David Liu
BAN 673

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 second quarter of 2020 (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 for the next two quarters of 2020 and the first two quarters of 2021.

Questions

1. Plot the data and visualize time series components.

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

```
walmart_rev.data <- read.csv("case2.csv")
walmart_rev.ts <- ts(walmart_rev.data$Revenue, start=c(5,1), end=c(20,2), freq=4)
walmart_rev.ts
```

Above are the codes for reading the data off the case2.csv file, and then forming it into a time series.

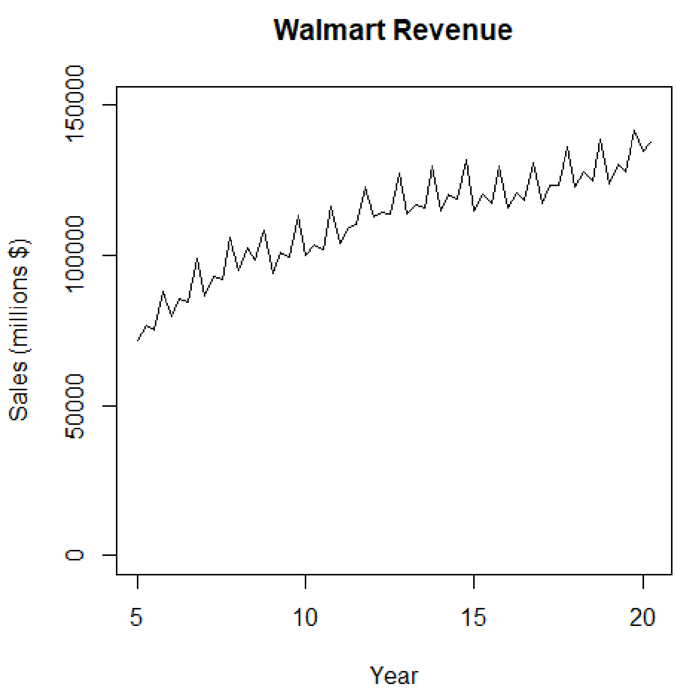
```
> walmart_rev.ts
     Qtr1   Qtr2   Qtr3   Qtr4
5   71680  76697  75397  88327
6   79676  85430  84467  98795
7   86410  92999  91865 105749
8   94940 102342  98345 108627
9   94242 100876  99373 113594
10  99811 103726 101952 116360
11 104189 109366 110226 122728
12 113010 114282 113800 127559
13 114070 116830 115688 129706
14 114960 120125 119001 131565
15 114826 120229 117408 129667
16 115904 120854 118179 130936
17 117542 123355 123179 136267
18 122690 128028 124894 138793
19 123925 130377 127991 141671
20 134622 137742
```

Above shows the output of the time series data.

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.

```
#1b. 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. A
plot(walmart_rev.ts, xlab="Year", ylab="Sales (millions $)", ylim=c(0,150000), main="Walmart Revenue")
```

Above shows the codes for building the plot below.



Above shows the plot of Walmart's quarterly revenue time series. Note that the Walmart revenue data tends to peak near the end of each year in a frequent and similar, non-multiplicative, pattern. The average revenue also gradually increases with each subsequent year, showing an obvious upward trend. Hence, the time series shows a trend component that is approximately linearly upwards, as well as a season component that can be described as additive seasonality.

David Liu
BAN 673

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.

```
validation <- 16
training <- length(walmart_rev.ts) - validation
walmart_rev_train.ts <- window(walmart_rev.ts, start = c(5,1), end = c(5, training))
walmart_rev_valid.ts <- window(walmart_rev.ts, start = c(5, training + 1),
                               end = c(5, training + validation))

walmart_rev_train.ts
walmart_rev_valid.ts
```

Codes above for partitioning 16 periods for validation data, and the rest for training.

```
> walmart_rev_train.ts
     Qtr1   Qtr2   Qtr3   Qtr4
5   71680  76697  75397  88327
6   79676  85430  84467  98795
7   86410  92999  91865 105749
8   94940 102342  98345 108627
9   94242 100876  99373 113594
10  99811 103726 101952 116360
11 104189 109366 110226 122728
12 113010 114282 113800 127559
13 114070 116830 115688 129706
14 114960 120125 119001 131565
15 114826 120229 117408 129667
16 115904 120854
```

Training data shown above

```
> walmart_rev_valid.ts
     Qtr1   Qtr2   Qtr3   Qtr4
16                118179 130936
17 117542 123355 123179 136267
18 122690 128028 124894 138793
19 123925 130377 127991 141671
20 134622 137742
```

Validation data shown above

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. Use each model to forecast revenues for the validation period using the forecast() function.

