Look into Monthly Production of Toyota Passenger cars

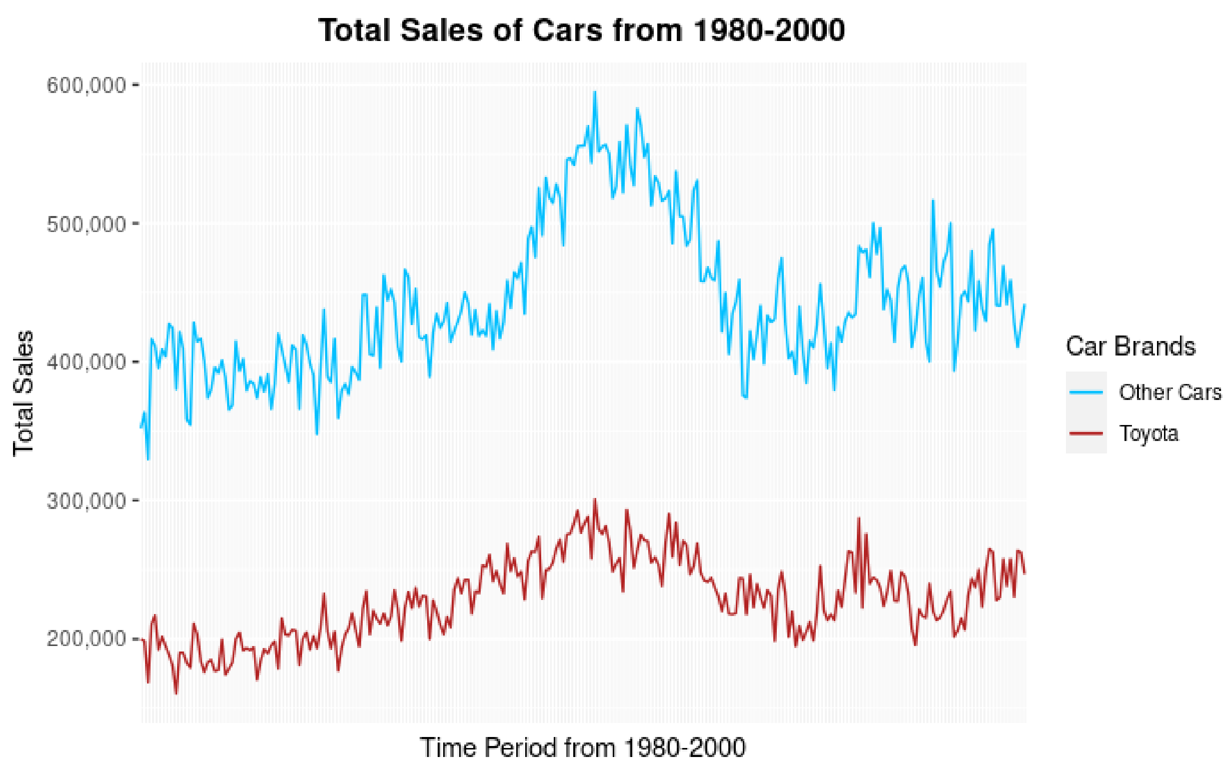
Sanjaya J Shetty

01/06/2021

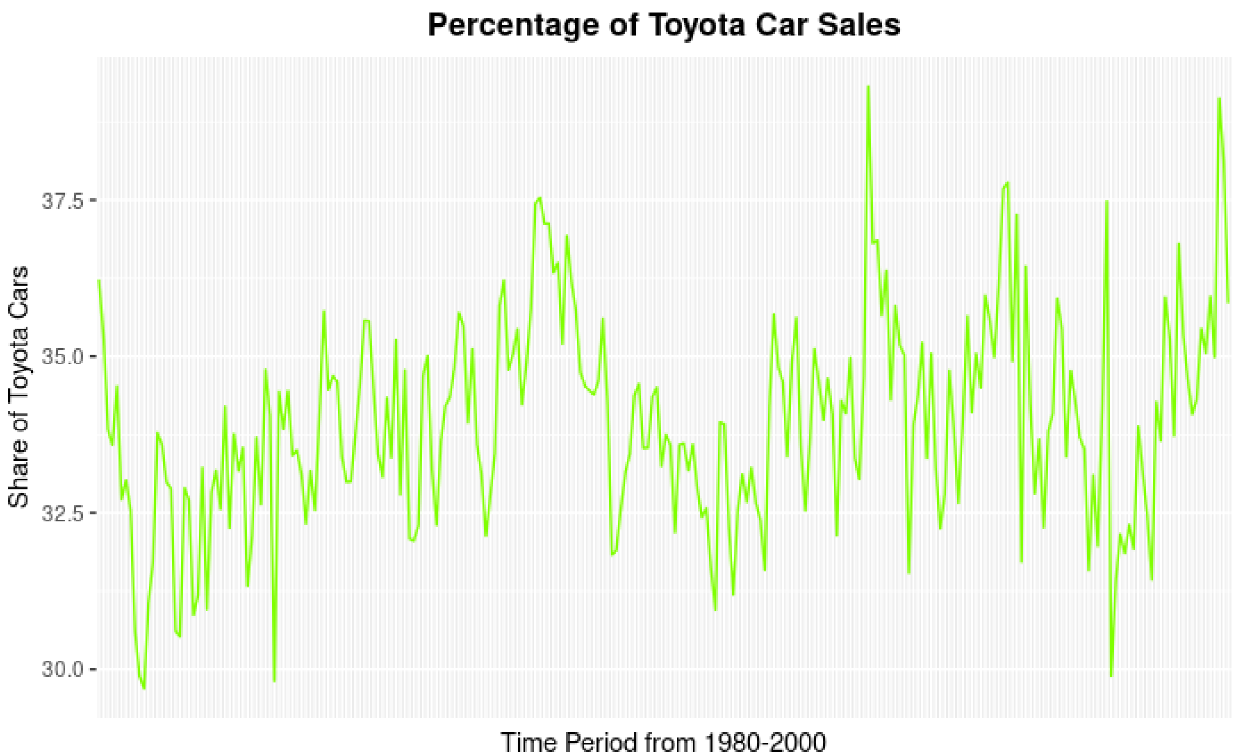
Questions

- **(a).** Make time series plots of the variables yt and xt, and also of the share of Toyota in all produced passenger cars, that is yt/(yt + xt). What conclusions do you draw from these plots?

Answers:



There is no defining trend in the above plot



The is no great flucation in the numbers

- **(b)**
  - **(i)** Perform the Augmented Dickey-Fuller (ADF) test for yt. In the ADF test equation, include a constant ( ) and three lags of yt, as well as the variable of interest, yt−1. Report the coefficient of yt−1 and its standard error and t-value, and draw your conclusion.

Answers:

Table 1	Estimate	Std. Error	t value
LagTOYOTA	0.07478	0.03673	2.036

- **(ii)** Perform a similar ADF test for xt.

Answer:

Table 2	Estimate	Std. Error	t value
LagOthers	0.05290	0.03356	1.576

- **(c).** Perform the two-step Engle-Granger test for cointegration of the time series yt and xt. In step 1, regress yt on a constant and xt. In step 2, perform a regression of the residuals et in the model et = + et−1 + 1 et−1 + 2 et−2 + 3 et−3 + t. What is your conclusion?

Answers:

step 1:

yt = 28428.168 + 0.4501 xt

step 2:

$dE = 191.27+0.28956*dE_t + 0.586\ dE_{t-1} + 0.14\ dE_{t-2} + 0.096\ dE_{t-3}$

using Engle-Granger for cointegration test we get

Engle-Granger Cointegration Test alternative: cointegrated

lag	EG	p.value
3.00	-4.73	0.01

Hence we could conclude that both variable are cointegrated.

- **(d).** Construct the first twelve sample autocorrelations and sample partial autocorrelations of yt and use the outcomes to motivate an AR(12) model for yt. Check that only the lagged terms at lags 1 to 5, 10, and 12 are significant, and estimate the following model: yt = + P5 j=1 j yt−j + 6 yt−10 + 7 yt−12 + t (recall that the estimation sample is Jan 1980 - Dec 1999).

