ARIMA Analysis

Haya Naviwala

5/12/2022

R Markdown

```
library(tidyverse)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.3.5     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   1.0.7
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#leveraging time series forecasting- future values depend on past values
#ARIMA- forecast AR(Autoregressive), I(Integrated or stationary), MA(Moving Average)
#regressed on its lagged values

dt <- read_csv("google.csv")
```

```
## Parsed with column specification:
## cols(
##   gvkey = col_double(),
##   datadate = col_double(),
##   fyearq = col_double(),
##   fqtr = col_double(),
##   tic = col_character(),
##   datafqtr = col_character(),
##   saleq = col_double()
## )
```

```
dt %>% glimpse()
```

```
## Observations: 72
## Variables: 7
## $ gvkey    <dbl> 160329, 160329, 160329, 160329, 160329, 160329, 16032...
## $ datadate <dbl> 20020331, 20020630, 20020930, 20021231, 20030331, 200...
## $ fyearq   <dbl> 2002, 2002, 2002, 2002, 2003, 2003, 2003, 2003, 2004,...
## $ fqtr     <dbl> 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...
## $ tic      <chr> "GOOGL", "GOOGL", "GOOGL", "GOOGL", "GOOGL", "GOOGL",...
## $ datafqtr <chr> "2002Q1", "2002Q2", "2002Q3", "2002Q4", "2003Q1", "20...
```

```
## $ saleq    <dbl> 42.285, 78.525, 130.787, 187.911, 248.618, 311.199, 3...
```

```
dt %>% head(2)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 7
##    gvkey datadate fyearq  fqtr tic   datafqtr saleq
##    <dbl>    <dbl>  <dbl> <dbl> <chr> <chr>    <dbl>
## 1 160329 20020331   2002     1 GOOGL 2002Q1    42.3
## 2 160329 20020630   2002     2 GOOGL 2002Q2    78.5
```

```
dt %>% tail(2)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 7
##    gvkey datadate fyearq  fqtr tic   datafqtr saleq
##    <dbl>    <dbl>  <dbl> <dbl> <chr> <chr>    <dbl>
## 1 160329 20190930   2019     3 GOOGL 2019Q3   40499
## 2 160329 20191231   2019     4 GOOGL 2019Q4   46075
```

```
#create the time series
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

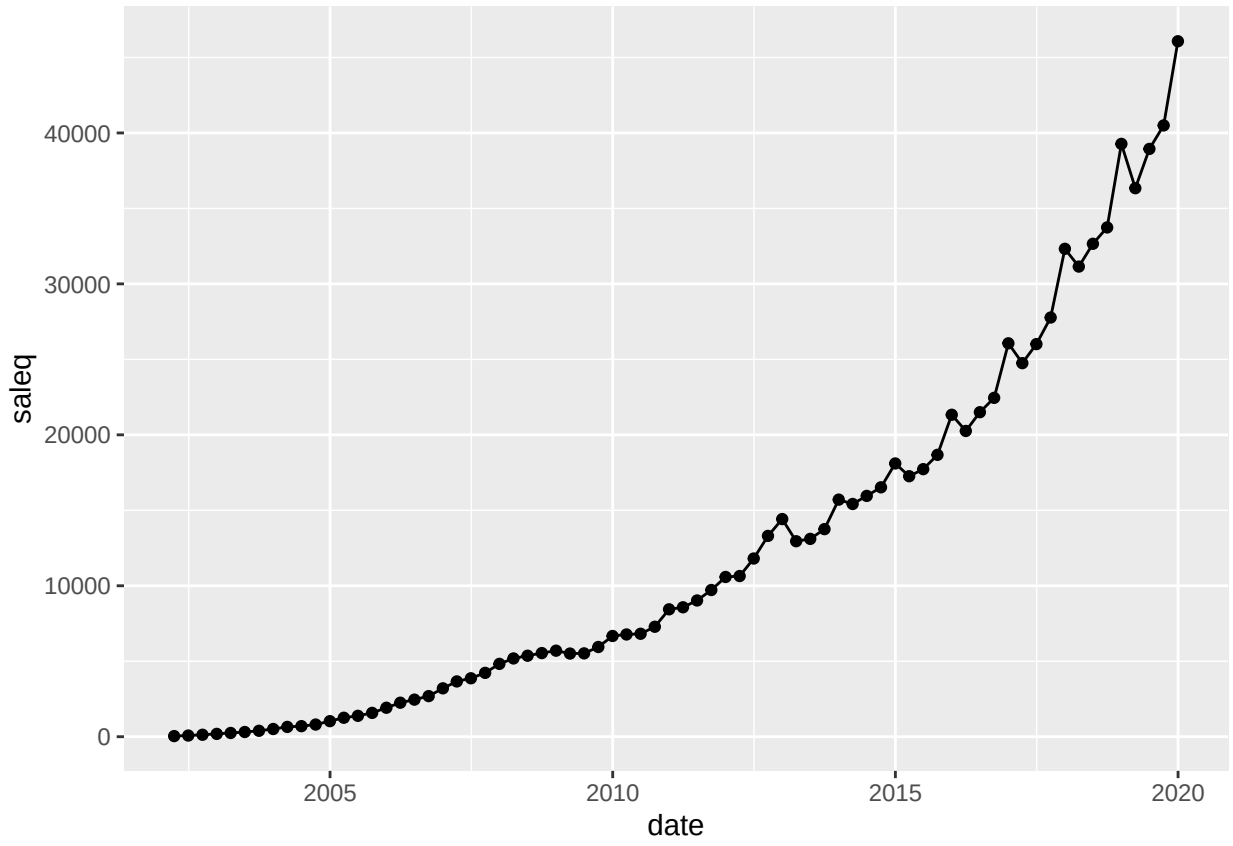
```
dt <- dt %>%
  mutate(date = yq(datafqtr) + months(3) - days(1))
dt %>% select(datafqtr, date) %>% head(2)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 2
##   datafqtr date
##   <chr>    <date>
```

