Case Study #2: Forecasting Walmart's Revenue

The data set for case study #2 represents quarterly revenues (in $million) in Walmart from the first quarter of 2005 through the fourth quarter of 2023 (*673_case2.csv*). This quarterly data is collected from *www.macrotrends.net/stocks/charts/WMT/walmart/revenue*. The goal is to forecast Walmart's quarterly revenue in the four quarters (Q1-Q4) of 2024-2025.

Questions

1. **Plot the data and visualize time series components.**
    a. Create time series data set in R using the *ts()* function.
    b. Apply the *plot()* function to create a data plot with the historical data, provide it in your report, and explain what time series components can be visualized in this plot.
2. **Apply five regression models using data partition.**

    Consider the following 5 regression-based models:
        i. Regression model with linear trend
        ii. Regression mode with quadratic trend
        iii. Regression model with seasonality
        iv. Regression model with linear trend and seasonality
        v. Regression model with quadratic trend and seasonality.

    a. Develop data partition with the validation partition of 16 periods and the rest for the training partition.
    b. Use the *tslm()* function for the training partition to develop each of the 5 regression models from the above list. Apply the *summary()* function to identify the model structure and parameters for each regression model, show them in your report, and also present the respective model equation and define its predictors. Briefly explain if the model is a good fit, statistically significant, and thus may be applied for forecasting. Use each model to forecast revenues for the validation period using the *forecast()* function, and present this forecast in your report.
    c. Apply the *accuracy()* function to compare performance measure of the 5 forecasts you developed in *2b*. Present the accuracy measures in your report, compare them, and, using MAPE and RMSE, identify the *three most accurate regression models for forecasting*. If the historical data contains trend and seasonality, then in selecting the regression models *give a preference to those models that include trend and seasonality*.

3. **Employ the entire data set to make time series forecast.**
    a. Apply the two most accurate regression models identified in *2c* to make the forecast in the four quarters (Q1-Q4) of 2024-2025. For that, use the entire data set to develop the regression model using the *tslm()* function. Apply the *summary()* function to identify the model structure and parameters, show them in your report, and also present the respective model equation and define its predictors. Briefly explain if the model is a good fit, statistically significant, and thus may be applied for forecasting. Use each model to forecast Walmart's revenue in Q1-Q4 of 2024-2025 using the *forecast()* function, and present this forecast in your report.

b. Apply the *accuracy()* function to compare the performance measures of the regression models developed in *3a* with those for naïve and seasonal naïve forecasts. Present the accuracy measures in your report, compare them, and identify, using MAPE and RMSE, which forecast is most accurate to forecast Walmart's quarterly revenue in Q1-Q4 of 2024-2025.