Table3	Estimate	Std. Error	t value	Pr(>
(Intercept)	518.49151	897.80065	0.578	0.564165
LagDelta_1	-0.61038	0.06391	-9.550	< 0.0000000000000002 ***
LagDelta_2	-0.30488	0.07550	-4.038	0.0000737 ***
LagDelta_3	-0.24722	0.07634	-3.238	0.001383 **
LagDelta_4	-0.23912	0.07793	-3.069	0.002412 **
LagDelta_5	-0.18718	0.07952	-2.354	0.019426 *
LagDelta_6	-0.10445	0.08001	-1.305	0.193076
LagDelta_7	-0.13765	0.08012	-1.718	0.087161 .
LagDelta_8	0.04058	0.07999	0.507	0.612411
LagDelta_9	0.02838	0.07869	0.361	0.718745
LagDelta_10	-0.26206	0.07675	-3.415	0.000756 ***
LagDelta_11	-0.03910	0.07596	-0.515	0.607243
LagDelta_12	0.23236	0.06400	3.631	0.000349 ***

- **(e).** Extend the model of part (d) by adding the Error Correcion (EC) term (yt −0.45xt), that is, estimate the ECM yt = + (yt−1 − 0.45xt−1) + P5 j=1 j yt−j + 6 yt−10 + 7 yt−12 + t (estimation sample is Jan 1980- Dec 1999). Check that the EC term is significant at the 5% level, but not at the 1% level.

Table 4	Estimate	Std. Error	t value	Pr(>
(Intercept)	-3039.85443	1922.07272	-1.582	0.11512
LagDelta_1	-0.54566	0.06449	-8.462	0.00000000000000305
LagDelta_2	-0.24207	0.07489	-3.232	0.00141
LagDelta_3	-0.19619	0.07321	-2.680	0.00789
LagDelta_4	-0.18636	0.06931	-2.689	0.00769
LagDelta_5	-0.09950	0.05918	-1.681	0.09406
LagDelta_10	-0.26582	0.05002	-5.314	0.00000025230739777
LagDelta_12	0.25515	0.05278	4.834	0.00000244194347778
ErrorCorrection	0.12523	0.06136	2.041	0.04239

- **(f).** Use the models of parts (d) and (e) to make two series of 12 forecasts of monthly changes in production of Toyota passenger cars in 2000. At each month, you should use the data that are then available, for example, to forecast production for September 2000 you can use the data up to and including August 2000. However, do not re-estimate the model and use the coefficients as obtained in parts (d) and (e). For each of the two forecast series, compute the values of the root mean squared error

(RMSE) and of the mean absolute error (MAE). Finally, give your interpretation of the outcomes.

Table 5	AR(12)	ECM
RMSE	14443.15	27657.25
MAE	11351.77	21305.06

hence, Not Ideal to predict the sale of TOYOTA

Methods

Load the required Library

```
library(readxl)
library(ggplot2)
options(scipen=999)
options(warn = -1)
```

