR(1)* model for residuals); (3) *ARIMA(1,1,1)(1,1,1)* model; (4) *auto ARIMA* model; and (5) *seasonal naïve* forecast for the entire data set. Present the accuracy measures in your report, compare them, and identify, using *MAPE* and *RMSE*, the best model to use for forecasting Walmart's revenue in Q1-Q4 of 2024-2025.