## Case Study #1: Sales Forecasting

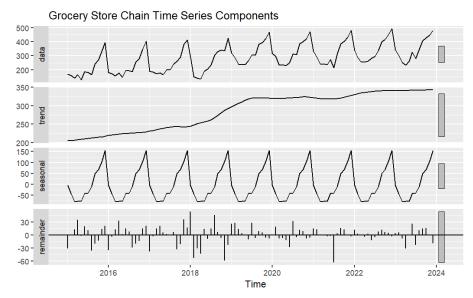
1. Identify time series components and plot the data.

## a. Create time series data set sales.ts in R using the ts() function.

We have used sales.ts to create a dataset from January 2015 until December 2023 with a frequency of 12 for monthly data.

We also used the autoplot() function to visualize the time series data:

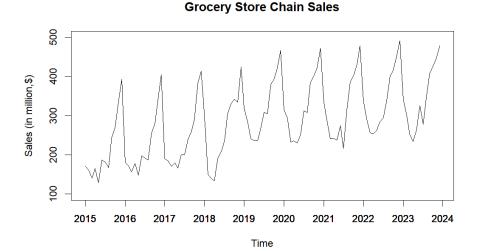
```
> #1(a)Creating time series data set sales.ts in R using the ts() function.
> sales.ts <-ts(Case1.data$Sales,
               start = c(2015, 1), end = c(2023, 12), freq = 12)
> sales.ts
            Feb
       Jan
                  Mar
                        Apr
                              May
                                     Jun
                                           Jul
                                                 Aug
                                                       Sep
                                                             Oct
2015 169.9 160.2 140.1 164.8 128.6 187.2 181.7 167.1 243.2 269.1 333.9 392.5
2016 180.5 172.4 156.1 178.3 147.3 197.1 192.6 186.7 255.2 279.8 347.6 404.4
2017 189.5 184.8 170.4 179.4 165.6 199.4 200.4 240.3 259.7 289.5 382.7 413.8
2018 294.3 149.5 140.1 133.3 191.4 207.0 236.0 303.6 329.9 341.6 334.5 424.5
2019 320.1 285.4 240.7 236.9 236.3 265.4 307.8 304.7 380.2 394.5 420.2 467.0
2020 316.3 294.1 232.2 235.1 229.9 250.8 312.4 308.1 384.7 400.3 421.9 472.0
2021 333.5 289.0 240.6 242.1 237.3 274.2 215.5 314.1 385.9 402.7 430.6 478.8
2022 338.8 290.3 255.7 253.9 262.1 284.0 294.2 338.6 401.4 414.4 450.1 491.5
2023 340.8 302.2 253.1 233.2 262.3 325.6 278.2 343.8 406.5 426.2 447.4 478.9
```



b. Employ the plot() function to create a data plot of the historical data, provide it in your report, and explain what data patterns can be visualized in this plot.

We used plot() function to create a data plot of the historical data. We observe an upward linear trend with mostly additive seasonality, except the sales were at the lowest in the initial months of 2018.

It is also observed that the sales are typically lower at the beginning of each year, and gradually increase through the months, with the highest during the months of October-December. This can be potentially due to the Holidays season and sales/discounts.

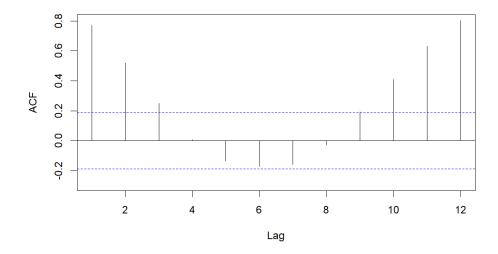


c. Apply the Acf() function to identify possible time series components. Provide in the report the autocorrelation chart and explain time series components existing in the historical data.

We observe positive autocorrelation for lags 1, 2, 3, 9, 10, 11, and 12. The positive correlation for lags 1 and 12 is the highest; it is negative for lags 5, 6, 7, and 8, and zero for 4.

We can assume that there's an upward trend, a level component, as well as seasonality in the given dataset.