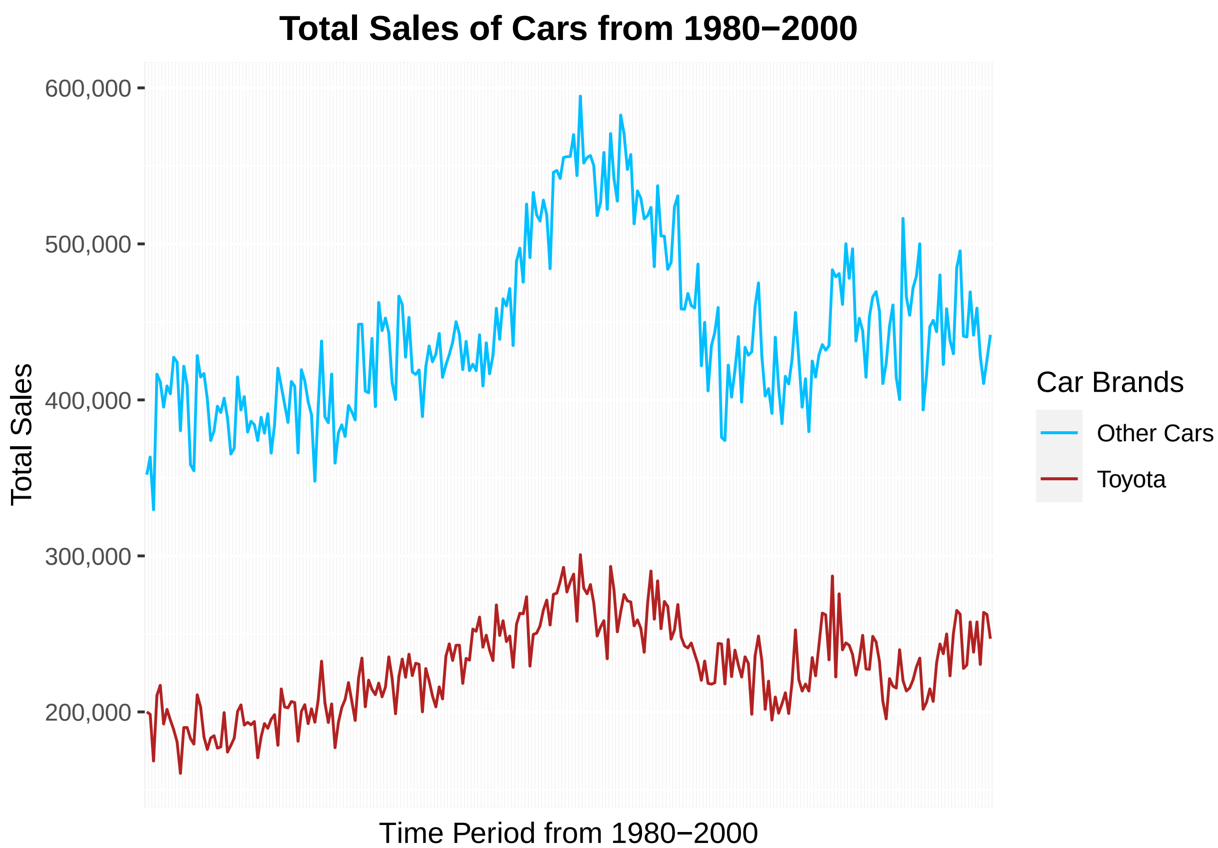
Load the Data

```
data <- read_xlsx('TestExer6-CARS-round2.xlsx')

head(data)
```

```
## # A tibble: 6 x 5
##   `YYYY-MM` TOYOTA  OTHER TOYOTA_SA OTHER_SA
##   <chr>      <dbl>  <dbl>     <dbl>    <dbl>
## 1 1980M01   175734 315111   200015.  352054.
## 2 1980M02   200479 377893   198443.  363393.
## 3 1980M03   200373 385236   168488.  329570.
## 4 1980M04   211636 404110   210573.  416497.
## 5 1980M05   208527 371930   217037.  411574.
## 6 1980M06   203901 420137   192241.  395403.
```

```
ggplot(data, aes(x = `YYYY-MM`))+
  geom_line(aes( y = OTHER_SA, group = 1, color = "Other Cars"))+
  geom_line(aes(y = TOYOTA_SA, group = 1, color = "Toyota"))+
  labs(y = "Total Sales", x = "Time Period from 1980-2000", title = "Total Sales of Cars from 1980-2000
       color = "Car Brands")+
  scale_color_manual(values = c(
    "Other Cars" = "deepskyblue1",
    "Toyota" = "firebrick"))+
  theme(axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        plot.title = element_text(hjust = 0.5, face = 'bold'))+
  scale_y_continuous(labels = scales::comma)
```




```
dataCarShare <- subset(data, select  =c('YYYY-MM', 'TOYOTA_SA', "OTHER_SA"))

head(dataCarShare)
```

