

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 a period of 2015-2019 and measured in millions of dollars. The goal is to identify the best forecasting model to predict monthly sales in 12 months of 2020.

Questions

1. Identify time series components and plot the data.

- Create time series data set in R using the `ts()` function.
- Employ the `plot()` function to create a data plot with the historical data, provide it in your report, and explain what data patterns can be visualized in this plot.
- 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.

- Use the `rollmean()` function to develop three trailing MAs (apply the entire data set with no partitioning) for the window width of 2, 6, and 12, respectively. Present the R code for these MAs in your report.
- Use the `forecast()` function to create a trailing MA forecast for each window width in 12 months of 2020, and present these forecasts in your report.
- Develop a seasonal naïve forecast for the entire historical data set, and apply the `accuracy()` function to compare accuracy of the four models: seasonal naïve forecast and trailing MAs with window width of 2, 6, and 12, respectively. Present the accuracy measures in your report, compare MAPE and RMSE of these forecasts, and identify the best forecasting model.

3. Apply the two-level forecast with regression and trailing MA for residuals.

- To de-trend and de-seasonalize the historical data for the entire data set, develop using the `tslm()` function a regression model with linear trend and seasonality and forecast sales in 2020 with the `forecast()` function. Present and briefly explain the model in your report.
- Identify regression residuals, apply a trailing MA (window width of 2) for these residuals using the `rollmean()` function, and forecast worldwide monthly sales in 12 months of 2020 (use the `forecast()` function). Combine the regression and trailing MA residuals' forecast for 2020, and present in your report a table that contains regression forecast, trailing MA forecast for residuals, and total (combined) forecast in 2020.
- Apply the `accuracy()` function to compare accuracy of the three forecasting models: seasonal naïve forecast (applied in question 2c), 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 MAPE and RMSE of these forecasts, and identify the best forecasting model.

4. Use advanced exponential smoothing methods.

- a. Develop data partition with the validation partition of 12 historical periods and training partition of the rest of the historical periods.
- b. For the training partition, use the *ets()* function to develop a Holt-Winter's model with multiplicative error, multiplicative trend, and multiplicative seasonality options, and automated selection of smoothing parameters for the training partition. Present and explain the model in your report. Use the model to forecast worldwide sales for the validation period using the *forecast()* function.
- c. To make a forecast for the 12 months of 2020, use the entire data set (no partitioning) to develop the Holt-Winter's model using the *ets()* function with the automated selection of error, trend, and seasonality options, and automated selection of smoothing parameters. Present and explain the model in your report. Use the model to forecast worldwide sales for the 12 months of 2020 using the *forecast()* function.
- d. Apply the *accuracy()* function to compare the two forecasting models: seasonal naïve forecast (applied in question 2c) and Holt-Winter's model developed in question 4c. Present the accuracy measures in your report, compare MAPE and RMSE of these forecasts, and identify the best forecasting model.
- e. Compare the best forecasts identified in questions 3c and 4c. Explain what your final choice of the forecasting model in this case will be.