1 REGRESSION MODEL WITH LINEAR TREND

```
#  i. Regression model with linear trend
walmart_rev.reg.lin <- tslm(walmart_rev_train.ts ~ trend)
summary(walmart_rev.reg.lin)

walmart_rev.reg.lin.pred <- forecast(walmart_rev.reg.lin, h = 16, level = 0)
walmart_rev.reg.lin.pred
```

Above, codes for creating a regression model with linear trend, and a forecast for the validation period.

```
> #  i. Regression model with linear trend
> walmart_rev.reg.lin <- tslm(walmart_rev_train.ts ~ trend)
> summary(walmart_rev.reg.lin)

Call:
tslm(formula = walmart_rev_train.ts ~ trend)

Residuals:
     Min       1Q   Median       3Q      Max 
-11944.2  -4520.5   -553.5   2316.8  13030.8 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 81740.51    2009.22   40.68   <2e-16 ***
trend        1024.62      74.44   13.76   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6703 on 44 degrees of freedom
Multiple R-squared:  0.8115,    Adjusted R-squared:  0.8072 
F-statistic: 189.5 on 1 and 44 DF,  p-value: < 2.2e-16
```

Above shows the summary of the regression model with linear trend. We can see that this model uses the formula: walmart_rev_train.ts ~ trend, which fits a linear model to the Walmart time series data. The multiple and adjusted r-squared values of 0.8115 and 0.8072, respectively, in addition to a significant p-value of 2.2e-16, indicates that a purely linear model is relatively good for use with the Walmart data, and that the data is quite linear. The regression equation for this model can be displayed as the following, with t as the time period index:

$y_t = 81740.51 + 1024.62t$

```
> walmart_rev.reg.lin.pred <- forecast(walmart_rev.reg.lin, h = 16, level = 0)
> walmart_rev.reg.lin.pred
      Point Forecast     Lo 0     Hi 0
16 Q3       129897.4 129897.4 129897.4
16 Q4       130922.1 130922.1 130922.1
17 Q1       131946.7 131946.7 131946.7
17 Q2       132971.3 132971.3 132971.3
17 Q3       133995.9 133995.9 133995.9
17 Q4       135020.5 135020.5 135020.5
18 Q1       136045.1 136045.1 136045.1
18 Q2       137069.8 137069.8 137069.8
18 Q3       138094.4 138094.4 138094.4
18 Q4       139119.0 139119.0 139119.0
19 Q1       140143.6 140143.6 140143.6
19 Q2       141168.2 141168.2 141168.2
19 Q3       142192.8 142192.8 142192.8
19 Q4       143217.5 143217.5 143217.5
20 Q1       144242.1 144242.1 144242.1
20 Q2       145266.7 145266.7 145266.7
```

Above, forecast values for the validation period using the regression model with linear trend.

2 REGRESSION MODEL WITH QUADRATIC TREND

```
#  ii. Regression mode with quadratic trend
walmart_rev.reg.quad <- tslm(walmart_rev_train.ts ~ trend + I(trend^2))
summary(walmart_rev.reg.quad)

walmart_rev.reg.quad.pred <- forecast(walmart_rev.reg.quad, h = 16, level = 0)
walmart_rev.reg.quad.pred
```