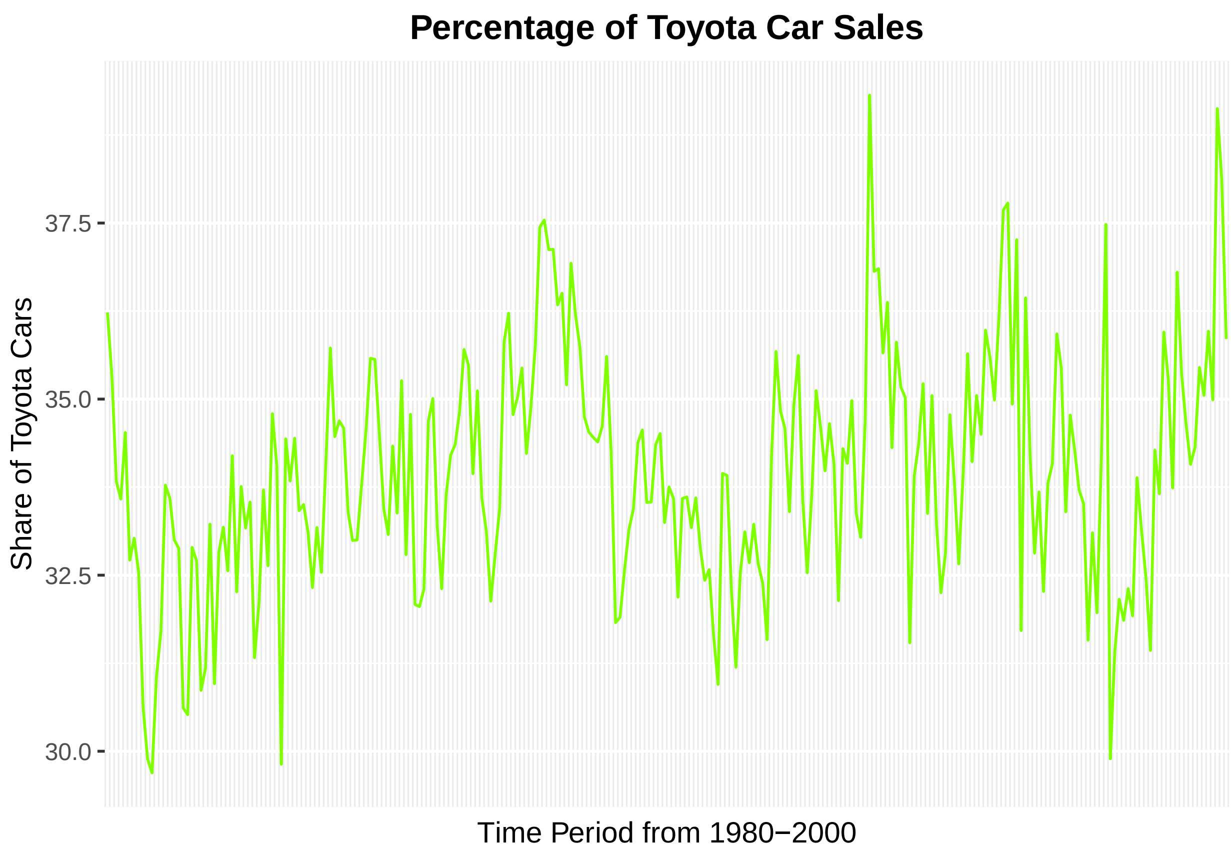
```
## # A tibble: 6 x 3
##   `YYYY-MM` TOYOTA_SA OTHER_SA
##   <chr>         <dbl>    <dbl>
## 1 1980M01     200015.  352054.
## 2 1980M02     198443.  363393.
## 3 1980M03     168488.  329570.
## 4 1980M04     210573.  416497.
## 5 1980M05     217037.  411574.
## 6 1980M06     192241.  395403.
```

```
dataCarShare$ShareTOYOTA <- dataCarShare$TOYOTA_SA/(data$TOYOTA_SA + data$OTHER_SA)*100

head(dataCarShare)
```

```
## # A tibble: 6 x 4
##   `YYYY-MM` TOYOTA_SA OTHER_SA ShareTOYOTA
##   <chr>         <dbl>    <dbl>       <dbl>
## 1 1980M01     200015.  352054.        36.2
## 2 1980M02     198443.  363393.        35.3
## 3 1980M03     168488.  329570.        33.8
## 4 1980M04     210573.  416497.        33.6
## 5 1980M05     217037.  411574.        34.5
## 6 1980M06     192241.  395403.        32.7
```

```
ggplot(dataCarShare, aes(`YYYY-MM`,ShareTOYOTA, group = 1))+
  geom_line(col = 'chartreuse1')+
  labs(y= 'Share of Toyota Cars', x=  'Time Period from 1980-2000', title = 'Percentage of Toyota Car Sa
  theme(axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        plot.title = element_text(hjust = 0.5, face = 'bold'))
```




```
library(tseries)

tseries::adf.test(data$TOYOTA_SA, k = 3)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  data$TOYOTA_SA
## Dickey-Fuller = -2.5096, Lag order = 3, p-value = 0.3612
## alternative hypothesis: stationary
```

```
tseries::adf.test(data$OTHER_SA, k = 3)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  data$OTHER_SA
## Dickey-Fuller = -2.0658, Lag order = 3, p-value = 0.5481
## alternative hypothesis: stationary
```

```r
library(aTSA)

coint.test( data$TOYOTA_SA,data$OTHER_SA, nlag = 3)
```

```
## Response: data$TOYOTA_SA
## Input: data$OTHER_SA
## Number of inputs: 1
## Model: y ~ X + 1
## -------------------------------
## Engle-Granger Cointegration Test
## alternative: cointegrated
##
## Type 1: no trend
##     lag      EG p.value
##    3.00   -4.73    0.01
## -----
##  Type 2: linear trend
##     lag      EG p.value
##    3.00    1.21    0.10
## -----
##  Type 3: quadratic trend
##     lag      EG p.value
##   3.000  -0.305   0.100
## -----------
## Note: p.value = 0.01 means p.value <= 0.01
##     : p.value = 0.10 means p.value >= 0.10
```

```r
for (i in 2:nrow(data)){
  data$DeltaTOYOTA[1] <- 0
  data$DeltaTOYOTA[i] <- data$TOYOTA_SA[i] - data$TOYOTA_SA[i-1]
}
```

```r
data$LagDelta_1 <- dplyr::lead(data$DeltaTOYOTA, n  = 1)
data$LagDelta_2 <- dplyr::lead(data$DeltaTOYOTA, n = 2)
data$LagDelta_3 <- dplyr::lead(data$DeltaTOYOTA, n = 3)

data$LagTOYOTA <- dplyr::lead(data$TOYOTA_SA, n = 1)
```

```r
model1 <- lm(DeltaTOYOTA~LagDelta_1+LagDelta_2+LagDelta_3+LagTOYOTA, data)

summary(model1)
```

```
