Case Study #1: Sales Forecasting

In this case, consider the data on worldwide monthly sales from a large chain of grocery stores (*673_case1.csv*). The monthly sales are given for the period of 2016-2025 and measured in millions of dollars. The goal is to identify the best forecasting model to predict monthly sales in the 12 months of 2026.

For answering the case questions below using R, you can apply some of the R code for the smoothing methods presented in Module 2. You may also use (but it is optional) an AI tool(s) to revise or edit your R code. However, the proper citation of AI use is required (see the Artificial Intelligence (AI) Policy in the Course Syllabus).

Questions

1. **Identify the time series components and plot the data.**
    a. Create time series data set *sales.ts* in R using the *ts()* function.
    b. Employ the *plot()* function to create a data plot of historical data, provide it in your report, and explain what data patterns can be visualized in this plot.
    c. Apply the *Acf()* function to identify possible time series components. Provide in the report the autocorrelation chart and explain the time series components existing in the historical data.
2. **Use trailing MA for forecasting time series.**
    a. Develop data partition with the validation partition of 36 monthly periods (3 years) and training partition of 84 monthly periods (7 years). Provide the data partition's R code in your report.
    b. Use the *rollmean()* function to develop 3 trailing MAs with the window width of 3, 5, and 12 for the training partition. Present the R code for these MAs in your report.
    c. Use the *forecast()* function to create a trailing MA forecast for each window width from question *2b* in the validation period, and present one of them, e.g., with window width of 3, in your report.
    d. Apply the *accuracy()* function to compare accuracy of the 3 trailing MA forecasts in the validation period. Present the accuracy measures in your report, compare MAPE and RMSE of these forecasts, and identify the best trailing MA forecast.
3. **Apply the two-level forecast with regression and trailing MA for residuals.**
    a. Develop using the *tslm()* function a regression model with linear trend and seasonality. Present the model summary in your report. Present and briefly explain the model equation in your report. Using this model, forecast monthly sales in the validation period with the *forecast()* function. Present the forecast in your report.
    b. Identify regression residuals in the training period, apply a trailing MA (window width of 3) for these residuals using the *rollmean()* function, and identify trailing MA forecast of these residuals in the validation period (use the *forecast()* function).
    c. Develop a two-level forecast for the validation period by combining the regression forecast and trailing MA forecast for residuals. Present in your report a table that contains

validation data, regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the validation period. Apply the *accuracy()* function to compare accuracy of the regression model with linear trend and seasonality and the two-level (combined) model with regression and trailing MA for residuals. Present the accuracy measures in your report, compare the MAPE and RMSE of these forecasts, and identify the best forecasting model for the validation period.

d. For the entire data set, identify the regression model with linear trend and seasonality and trailing MA with the window width of 3 for the regression residuals. Use these models to forecast 12 future months of 2026 and develop a two-level forecast for the 12 future months as a combination of the specified forecasts. Present in your report a table that contains regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the 12 future months of 2026.
e. Develop a seasonal naïve forecast for the entire historical data set and apply the *accuracy()* function to compare accuracy of the three forecasting models: seasonal naïve forecast, regression model with trend and seasonality, and two-level (combined) model with regression and trailing MA for residuals. Present the accuracy measures in your report, compare the MAPE and RMSE of these forecasts, and identify the best forecasting model for forecasting monthly sales in 2026.

4. Use advanced exponential smoothing methods.

a. For the training partition (from question *2a*), use the *ets()* function to develop a Holt-Winter's (HW) model with automated selection of error, trend, and seasonality options, and automated selection of smoothing parameters for the training partition. Present the model summary (output) and explain the model in your report. Use the model to forecast monthly sales for the validation period using the *forecast()* function, and present this forecast in your report.
b. To make a forecast for the 12 future months of 2026, use the entire data set (no partitioning) to develop the HW model using the *ets() function* for the model with the automated selection of error, trend, and seasonality options, and automated selection of smoothing parameters. Present the model summary (output) and explain this model in your report. Use the model to forecast monthly sales in the 12 months of 2026 using the *forecast()* function, and present the forecast in your report.
c. Apply the *accuracy()* function to compare the two models: seasonal naïve forecast (applied in question *3e*) and the HW model developed in question *4b*. Present the accuracy measures in your report, compare MAPE and RMSE of these forecasts, and identify the best forecasting model.
d. Compare the best forecasts identified in questions *3e* and *4c*. Explain what your final choice of the forecasting model in this case will be.