## **Autocorrelation for Grocery Store Chain Sales**



```
main = Au
 > data.frame(Lag, ACF)
   Lag ACF
   0 1.000
    1 0.770
    2 0.518
    3 0.248
 5
    4 0.005
 6
    5 -0.136
    6 -0.170
 8
    7 -0.158
 9
   8 -0.028
 10 9 0.192
 11 10 0.410
 12 11 0.630
 13 12 0.801
| > |
```

- 2. Use trailing MA for forecasting time series.
- a. Develop data partition with the validation partition of 24 monthly periods (2 years) and training partition of 84 monthly periods (7 years). Provide the data partition's R code in your report.

We have used the following code to develop the data partition:

b. Use the rollmean() function to develop 3 trailing MAs with the window width of 4, 6, and 12 for the training partition. Present the R code for these MAs in your report.

We have used the following code to develop the trailing MAs:

```
\#2(b) \# Using rollmean() to create trailing moving average with window widths \# of k=4, 6, and 12. \# ma.trailing_4 <- rollmean(train.ts, k=4, align = "right") \# ma.trailing_6 <- rollmean(train.ts, k=6, align = "right") \# ma.trailing_12 <- rollmean(train.ts, k=12, align = "right")
```

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 4, in your report.

We created the trailing MA forecast for k = 4, 6, and 12. The same for the window width of 4 is as follows:

d. Apply the accuracy() function to compare accuracy of the three 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.

The accuracy measures have been given below. Comparing the MAPE and RMSE of the 3 forecasts with windows width of 4, 6, and 12, we notice that the most accurate forecast comes out to be the trailing MA forecast with window width of 4, with the lowest MAPE and RMSE of 16.198% and 59.89, respectively.

```
> round(accuracy(ma.trail_4.pred$mean, valid.ts), 3)

ME RMSE MAE MPE MAPE ACF1 Theil's U

Test set -6.869 59.89 52.245 -4.873 16.198 0.714 1.273

> round(accuracy(ma.trail_6.pred$mean, valid.ts), 3)

ME RMSE MAE MPE MAPE ACF1 Theil's U

Test set 7.812 77.742 71.217 -2.136 21.529 0.779 1.728

> round(accuracy(ma.trail_12.pred$mean, valid.ts), 3)

ME RMSE MAE MPE MAPE ACF1 Theil's U

Test set 20.257 80.55 66.439 1.116 18.847 0.737 1.609
```

- 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.

We have developed a regression model with linear trend and seasonality. The regression model includes a trend as well as 11 seasonal variables from February until December.

```
Call:
tslm(formula = train.ts ~ trend + season)
Residuals:
           1Q Median
   Min
                         30
-84.176 -13.857 0.967 17.228 51.108
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
0.1276 14.036 < 2e-16 ***
trend
           1.7910
          -40.1767 15.0027 -2.678 0.00920 **
-72.7106 15.0044 -4.846 7.16e-06 ***
season2
season3
season4
         -67.4016 15.0071 -4.491 2.68e-05 ***
          -73.9783 15.0109 -4.928 5.24e-06 ***
season5
season6
          -40.8121
                     15.0157 -2.718 0.00825 **
         -33.2746 15.0217 -2.215 0.02996 *
season7
season8
           -9.6084 15.0288 -0.639 0.52466
           47.7720
                     15.0369
                              3.177 0.00220 **
season9
           65.7953 15.0461
                               4.373 4.12e-05 ***
season10
season11 105.9900 15.0564 7.040 9.94e-10 ***
season12 158.7133 15.0677 10.533 3.71e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
Residual standard error: 28.07 on 71 degrees of freedom
Multiple R-squared: 0.9216,
                            Adjusted R-squared: 0.9083
F-statistic: 69.55 on 12 and 71 DF, p-value: < 2.2e-16
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 28.07 on 71 degrees of freedom
Multiple R-squared: 0.9216,
                                     Adjusted R-squared: 0.9083
F-statistic: 69.55 on 12 and 71 DF, p-value: < 2.2e-16
```