##
## Call:
## lm(formula = DeltaTOYOTA ~ LagDelta_1 + LagDelta_2 + LagDelta_3 +
##     LagTOYOTA, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -45616 -10266    264   9029  39114
##
## Coefficients:
##                 Estimate   Std. Error t value             Pr(>|t|)
## (Intercept) -16669.26905   8450.46348  -1.973               0.0497 *
## LagDelta_1      -0.65786      0.06342 -10.373 < 0.0000000000000002 ***
```

```
## LagDelta_2      -0.34576      0.07303  -4.735          0.00000372 ***
## LagDelta_3      -0.09677      0.06388  -1.515              0.1311
## LagTOYOTA        0.07478      0.03673   2.036              0.0428 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15450 on 244 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3119, Adjusted R-squared:  0.3006
## F-statistic: 27.65 on 4 and 244 DF,  p-value: < 0.00000000000000022
```

```r
for (i in 2:nrow(data)){
  data$DeltaOthers[1] <- 0
  data$DeltaOthers[i] <- data$OTHER_SA[i] - data$OTHER_SA[i-1]
}
```

```r
data$LagOtherDelta_1 <- dplyr::lead(data$DeltaOthers, n  = 1)
data$LagOtherDelta_2 <- dplyr::lead(data$DeltaOthers, n = 2)
data$LagOtherDelta_3 <- dplyr::lead(data$DeltaOthers, n = 3)

data$LagOthers <- dplyr::lead(data$OTHER_SA, n = 1)
```

```r
model2 <- lm(DeltaOthers~LagOtherDelta_1+LagOtherDelta_2+LagOtherDelta_3+LagOthers, data)

summary(model2)
```

```
##
## Call:
## lm(formula = DeltaOthers ~ LagOtherDelta_1 + LagOtherDelta_2 +
##     LagOtherDelta_3 + LagOthers, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -80708 -17962  -2401  16367  82867
##
## Coefficients:
##                    Estimate   Std. Error t value             Pr(>|t|)
## (Intercept)     -22953.12547  15030.68236  -1.527                0.128
## LagOtherDelta_1     -0.57608      0.06370  -9.044 < 0.0000000000000002 ***
## LagOtherDelta_2     -0.37494      0.07046  -5.322          0.000000233 ***
## LagOtherDelta_3     -0.10087      0.06423  -1.570                0.118
## LagOthers            0.05290      0.03356   1.576                0.116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26220 on 244 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.2696, Adjusted R-squared:  0.2576
## F-statistic: 22.52 on 4 and 244 DF,  p-value: 0.0000000000000007666
```

```r
model3 <- lm(data$TOYOTA_SA~data$OTHER_SA)

summary(model3)
```

```
##
## Call:
```

```
## lm(formula = data$TOYOTA_SA ~ data$OTHER_SA)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -40684 -11089    558   9837  50631
##
## Coefficients:
##                Estimate  Std. Error t value             Pr(>|t|)
## (Intercept)   28428.16819 8686.84193   3.273              0.00122 **
## data$OTHER_SA     0.45013     0.01941  23.196 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16460 on 250 degrees of freedom
## Multiple R-squared:  0.6828, Adjusted R-squared:  0.6815
## F-statistic:   538 on 1 and 250 DF,  p-value: < 0.00000000000000022
```