Above, codes for creating a regression model with quadratic trend, and a forecast for the validation period.

```
> #  ii. Regression mode with quadratic trend
> walmart_rev.reg.quad <- tslm(walmart_rev_train.ts ~ trend + I(trend^2))
> summary(walmart_rev.reg.quad)

Call:
tslm(formula = walmart_rev_train.ts ~ trend + I(trend^2))

Residuals:
   Min     1Q Median     3Q    Max 
 -7833  -3881  -1569   3575  10929 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 73558.345   2671.671  27.533  < 2e-16 ***
trend        2047.386    262.210   7.808 8.88e-10 ***
I(trend^2)    -21.761      5.409  -4.023 0.000228 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5779 on 43 degrees of freedom
Multiple R-squared:  0.8631,	Adjusted R-squared:  0.8567 
F-statistic: 135.5 on 2 and 43 DF,  p-value: < 2.2e-16
```

Above shows a summary of the regression model with quadratic trend. This model uses the formula: walmart_rev_train.ts ~ trend I(trend^2), which fits a purely quadratic model to the

Walmart time series data. The model has a high multiple r-squared value of 0.8631, a high adjusted r-squared value of 0.8567, and a significant p-value of 2.2e-16, indicating that a quadratic model is good for use with the Walmart time series data. The regression equation for this model can be displayed as the following, with t as the time period index:

$y_t = 73558.345 + 2047.386t – 21.761t^2$

```
> walmart_rev.reg.quad.pred <- forecast(walmart_rev.reg.quad, h = 16, level = 0)
> walmart_rev.reg.quad.pred
      Point Forecast     Lo 0     Hi 0
16 Q3       121715.3 121715.3 121715.3
16 Q4       121695.4 121695.4 121695.4
17 Q1       121631.9 121631.9 121631.9
17 Q2       121525.0 121525.0 121525.0
17 Q3       121374.5 121374.5 121374.5
17 Q4       121180.5 121180.5 121180.5
18 Q1       120943.0 120943.0 120943.0
18 Q2       120661.9 120661.9 120661.9
18 Q3       120337.3 120337.3 120337.3
18 Q4       119969.3 119969.3 119969.3
19 Q1       119557.6 119557.6 119557.6
19 Q2       119102.5 119102.5 119102.5
19 Q3       118603.8 118603.8 118603.8
19 Q4       118061.7 118061.7 118061.7
20 Q1       117476.0 117476.0 117476.0
20 Q2       116846.7 116846.7 116846.7
```

Above, forecast values for the validation period using the regression model with quadratic trend.

3 REGRESSION MODEL WITH SEASONALITY

```
#  iii. Regression model with seasonality
walmart_rev.reg.seas <- tslm(walmart_rev_train.ts ~ season)
summary(walmart_rev.reg.seas)

walmart_rev.reg.seas.pred <- forecast(walmart_rev.reg.seas, h = 16, level = 0)
walmart_rev.reg.seas.pred
```

Above, codes for creating a regression model with seasonality, and a forecast for the validation period.

```
> #  iii. Regression model with seasonality
> walmart_rev.reg.seas <- tslm(walmart_rev_train.ts ~ season)
> summary(walmart_rev.reg.seas)

Call:
tslm(formula = walmart_rev_train.ts ~ season)

Residuals:
     Min       1Q   Median       3Q      Max
-28629.8  -9229.4     81.6  13616.6  16499.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   100310       4208  23.835   <2e-16 ***
season2         5003       5952   0.841   0.4053
season3         2192       6086   0.360   0.7205
season4        15388       6086   2.529   0.0153 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14580 on 42 degrees of freedom
Multiple R-squared:  0.1489,    Adjusted R-squared:  0.08812
F-statistic:  2.45 on 3 and 42 DF,  p-value: 0.07681
```

Above shows the summary of the regression model with seasonality. This model uses the formula: walmart_rev_train.ts ~ season, which fits a purely seasonal model to the Walmart time series data. The model has a low multiple r-squared value of 0.1489, a low adjusted r-squared value of 0.08812, and a not very significant p-value of 0.07681, indicating that a seasonal model with no trend is not good for use with the Walmart time series data. The regression equation for this model can be displayed as the following, with D_n as a binary categorical variable that describes the season (or quarter in this case) of the Walmart revenue data:

$y_t = 100310 + 5003D_2 + 2192D_3 + 15388D_4$

D_1 is not included in the equation because having D_2 to D_4 as 0 would already give the assumption that D_1 is 1.

```
> walmart_rev.reg.seas.pred <- forecast(walmart_rev.reg.seas, h = 16, level = 0)
> walmart_rev.reg.seas.pred
      Point Forecast     Lo 0     Hi 0
16 Q3       102502.0 102502.0 102502.0
16 Q4       115697.9 115697.9 115697.9
17 Q1       100309.8 100309.8 100309.8
17 Q2       105313.0 105313.0 105313.0
17 Q3       102502.0 102502.0 102502.0
17 Q4       115697.9 115697.9 115697.9
18 Q1       100309.8 100309.8 100309.8
18 Q2       105313.0 105313.0 105313.0
18 Q3       102502.0 102502.0 102502.0
18 Q4       115697.9 115697.9 115697.9
19 Q1       100309.8 100309.8 100309.8
19 Q2       105313.0 105313.0 105313.0
19 Q3       102502.0 102502.0 102502.0
19 Q4       115697.9 115697.9 115697.9
20 Q1       100309.8 100309.8 100309.8
20 Q2       105313.0 105313.0 105313.0
```

Above, forecast values for the validation period using the regression model with seasonality.

David Liu
BAN 673

4 REGRESSION MODEL WITH LINEAR TREND AND SEASONALITY

```
#  iv. Regression model with linear trend and seasonality
walmart_rev.reg.lin.seas <- tslm(walmart_rev_train.ts ~ trend + season)
summary(walmart_rev.reg.lin.seas)

walmart_rev.reg.lin.seas.pred <- forecast(walmart_rev.reg.lin.seas, h = 16, level = 0)
walmart_rev.reg.lin.seas.pred
```

Above, codes for creating a regression model with linear trend and seasonality, and a forecast for the validation period.

```
> #  iv. Regression model with linear trend and seasonality
> walmart_rev.reg.lin.seas <- tslm(walmart_rev_train.ts ~ trend + season)
> summary(walmart_rev.reg.lin.seas)

Call:
tslm(formula = walmart_rev_train.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-7102.9 -1775.9   793.3  2007.1  7163.0 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 77001.66    1489.86  51.684  < 2e-16 ***
trend        1013.40      42.94  23.599  < 2e-16 ***
season2      3989.77    1578.02   2.528   0.0154 *  
season3      2192.17    1612.89   1.359   0.1815    
season4     14374.68    1613.46   8.909 3.88e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3864 on 41 degrees of freedom
Multiple R-squared:  0.9416,	Adjusted R-squared:  0.9359 
F-statistic: 165.4 on 4 and 41 DF,  p-value: < 2.2e-16
```

Above shows a summary of the regression model with linear trend and seasonality. This model uses the formula: walmart_rev_train.ts ~ trend + season, which fits a model with linear trend and seasonality to the Walmart time series data. The model has a high multiple r-squared value of 0.9416, a high adjusted r-squared value of 0.9359, and a significant p-value of 2.2e-16, indicating that a model with linear trend and seasonality is good for use with the Walmart time series data. The regression equation for this model can be displayed as the following, with t as the time period index, and D_n as a binary categorical variable that describes the season (or quarter in this case) of the Walmart revenue data:

$y_t = 77001.66 + 1013.40t + 3989.77D_2 + 2192.17D_3 + 14374.68D_4$

```
> walmart_rev.reg.lin.seas.pred <- forecast(walmart_rev.reg.lin.seas, h = 16, level = 0)
> walmart_rev.reg.lin.seas.pred
      Point Forecast     Lo 0     Hi 0
16 Q3       126823.6 126823.6 126823.6
16 Q4       140019.5 140019.5 140019.5
17 Q1       126658.2 126658.2 126658.2
17 Q2       131661.4 131661.4 131661.4
17 Q3       130877.2 130877.2 130877.2
17 Q4       144073.1 144073.1 144073.1
18 Q1       130711.8 130711.8 130711.8
18 Q2       135715.0 135715.0 135715.0
18 Q3       134930.8 134930.8 134930.8
18 Q4       148126.7 148126.7 148126.7
19 Q1       134765.4 134765.4 134765.4
19 Q2       139768.6 139768.6 139768.6
19 Q3       138984.4 138984.4 138984.4
19 Q4       152180.3 152180.3 152180.3
20 Q1       138819.0 138819.0 138819.0
20 Q2       143822.2 143822.2 143822.2
```

Above, forecast values for the validation period using the regression model with linear trend and seasonality.