We have also forecasted monthly sales in the validation period with the forecast() function:

```
> trend.seas.pred
        Point Forecast
                           Lo 0
Jan 2022
           343.6964 343.6964 343.6964
Feb 2022
              305.3107 305.3107 305.3107
Mar 2022
              274.5679 274.5679 274.5679
Apr 2022
             281.6679 281.6679 281.6679
May 2022
              276.8821 276.8821 276.8821
Jun 2022
             311.8393 311.8393 311.8393
Jul 2022
             321.1679 321.1679 321.1679
Aug 2022
              346.6250 346.6250 346.6250
             405.7964 405.7964 405.7964
Sep 2022
             425.6107 425.6107 425.6107
Oct 2022
Nov 2022
             467.5964 467.5964 467.5964
Dec 2022
             522.1107 522.1107 522.1107
Jan 2023
              365.1884 365.1884 365.1884
              326.8027 326.8027 326.8027
Feb 2023
Mar 2023
             296.0598 296.0598 296.0598
Apr 2023
              303.1598 303.1598 303.1598
May 2023
             298.3741 298.3741 298.3741
              333.3313 333.3313 333.3313
Jun 2023
Jul 2023
              342.6598 342.6598 342.6598
Aug 2023
             368.1170 368.1170 368.1170
Sep 2023
             427.2884 427.2884 427.2884
             447.1027 447.1027 447.1027
Oct 2023
Nov 2023
             489.0884 489.0884 489.0884
Dec 2023
              543.6027 543.6027 543.6027
```

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 forecasts of these residuals in the validation period (use the forecast() function). Provide the trailing MA forecast for residuals in the validation period in your report.

The trailing MA forecast for residuals is as follows:

```
> #Residuals forecast for validation period.
> ma.trail.res.pred <- forecast(ma.trail.res, h = nValid, level = 0)</pre>
> ma.trail.res.pred
        Point Forecast
                            100
                                       Hi O
Jan 2022 -12.91321 -12.91321 -12.91321
Feb 2022
             -12.91321 -12.91321 -12.91321
Mar 2022
             -12.91321 -12.91321 -12.91321
Apr 2022
            -12.91321 -12.91321 -12.91321
May 2022
            -12.91321 -12.91321 -12.91321
Jun 2022
             -12.91321 -12.91321 -12.91321
Jul 2022
             -12.91321 -12.91321 -12.91321
            -12.91321 -12.91321 -12.91321
Aug 2022
            -12.91321 -12.91321 -12.91321
-12.91321 -12.91321 -12.91321
Sep 2022
Oct 2022
            -12.91321 -12.91321 -12.91321
Nov 2022
Dec 2022
             -12.91321 -12.91321 -12.91321
Jan 2023
             -12.91321 -12.91321 -12.91321
Feb 2023
             -12.91321 -12.91321 -12.91321
Mar 2023
            -12.91321 -12.91321 -12.91321
            -12.91321 -12.91321 -12.91321
-12.91321 -12.91321 -12.91321
Apr 2023
May 2023
Jun 2023
            -12.91321 -12.91321 -12.91321
Jul 2023
             -12.91321 -12.91321 -12.91321
Aug 2023
             -12.91321 -12.91321 -12.91321
Sep 2023
            -12.91321 -12.91321 -12.91321
            -12.91321 -12.91321 -12.91321
Oct 2023
             -12.91321 -12.91321 -12.91321
Nov 2023
            -12.91321 -12.91321 -12.91321
Dec 2023
```

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 the 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 for the validation period.