```
## 1 2002Q1   2002-03-31
## 2 2002Q2   2002-06-30
```

```
#Time series trend- making a plot
dt %>%
  ggplot(aes(x=date, y=saleq)) +
  geom_point() + geom_line()
```




```
#training and validation sets
#descending order and using the first 4 observations
dtTest <- dt %>%
  arrange(desc(date)) %>%
  top_n(4, date)
dtTest
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 4 x 8
##    gvkey datadate fyearq  fqtr tic   datafqtr saleq date
##    <dbl>    <dbl>  <dbl> <dbl> <chr> <chr>    <dbl> <date>
## 1 160329 20191231   2019     4 GOOGL 2019Q4   46075 2019-12-31
## 2 160329 20190930   2019     3 GOOGL 2019Q3   40499 2019-09-30
## 3 160329 20190630   2019     2 GOOGL 2019Q2   38944 2019-06-30
```

```
## 4 160329 20190331   2019     1 GOOGL 2019Q1   36339 2019-03-31
```

```
dtTrain <- dt %>%
  filter(!(date %in% dtTest$date)) %>%
  arrange(desc(date))
head(dtTrain, 2)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 8
##    gvkey datadate fyearq  fqtr tic   datafqtr saleq date
##    <dbl>    <dbl>  <dbl> <dbl> <chr> <chr>    <dbl> <date>
## 1 160329 20181231   2018     4 GOOGL 2018Q4   39276 2018-12-31
## 2 160329 20180930   2018     3 GOOGL 2018Q3   33740 2018-09-30
```

```
tail(dtTrain, 2)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 8
##    gvkey datadate fyearq  fqtr tic   datafqtr saleq date
##    <dbl>    <dbl>  <dbl> <dbl> <chr> <chr>    <dbl> <date>
## 1 160329 20020630   2002     2 GOOGL 2002Q2    78.5 2002-06-30
## 2 160329 20020331   2002     1 GOOGL 2002Q1    42.3 2002-03-31
```

```
# auto.arima- estimate best arima model- estimate
# best # of lags for the AR and MA components of the model and the I
# takes as imput a univariate time series- meaning that our data
# should only have one variable(saleq) and the data points should be
# time ordered. need to use xts

library(xts)
```

```
## Warning: package 'xts' was built under R version 3.6.2
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 3.6.2
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
dtxts_Train <- xts(dtTrain$saleq,
                   order.by = dtTrain$date)
head(dtxts_Train, 4)
```

```
##               [,1]
## 2002-03-31  42.285
## 2002-06-30  78.525
## 2002-09-30 130.787
## 2002-12-31 187.911
```

```
dtxts_Test <- xts(dtTest$saleq,
                  order.by = dtTest$date)
head(dtxts_Test)
```

```
##             [,1]
## 2019-03-31 36339
## 2019-06-30 38944
## 2019-09-30 40499
## 2019-12-31 46075
```

```
#ARIMA model
M1 <- auto.arima(dtxts_Train)
```

```
## Error in auto.arima(dtxts_Train): could not find function "auto.arima"
```