```
data$Errors <- model3$residuals


for (i in 2:nrow(data)){
  data$DeltaErrors[1] <- 0
  data$DeltaErrors[i] <- data$Errors[i] - data$Errors[i-1]
}
```

```
data$LagDeltaErrors_1 <- dplyr::lead(data$DeltaErrors, n  = 1)
data$LagDeltaErrors_2 <- dplyr::lead(data$DeltaErrors, n = 2)
data$LagDeltaErrors_3 <- dplyr::lead(data$DeltaErrors, n = 3)

data$LagErrors <- dplyr::lead(data$Errors, n = 1)
```

```
model4 <- lm(DeltaErrors~LagDeltaErrors_1+LagDeltaErrors_2+LagDeltaErrors_3+LagErrors,data)

summary(model4)
```

```
##
## Call:
## lm(formula = DeltaErrors ~ LagDeltaErrors_1 + LagDeltaErrors_2 +
##     LagDeltaErrors_3 + LagErrors, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -54952  -8553     85   7812  38534
##
## Coefficients:
##                   Estimate Std. Error t value             Pr(>|t|)
## (Intercept)      191.26861  816.15563   0.234               0.8149
## LagDeltaErrors_1  -0.58609    0.06343  -9.240 < 0.0000000000000002 ***
## LagDeltaErrors_2  -0.14254    0.07105  -2.006               0.0459 *
## LagDeltaErrors_3  -0.09610    0.06219  -1.545               0.1236
## LagErrors          0.28956    0.06374   4.543           0.00000874 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12870 on 244 degrees of freedom
```

```
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.2683, Adjusted R-squared:  0.2563
## F-statistic: 22.37 on 4 and 244 DF,  p-value: 0.0000000000000009446
```

```
data$LagDelta_4 <- dplyr::lead(data$DeltaTOYOTA, n  = 4)
data$LagDelta_5 <- dplyr::lead(data$DeltaTOYOTA, n  = 5)
data$LagDelta_6 <- dplyr::lead(data$DeltaTOYOTA, n  = 6)
data$LagDelta_7 <- dplyr::lead(data$DeltaTOYOTA, n  = 7)
data$LagDelta_8 <- dplyr::lead(data$DeltaTOYOTA, n  = 8)
data$LagDelta_9 <- dplyr::lead(data$DeltaTOYOTA, n  = 9)
data$LagDelta_10 <- dplyr::lead(data$DeltaTOYOTA, n  = 10)
data$LagDelta_11 <- dplyr::lead(data$DeltaTOYOTA, n  = 11)
data$LagDelta_12 <- dplyr::lead(data$DeltaTOYOTA, n  = 12)

data$ErrorCorrection <- (data$TOYOTA_SA - (0.45*(data$OTHER_SA)))
```

```
model6 <- lm(DeltaTOYOTA~LagDelta_1+LagDelta_2+LagDelta_3+LagDelta_4+LagDelta_5+LagDelta_6+LagDelta_7+La

summary(model6)
```

```
##
## Call:
## lm(formula = DeltaTOYOTA ~ LagDelta_1 + LagDelta_2 + LagDelta_3 +
##     LagDelta_4 + LagDelta_5 + LagDelta_6 + LagDelta_7 + LagDelta_8 +
##     LagDelta_9 + LagDelta_10 + LagDelta_11 + LagDelta_12, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43343  -8616   -228   8170  40343
##
## Coefficients:
##              Estimate Std. Error t value             Pr(>|t|)
## (Intercept) 518.49151  897.80065   0.578             