5 REGRESSION MODEL WITH QUADRATIC TREND AND SEASONALITY

```
#  v. Regression model with quadratic trend and seasonality.
walmart_rev.reg.quad.seas <- tslm(walmart_rev_train.ts ~ trend + I(trend^2) + season)
summary(walmart_rev.reg.quad.seas)

walmart_rev.reg.quad.seas.pred <- forecast(walmart_rev.reg.quad.seas, h = 16, level = 0)
walmart_rev.reg.quad.seas.pred
```

Above, codes for creating a regression model with quadratic trend and seasonality, and a forecast for the validation period.

```
> #  v. Regression model with quadratic trend and seasonality.
> walmart_rev.reg.quad.seas <- tslm(walmart_rev_train.ts ~ trend + I(trend^2) + season)
> summary(walmart_rev.reg.quad.seas)

Call:
tslm(formula = walmart_rev_train.ts ~ trend + I(trend^2) + season)

Residuals:
   Min     1Q Median     3Q    Max 
 -3752  -1380     75   1360   5148 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 69770.953   1025.702  68.023  < 2e-16 ***
trend        1953.925     90.956  21.482  < 2e-16 ***
I(trend^2)    -20.011      1.877 -10.662 2.94e-13 ***
season2      3989.768    815.059   4.895 1.65e-05 ***
season3      1578.490    835.055   1.890    0.066 .  
season4     13761.001    835.349  16.473  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1996 on 40 degrees of freedom
Multiple R-squared:  0.9848,	Adjusted R-squared:  0.9829 
F-statistic: 518.7 on 5 and 40 DF,  p-value: < 2.2e-16
```

Above shows the summary of the regression model with quadratic trend and seasonality. This model uses the formula: walmart_rev_train.ts ~ trend + I(trend^2) + season, which fits a model with quadratic trend and seasonality to the Walmart time series data. The model has a high multiple r-squared value of 0.9848, a high adjusted r-squared value of 0.9829, and a significant p-value of 2.2e-16, indicating that a model with quadratic trend and seasonality is good for use with the Walmart time series data. The regression equation for this model can be displayed as the following, with t as the time period index, and D_n as a binary categorical variable that describes the season (or quarter in this case) of the Walmart revenue data:

$y_t = 69770.953 + 1953.925t - 20.011t^2 + 3989.768D_2 + 1578.490D_3 + 13761.001D_4$

```
> walmart_rev.reg.quad.seas.pred <- forecast(walmart_rev.reg.quad.seas, h = 16, level = 0)
> walmart_rev.reg.quad.seas.pred
      Point Forecast     Lo 0     Hi 0
16 Q3       118979.2 118979.2 118979.2
16 Q4       131214.6 131214.6 131214.6
17 Q1       117466.4 117466.4 117466.4
17 Q2       121429.0 121429.0 121429.0
17 Q3       118950.5 118950.5 118950.5
17 Q4       131025.8 131025.8 131025.8
18 Q1       117117.5 117117.5 117117.5
18 Q2       120920.0 120920.0 120920.0
18 Q3       118281.4 118281.4 118281.4
18 Q4       130196.6 130196.6 130196.6
19 Q1       116128.3 116128.3 116128.3
19 Q2       119770.7 119770.7 119770.7
19 Q3       116972.0 116972.0 116972.0
19 Q4       128727.1 128727.1 128727.1
20 Q1       114498.7 114498.7 114498.7
20 Q2       117981.0 117981.0 117981.0
```

Above, forecast values for the validation period using the regression model with quadratic trend and seasonality.