```
summary(M1)
```

```
## Error in summary(M1): object 'M1' not found
```

```
fM1 <- forecast(M1,4)
```

```
## Error in forecast(M1, 4): could not find function "forecast"
```

```
fM1
```

```
## Error in eval(expr, envir, enclos): object 'fM1' not found
```

```
plot(fM1, xlab =  'Quarter' , ylab =  'Sales')
```

```
## Error in plot(fM1, xlab = "Quarter", ylab = "Sales"): object 'fM1' not found
```

```
accuracy(fM1, dtxts_Test)
```

```
## Error in accuracy(fM1, dtxts_Test): could not find function "accuracy"
```

```
dtxts <- xts(dt$saleq, order.by = dt$date)
head(dtxts,2)
```

```
##             [,1]
## 2002-03-31 42.285
## 2002-06-30 78.525
```

```
tail(dtxts,2)
```

```
##             [,1]
## 2019-09-30 40499
## 2019-12-31 46075
```

```
M2 <- auto.arima(dtxts)
```

```
## Error in auto.arima(dtxts): could not find function "auto.arima"
```

```
summary(M2)
```

```
## Error in summary(M2): object 'M2' not found
```

```
fM2 <- forecast(M2, 2)
```

```
## Error in forecast(M2, 2): could not find function "forecast"
```

```
fM2
```

```
## Error in eval(expr, envir, enclos): object 'fM2' not found
```

```
plot(fM2)
```

```
## Error in plot(fM2): object 'fM2' not found
```

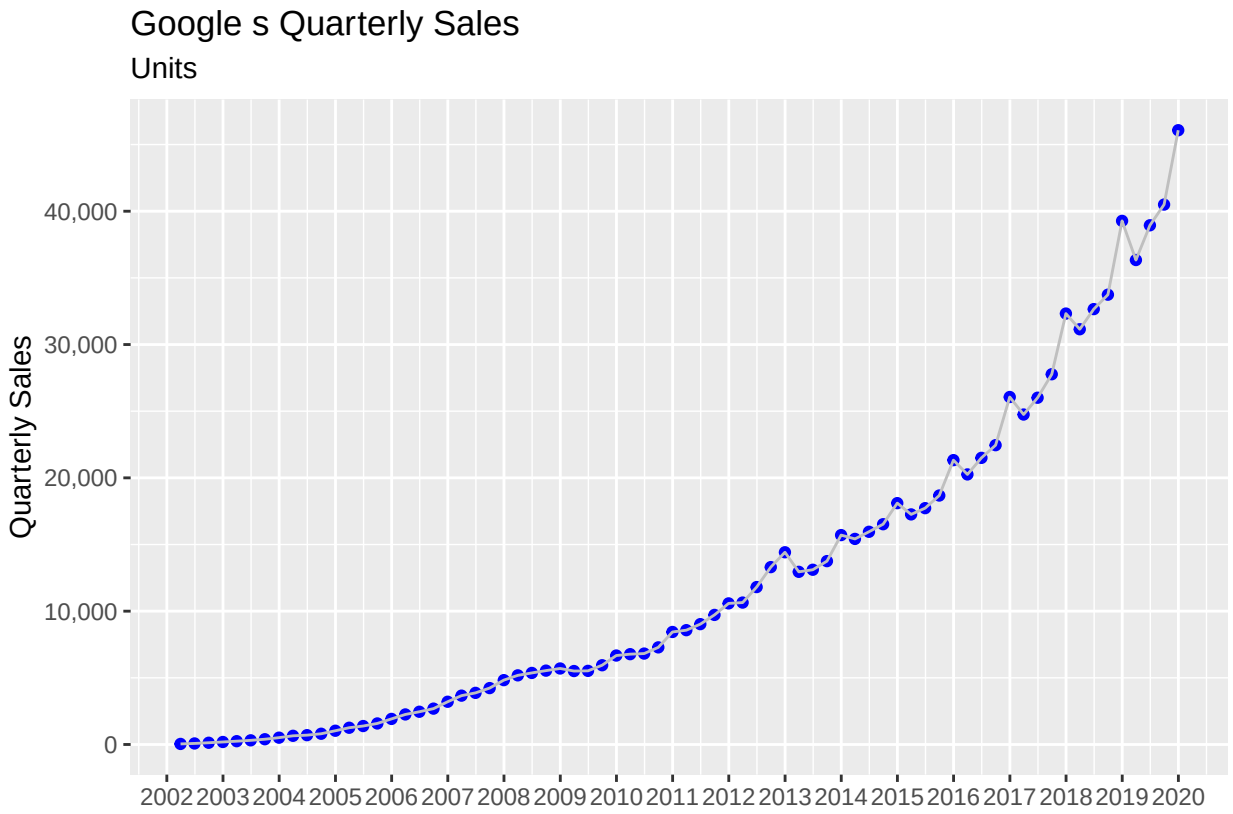
```
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
dt %>%
  ggplot(aes(x=date, y=saleq)) +
  geom_point(colour = "blue") +
  geom_line(colour = "grey75") +
  scale_x_date(breaks = date_breaks(width = "1 year"),
               labels = date_format( '%Y' )) +
  scale_y_continuous(labels = comma_format()) +
  xlab("") + ylab("Quarterly Sales") +
  ggtitle("Google s Quarterly Sales", subtitle = "Units")
```