We have represented validation data, regression forecast, MA forecast for residuals, and two-level combined forecast in the validation period. We also calculated the accuracy measures, from which we can derive that out of the two forecasting models in the validation period, the two-level model is more accurate with MAPE and RMSE as 5.313% and 22.907, respectively.

```
> #Developing two-level forecast for validation period
> #by combining regression forecast and trailing MA forecast for residuals
> fst.2level <- trend.seas.pred$mean + ma.trail.res.pred$mean</pre>
> fst.2level
                      Feb
                                 Mar
                                                                           Jul
                                                                                      Aug
            Jan
                                            Apr
                                                      May
                                                                                                Sep
2022 330.7832 292.3975 261.6546 268.7546 263.9689 298.9261 308.2546 333.7118 392.8832
2023 352.2752 313.8895 283.1466 290.2466 285.4609 320.4180 329.7466 355.2038 414.3752
            0ct
                    Nov
                                 Dec
2022 412.6975 454.6832 509.1975
2023 434.1895 476.1752 530.6895
/ variu.ui
   Sales Regression.Fst MA.Residuals.Fst Combined.Fst
1 338.8 343.6964 -12.91321
                                                   330.7832
                            -12.91321
-12.01321
   290.3
                 305.3107
                                                   292.3975
   255.7
                 274.5679
                                                   261.6546
   253.9
                 281.6679
                                   -12.91321
                                                   268.7546
                276.8821
5
   262.1
                                   -12.91321
                                                   263.9689
                311.8393
6 284.0
                                   -12.91321
                                                   298.9261
6 284.0 311.8393
7 294.2 321.1679
8 338.6 346.6250
9 401.4 405.7964
10 414.4 425.6107
11 450.1 467.5964
12 491.5 522.1107
13 340.8 365.1884
14 302.2 326.8027
15 253.1 296.0598
16 233.2 303.1598
17 262.3 298.3741
18 325.6 333.3313
19 278.2 342.6598
20 343.8 368.1170
21 406.5 427.2884
22 426.2 447.1027
23 447.4 489.0884
               321.1679
   294.2
                                   -12.91321
                                                  308.2546
                                  -12.91321
                                                  333.7118
                                 -12.91321
                                                 392.8832
                                 -12.91321 412.6975
                                 -12.91321 454.6832
                               23 447.4
                489.0884
                                   -12.91321
                                                   476.1752
24 478.9
                543.6027
                                   -12.91321
                                                   530.6895
> round(accuracy(trend.seas.pred$mean, valid.ts), 3)
               ME RMSE MAE MPE MAPE ACF1 Theil's U
 Test set -27.102 32.526 27.102 -8.638 8.638 0.231 0.763
 > round(accuracy(fst.2level, valid.ts), 3)
               ME RMSE MAE MPE MAPE ACF1 Theil's U
 Test set -14.189 22.907 16.548 -4.651 5.313 0.231 0.538
```

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 the 12 months of 2024 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 the regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the 12 months of 2024.

We have developed the regression model with linear trend and seasonality and trailing MA with the window width of 3 for the regression residuals. We then used it to develop models to forecast the 12 months of 2024 and developed a two-level forecast for the 12 future months as a combination of the specified forecasts. The output is as follows:

```
> #Creating a table with regression forecast, trailing MA for residuals,
> #total forecast for future 12 periods.
> future12.df <- data.frame(tot.trend.seas.pred$mean, tot.ma.trail.res.pred$mean,</pre>
                          tot.fst.2level)
> future12.df
  tot.trend.seas.pred.mean tot.ma.trail.res.pred.mean tot.fst.2level
                 366.8021
                                          -22.21898 344.5831
1
2
                 327.2688
                                          -22.21898
                                                         305.0498
                                                         271.8387
3
                 294.0576
                                          -22.21898
                                          -22.21898
4
                  297.1688
                                                         274.9498
                                          -22.21898
5
                 297.5910
                                                         275.3720
                                          -22.21898
                                                        312.0275
                 334.2465
6
                                                       312.02/3
315.1498
347.1720
407.1387
                 337.3688
                                          -22.21898
7
                 369.3910
                                         -22.21898
8
9
                429.3576
                                         -22.21898
                                                       426.1831
10
                448.4021
                                         -22.21898
11
                487.3799
                                         -22.21898
                                                       465.1609
                537.8799
                                         -22.21898
                                                       515.6609
```

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 linear trend and seasonality, and two-level (combined) model with the 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 for forecasting monthly sales in 2024.

The accuracy measures are as follows. The two-level model is more accurate with MAPE and RMSE as 5.227% and 17.232, respectively.

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.

The Holt-Winters (HW) model with the automated selection of the model options and automated selection of the smoothing parameters for the training period has been made and the summary is as follows.