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 two most accurate regression models for forecasting.

```
#2c. 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 two most accurate regression models for forecasting.
reg.lin.measures <- accuracy(walmart_rev.reg.lin.pred, walmart_rev_valid.ts)
reg.quad.measures <- accuracy(walmart_rev.reg.quad.pred, walmart_rev_valid.ts)
reg.seas.measures <- accuracy(walmart_rev.reg.seas.pred, walmart_rev_valid.ts)
reg.lin.seas.measures <- accuracy(walmart_rev.reg.lin.seas.pred, walmart_rev_valid.ts)
reg.quad.seas.measures <- accuracy(walmart_rev.reg.quad.seas.pred, walmart_rev_valid.ts)

models <- c("Linear Trend","Quadratic Trend","Seasonality","Linear Trend with Seasonality",
            "Quadratic Trend with Seasonality")
RMSE <- c(reg.lin.measures[4], reg.quad.measures[4], reg.seas.measures[4],
          reg.lin.seas.measures[4], reg.quad.seas.measures[4])
MAPE <- c(reg.lin.measures[10], reg.quad.measures[10], reg.seas.measures[10],
          reg.lin.seas.measures[10], reg.quad.seas.measures[10])
accuracy_table <- data.frame(models, RMSE, MAPE)

accuracy_table
```

Above, codes for creating the forecast performance measures of the 5 forecasts in 2b.

David Liu
BAN 673

```
> reg.lin.measures
                        ME      RMSE      MAE        MPE     MAPE     MASE        ACF1 Theil's U
Training set -3.165004e-13  6555.452 5127.986 -0.4509279 4.931348 1.174925 -0.08439163        NA
Test set     -8.820130e+03 10377.449 8977.681 -7.0459390 7.161610 2.056967 -0.53002856  1.020057
> reg.quad.measures
                       ME      RMSE      MAE        MPE     MAPE     MASE       ACF1 Theil's U
Training set 9.490186e-13  5587.708 4577.867 -0.2780623 4.296914 1.048881 -0.4246096        NA
Test set     8.719292e+03 11998.816 9672.569  6.4419285 7.250912 2.216180  0.1766465  1.238746
> reg.seas.measures
                       ME     RMSE      MAE       MPE     MAPE     MASE      ACF1 Theil's U
Training set 1.893867e-12 13930.36 11822.55 -1.974843 11.93114 2.708782 0.9285386        NA
Test set     2.280625e+04 23374.42 22806.25 17.636390 17.63639 5.225370 0.7200383  2.374952
> reg.lin.seas.measures
                        ME     RMSE      MAE        MPE     MAPE      MASE      ACF1 Theil's U
Training set  3.164618e-13 3647.876 2878.373 -0.1809332 2.844026 0.6594931 0.8270902        NA
Test set     -8.609108e+03 8776.477 8609.108 -6.7195667 6.719567 1.9725195 0.3087670 0.8925647
> reg.quad.seas.measures
                       ME     RMSE      MAE         MPE     MAPE      MASE      ACF1 Theil's U
Training set 6.327577e-13 1861.038 1517.708 -0.03089353 1.463339 0.3477373 0.6694205        NA
Test set     7.533251e+03 9703.954 7668.093  5.69147321 5.802703 1.7569140 0.7495744 0.9754404
```

```
> accuracy_table
                             models      RMSE      MAPE
1                      Linear Trend 10377.449  7.161610
2                   Quadratic Trend 11998.816  7.250912
3                       Seasonality 23374.423 17.636390
4    Linear Trend with Seasonality  8776.477  6.719567
5 Quadratic Trend with Seasonality  9703.954  5.802703
```

Above table shows RMSE and MAPE of the 5 forecasts. It is clear that the model with linear trend with seasonality and the model with quadratic trend with seasonality are the two models with the lowest RMSE and MAPE; Hence, these are the two best models for forecasting.

David Liu
BAN 673

3. Employ the entire data set to make time series forecast.

a. Apply the two most accurate regression models identified in question to make the forecast for the last two quarters of 2020 and first two quarters of 2021. 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. Use each model to forecast Walmart's revenue in the 4 quarters of 2020 and 2021 using the forecast() function, and present this forecast in your report.