```
##dt0 <- read_csv("finStatements4All_FY2009-20.csv")
names(dt0)
```

```
## Error in eval(expr, envir, enclos): object 'dt0' not found
```

```
# The ticker symbol for Tesla is TSLA

dt0 %>% filter(tic=="TSLA") %>% select(fyear, tic, conm, naics)
```

```
## Error in eval(lhs, parent, parent): object 'dt0' not found
```

```
dt1 <- dt0 %>%
  filter(naics==336111 & fyear>2009 & fyear < 2020) %>%
  select(fyear, tic, conm, sale, at, oibdp, naics) %>%
  mutate(ROA = oibdp/at, PM = oibdp/sale, TATO=sale/at)
```

```
## Error in eval(lhs, parent, parent): object 'dt0' not found
```

```
dt1 %>% select(sale, at, oibdp,ROA, PM, TATO) %>% summary()
```

```
## Error in eval(lhs, parent, parent): object 'dt1' not found
```

```
dt2 <- dt1 %>% filter(sale>=100)
```

```
## Error in eval(lhs, parent, parent): object 'dt1' not found
```

```
dt2 %>% select(sale, at, oibdp,ROA, PM, TATO) %>% summary()
```

```
## Error in eval(lhs, parent, parent): object 'dt2' not found
```

```
dt3_indROA <- dt2 %>%
  group_by(fyear) %>%
  summarise(ind_minROA = min(ROA),
```

```
            ind_q1ROA  = quantile(ROA, 0.25),
            ind_medROA = median(ROA),
            ind_q3ROA  = quantile(ROA, 0.75),
            ind_maxROA = max(ROA))
```

```
## Error in eval(lhs, parent, parent): object 'dt2' not found
```

```
dt3_indROA
```

```
## Error in eval(expr, envir, enclos): object 'dt3_indROA' not found
```

```
dt3_indPM <- dt2  %>% group_by(fyear) %>%
  summarise(ind_minPM = min(PM),
            ind_q1PM  = quantile(PM, 0.25),ind_medPM = median(PM),
            ind_q3PM  = quantile(PM, 0.75),ind_maxPM = max(PM))
```

```
## Error in eval(lhs, parent, parent): object 'dt2' not found
```

```
dt3_indPM
```

```
## Error in eval(expr, envir, enclos): object 'dt3_indPM' not found
```

```
dt3_indTATO <- dt2 %>% group_by(fyear) %>%
  summarise(ind_minTATO = min(TATO),
            ind_q1TATO  = quantile(TATO, 0.25),ind_medTATO = median(TATO),
            ind_q3TATO  = quantile(TATO, 0.75), ind_maxTATO = max(TATO))
```

```
## Error in eval(lhs, parent, parent): object 'dt2' not found
```

```
dt3_indTATO
```

```
## Error in eval(expr, envir, enclos): object 'dt3_indTATO' not found
```

```
dt4Tesla <- dt1 %>% filter(tic=="TSLA") %>%
  select(fyear, tic, ROA, PM, TATO)
```

```
## Error in eval(lhs, parent, parent): object 'dt1' not found
```

```
dt4Tesla
```

```
## Error in eval(expr, envir, enclos): object 'dt4Tesla' not found
```

```
dt5 <- bind_cols(
  dt3_indROA %>% select(fyear, ind_q1ROA, ind_q3ROA),
  dt3_indPM %>% select(ind_q1PM, ind_q3PM),
  dt3_indTATO %>% select(ind_q1TATO, ind_q3TATO),
  dt4Tesla %>% select(tic, ROA, PM, TATO))
```

```
## Error in eval(lhs, parent, parent): object 'dt3_indROA' not found
```

```
dt5
```

```
## Error in eval(expr, envir, enclos): object 'dt5' not found
```

Including Plots

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.