0.564165
## LagDelta_1   -0.61038    0.06391  -9.550 < 0.0000000000000002 ***
## LagDelta_2   -0.30488    0.07550  -4.038            0.0000737 ***
## LagDelta_3   -0.24722    0.07634  -3.238             0.001383 **
## LagDelta_4   -0.23912    0.07793  -3.069             0.002412 **
## LagDelta_5   -0.18718    0.07952  -2.354             0.019426 *
## LagDelta_6   -0.10445    0.08001  -1.305             0.193076
## LagDelta_7   -0.13765    0.08012  -1.718             0.087161 .
## LagDelta_8    0.04058    0.07999   0.507             0.612411
## LagDelta_9    0.02838    0.07869   0.361             0.718745
## LagDelta_10  -0.26206    0.07675  -3.415             0.000756 ***
## LagDelta_11  -0.03910    0.07596  -0.515             0.607243
## LagDelta_12   0.23236    0.06400   3.631             0.000349 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13760 on 227 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.4638, Adjusted R-squared:  0.4355
## F-statistic: 16.37 on 12 and 227 DF,  p-value: < 0.00000000000000022
```

```
model5 <- lm(DeltaTOYOTA~LagDelta_1+LagDelta_2+LagDelta_3+LagDelta_4+LagDelta_5+LagDelta_10+LagDelta_12+
```

```r
summary(model5)
```

```
##
## Call:
## lm(formula = DeltaTOYOTA ~ LagDelta_1 + LagDelta_2 + LagDelta_3 +
##     LagDelta_4 + LagDelta_5 + LagDelta_10 + LagDelta_12 + ErrorCorrection,
##     data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -41336  -9247    639   8662  37740
##
## Coefficients:
##                    Estimate  Std. Error t value            Pr(>|t|)
## (Intercept)     -3039.85443  1922.07272  -1.582             0.11512
## LagDelta_1         -0.54566     0.06449  -8.462 0.00000000000000305 ***
## LagDelta_2         -0.24207     0.07489  -3.232             0.00141 **
## LagDelta_3         -0.19619     0.07321  -2.680             0.00789 **
## LagDelta_4         -0.18636     0.06931  -2.689             0.00769 **
## LagDelta_5         -0.09950     0.05918  -1.681             0.09406 .
## LagDelta_10        -0.26582     0.05002  -5.314 0.00000025230739777 ***
## LagDelta_12         0.25515     0.05278   4.834 0.00000244194347778 ***
## ErrorCorrection     0.12523     0.06136   2.041             0.04239 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13780 on 231 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.4535, Adjusted R-squared:  0.4345
## F-statistic: 23.96 on 8 and 231 DF,  p-value: < 0.000000000000000022
```

```r
library(Metrics)

rmse(data$DeltaTOYOTA, model5$fitted.values)
```

```
## [1] 14443.15
```

```r
rmse(data$DeltaOthers, model6$fitted.values)
```

```
## [1] 27657.25
```

```r
mae(data$DeltaTOYOTA, model5$fitted.values)
```

```
## [1] 11351.77
```

```r
mae(data$DeltaOthers, model6$fitted.values)
```

```
## [1] 21305.06
```