1 REGRESSION MODEL WITH LINEAR TREND AND SEASONALITY FITTED TO ALL DATA

```
#Regression model with linear trend and seasonality using all data
walmart_rev.reg.lin.seas.all <- tslm(walmart_rev.ts ~ trend + season)
summary(walmart_rev.reg.lin.seas.all)

walmart_rev.reg.lin.seas.all.pred <- forecast(walmart_rev.reg.lin.seas.all, h=4, level=0)
walmart_rev.reg.lin.seas.all.pred
```

Above codes create the model with linear trend and seasonality using all Walmart revenue data

```
> walmart_rev.reg.lin.seas.all <- tslm(walmart_rev.ts ~ trend + season)
> summary(walmart_rev.reg.lin.seas.all)

Call:
tslm(formula = walmart_rev.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-9124.2 -3125.8   499.2  3413.9  8312.8 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 79919.20    1418.21  56.352  < 2e-16 ***
trend         854.41      30.23  28.264  < 2e-16 ***
season2      4193.15    1505.72   2.785  0.00726 ** 
season3      1711.60    1530.30   1.118  0.26806    
season4     14095.79    1530.60   9.209 7.04e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4258 on 57 degrees of freedom
Multiple R-squared:  0.9411,	Adjusted R-squared:  0.937 
F-statistic: 227.9 on 4 and 57 DF,  p-value: < 2.2e-16
```

In the above summary, the formula: walmart_rev.ts ~ trend + season fits a model with a linear trend and includes seasonality. The model has a multiple r-squared value of 0.9411, an adjusted r-squared value of 0.937, and a significant p-value of 2.2e-16, indicating that the model fits well with all the Walmart revenue data. Shown in the coefficients portion of the summary, the model has the following equation, with t as the time period index and D_n as a binary categorical variable that describes the season (or quarter in this case) of the Walmart revenue data:

$y_t = 79919.20 + 854.41t + 4193.15D_2 + 1711.60D_3 + 14095.79D_4$

David Liu
BAN 673

```
> walmart_rev.reg.lin.seas.all.pred <- forecast(walmart_rev.reg.lin.seas.all, h=4, level=0)
> walmart_rev.reg.lin.seas.all.pred
      Point Forecast     Lo 0     Hi 0
20 Q3       135458.9 135458.9 135458.9
20 Q4       148697.5 148697.5 148697.5
21 Q1       135456.2 135456.2 135456.2
21 Q2       140503.7 140503.7 140503.7
```

Above shows the forecasted revenues for quarter 3 and 4 of 2020, and quarter 1 and 2 of 2021 using the model with linear trend and seasonality.

2 REGRESSION MODEL WITH QUADRATIC TREND AND SEASONALITY FITTED TO ALL DATA

```
#Regression model with quadratic trend and seasonality using all data
walmart_rev.reg.quad.seas.all <- tslm(walmart_rev.ts ~ trend + I(trend^2) + season)
summary(walmart_rev.reg.quad.seas.all)

walmart_rev.reg.quad.seas.all.pred <- forecast(walmart_rev.reg.quad.seas.all, h=4, level=0)
walmart_rev.reg.quad.seas.all.pred
```

Above codes create the model with quadratic trend and seasonality using all Walmart revenue data

```
> summary(walmart_rev.reg.quad.seas.all)

Call:
tslm(formula = walmart_rev.ts ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-4605.0 -1701.3    25.3  1596.7  8218.6 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 72987.016   1268.134  57.555  < 2e-16 ***
trend        1524.243     82.998  18.365  < 2e-16 ***
I(trend^2)    -10.632      1.277  -8.325 2.26e-11 ***
season2      4193.148   1015.506   4.129 0.000122 ***
season3      1272.140   1033.433   1.231 0.223474    
season4     13656.326   1033.634  13.212  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2872 on 56 degrees of freedom
Multiple R-squared:  0.9737,	Adjusted R-squared:  0.9713 
F-statistic: 414.6 on 5 and 56 DF,  p-value: < 2.2e-16
```

In the above summary, the formula: walmart_rev.ts ~ trend + I(trend^2) + season fits a model with a quadratic trend and includes seasonality. The model has a multiple r-squared value of 0.9737, an adjusted r-squared value of 0.9713, and a significant p-value of 2.2e-16, indicating that the model fits well with all the Walmart revenue data. Shown in the coefficients portion of the summary, the model has the following equation, with t as the time period index and D_n as a binary categorical variable that describes the season (or quarter in this case) of the Walmart revenue data:

$y_t = 72987.016 + 1524.243t - 10.632t^2 + 4193.148D_2 + 1272.140D_3 + 13656.326D_4$

```
> walmart_rev.reg.quad.seas.all.pred <- forecast(walmart_rev.reg.quad.seas.all, h=4, level=0)
> walmart_rev.reg.quad.seas.all.pred
      Point Forecast    Lo 0     Hi 0
20 Q3       128087.3 128087.3 128087.3
20 Q4       140645.4 140645.4 140645.4
21 Q1       127141.8 127141.8 127141.8
21 Q2       131466.4 131466.4 131466.4
```

Above shows the forecasted revenues for quarter 3 and 4 of 2020, and quarter 1 and 2 of 2021 using the model with quadratic trend and seasonality.

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 2020 and 2021.

```
#3b.  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 2020 and 2021.

#naive forecast
walmart_rev.naive <- naive(walmart_rev.ts)

#seasonal naive forecast
walmart_rev.snaive <- snaive(walmart_rev.ts)

# accuracy measures between regression model with linear trend and seasonality, regression model
# with quadratic trend and seasonality, naive model, and seasonal naive model. All models uses
# entire walmart revenue data.
linear.seas.measures <- accuracy(walmart_rev.reg.lin.seas.all.pred$fitted, walmart_rev.ts)
quad.seas.measures <- accuracy(walmart_rev.reg.quad.seas.all.pred$fitted, walmart_rev.ts)
naive.measures <- accuracy(walmart_rev.naive$fitted, walmart_rev.ts)
snaive.measures <- accuracy(walmart_rev.snaive$fitted, walmart_rev.ts)

linear.seas.measures
quad.seas.measures
naive.measures
snaive.measures

all_data.models <- c("Linear with seasonality","Quadratic with seasonality","Naive","Seasonal naive")
all_data.RMSE <- c(linear.seas.measures[2], quad.seas.measures[2], naive.measures[2], snaive.measures[2])
all_data.MAPE <- c(linear.seas.measures[5], quad.seas.measures[5], naive.measures[5], snaive.measures[5])
all_data.accuracy_table <- data.frame(all_data.models, all_data.RMSE, all_data.MAPE)

all_data.accuracy_table
```

Codes for finding the accuracy measures between regression model with linear trend and seasonality, regression model with quadratic trend and seasonality, naive model, and seasonal naive model.

David Liu
BAN 673

```
> linear.seas.measures
                    ME     RMSE     MAE        MPE     MAPE      ACF1 Theil's U
Test set -4.697175e-13 4082.663 3416.38 -0.2241946 3.247459 0.8748842 0.4631695
> quad.seas.measures
                   ME     RMSE      MAE         MPE     MAPE      ACF1 Theil's U
Test set 2.348158e-13 2729.216 2077.481 -0.07588254 1.888066 0.7545759  0.281903
> naive.measures
               ME    RMSE      MAE       MPE     MAPE       ACF1 Theil's U
Test set 1082.984 9614.53 8103.902 0.6996276 7.246587 -0.7155763         1
> snaive.measures
               ME     RMSE      MAE      MPE     MAPE      ACF1 Theil's U
Test set 3964.224 5074.236 4163.845 3.695594 3.872531 0.7530472 0.5576718
```

Original accuracy measures for the four models.

```
> all_data.accuracy_table
             all_data.models all_data.RMSE all_data.MAPE
1    Linear with seasonality      4082.663      3.247459
2 Quadratic with seasonality      2729.216      1.888066
3                      Naive      9614.530      7.246587
4             Seasonal naive      5074.236      3.872531
```

The above table shows all relevant accuracy measures (RMSE and MAPE) for the four models. Using RMSE and MAPE as metrics, both models performed better than the naïve and seasonal naïve models. Between the two models, the model with quadratic trend and seasonality has a lower RMSE and MAPE, making it the most accurate model for forecasting the four quarters from 2020 to 2021.