This HW model has the (A, A, A) options, i.e., additive error, additive trend, and additive seasonality. The optimal value for exponential smoothing constant (alpha) is 0.1951, beta is 0.0001, and gamma is the same 0.0001.

```
ETS(A,A,A)
 Call:
   ets(y = train.ts, model = "ZZZ")
    Smoothing parameters:
        alpha = 0.1951
        beta = 1e-04
        gamma = 1e-04
    Initial states:
        1 = 201.4349
        b = 1.569
        s = 157.8048 \ 103.5611 \ 60.5357 \ 42.4259 \ -13.9408 \ -25.3492
                      -46.1451 -78.2233 -72.2383 -78.0027 -43.2154 -7.2129
    sigma: 27.9821
          AIC
                       AICC
 948.1414 957.4142 989.4653
            Point Forecast
                                              Lo 0
Jan 2022 323.1783 323.1783 323.1783
Feb 2022
                      288.7394 288.7394 288.7394
Mar 2022
                      255.5209 255.5209 255.5209
Apr 2022
May 2022
                        262.8517 262.8517 262.8517
                         258.4331 258.4331 258.4331
                     292.0789 292.0789 292.0789
Jun 2022
Jul 2022
                     314.4322 314.4322 314.4322

    Jul 2022
    314.4322
    314.4322
    314.4322

    Aug 2022
    327.4177
    327.4177
    327.4177

    Sep 2022
    385.3516
    385.3516
    385.3516

    Oct 2022
    405.0276
    405.0276
    405.0276

    Nov 2022
    449.6175
    449.6175
    449.6175

    Dec 2022
    505.4270
    505.4270
    505.4270

    Jan 2023
    341.9785
    341.9785
    341.9785

    Feb 2023
    307.5396
    307.5396
    307.5396

    Mar 2023
    374.3210
    374.3210
    374.3210

Mar 2023
                         274.3210 274.3210 274.3210
Apr 2023
                        281.6519 281.6519 281.6519
May 2023
                      277.2332 277.2332 277.2332
Jun 2023
                      310.8791 310.8791 310.8791
Jul 2023 333.2323 333.2323 333.2323
Aug 2023 346.2178 346.2178 346.2178
Sep 2023 404.1517 404.1517 404.1517
Oct 2023 423.8278 423.8278 423.8278
Nov 2023 468.4177 468.4177 468.4177
Dec 2023 524.2271 524.2271
```

b. To make a forecast in the 12 months of 2024, 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 2024 using the forecast() function, and present the forecast in your report.

We developed a forecast using ets() function with the automated selections mentioned. This HW model has the (A, A, A) options, i.e., additive error, additive trend, and additive seasonality. The optimal value for exponential smoothing constant (alpha) is 0.2017, beta is 0.0001, and gamma is the same 0.0001.

Also, the summary is as follows:

```
> #Making a forecast in the 12 months of 2024
> HW.ZZZ <- ets(sales.ts, model = "ZZZ")
> HW.ZZZ
ETS(A,A,A)
ets(y = sales.ts, model = "ZZZ")
 Smoothing parameters:
   alpha = 0.2017
    beta = 1e-04
   gamma = 1e-04
 Initial states:
    1 = 202.496
   b = 1.3772
   s = 156.7496 \ 105.029 \ 65.6201 \ 47.3008 \ -11.9245 \ -38.2964
           -43.5536 -77.6554 -77.1076 -80.0095 -42.8196 -3.3328
  sigma: 25.1588
    ATC
                      BTC
             ATCC
1218.998 1225.798 1264.594
> #Using forecast() function to make predictions using the model for 12 months
> HW.ZZZ.pred <- forecast(HW.ZZZ, h = 12 , level = 0)
> HW.ZZZ.pred
        Point Forecast
                             100
Jan 2024 345.1928 345.1928 345.1928
              307.0794 307.0794 307.0794
Feb 2024
              271.2673 271.2673 271.2673 275.5425 275.5425 275.5425
Mar 2024
Apr 2024
May 2024
              276.3692 276.3692 276.3692
Jun 2024
             311.8473 311.8473 311.8473
             318.4755 318.4755 318.4755
Jul 2024
              346.2273 346.2273 346.2273
406.8265 406.8265 406.8265
Aug 2024
Sep 2024
Oct 2024
              426.5198 426.5198 426.5198
Nov 2024 467.3017 467.3017 467.3017
Dec 2024
              520.3950 520.3950 520.3950
> |
```

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.

The MAPE and RMSE in the HW are better than those for the seasonal naïve forecast. We can conclude that the HW is more accurate than the others.

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.

As we can see from the accuracy measures, the HW model (the bottom one) is better with MAPE and RMSE as 5.313% and 22.907, respectively.