

Capstone Time-Series

Claudio Rivas Z.

April 7, 2019

Contents

Introduction	2
Analysis	3
Data Inspection	3
Data Visualization	4
Data Preparation	5
Analyzing Time Series	6
Classical Seasonal Decomposition by Moving Averages	6
Seasonal Decomposition of Time Series by Loess	7
Forecasting	9
Exponential Smoothing Forecasts	9
ARIMA Modelling of Time Series	10
Forecasting using Holt-Winters objects	11
Results	12
Conclusions	14

Introduction

This report is part of the **Data Science Professional Certificate** program, specifically the final course “Capstone”.

For this particular project, I chose a topic that interested me before enrolling in this program, in this project we will forecast the stock prices of a certain company, we will analyze **time series**.

To make it more interesting, we will select two similar companies and compare the results in order to determine in which it is convenient to invest. The **data** can be downloaded from the yahoo financial website <https://finance.yahoo.com/>.

During this work, we will analyze the data, complete the missing non-working days with interpolated values, decompose the series using loess and moving averages, use different forecasting methods and select the best one.

Finally, we will compare the forecast obtained with the real values.

Analysis

Data Inspection

We will analyze the data of American Airline (AAL) and Delta Air Line (DAL)

```
str(AAL)
## 'data.frame': 563 obs. of 2 variables:
## $ Date : Date, format: "2017-01-03" "2017-01-04" ...
## $ Close: num 46.3 46.7 45.9 46.2 47.1 ...
```

```
str(DAL)
## 'data.frame': 563 obs. of 2 variables:
## $ Date : Date, format: "2017-01-03" "2017-01-04" ...
## $ Close: num 49.5 50.7 49.7 49.7 49.7 ...
```

In both cases we have 563 observations of 2 variables, one of them is the date of registration and the other is the closing value. The observations are filtered from the beginning of 2017 until May 2019.

```
head(AAL)
##           Date Close
## 1 2017-01-03 46.30
## 2 2017-01-04 46.70
## 3 2017-01-05 45.89
## 4 2017-01-06 46.21
## 5 2017-01-09 47.08
## 6 2017-01-10 48.48
```

```
head(DAL)
##           Date Close
## 1 2017-01-03 49.48
## 2 2017-01-04 50.70
## 3 2017-01-05 49.73
## 4 2017-01-06 49.68
## 5 2017-01-09 49.69
## 6 2017-01-10 50.93
```

We also note that certain days have no observations, which corresponds to the holidays and non-working days of the stock exchange, for example, on days 7 and 8 of January 2017.

```
summary(AAL)
##           Date             Close
##  Min.   :2017-01-03   Min.   :29.72
##  1st Qu.:2017-07-25   1st Qu.:38.26
##  Median :2018-02-14   Median :44.74
##  Mean   :2018-02-13   Mean   :43.89
##  3rd Qu.:2018-09-05   3rd Qu.:49.40
##  Max.   :2019-03-29   Max.   :58.47
```

The observations begin on 2017-01-03 and end on 2019-03-29

The minimum closing price was 29.72 and the maximum was 58.47

The average is 43.89

```

summary(DAL)
##          Date            Close
##  Min.   :2017-01-03   Min.   :44.03
##  1st Qu.:2017-07-25  1st Qu.:49.59
##  Median :2018-02-14  Median :51.84
##  Mean   :2018-02-13  Mean   :51.95
##  3rd Qu.:2018-09-05  3rd Qu.:54.41
##  Max.   :2019-03-29  Max.   :60.71

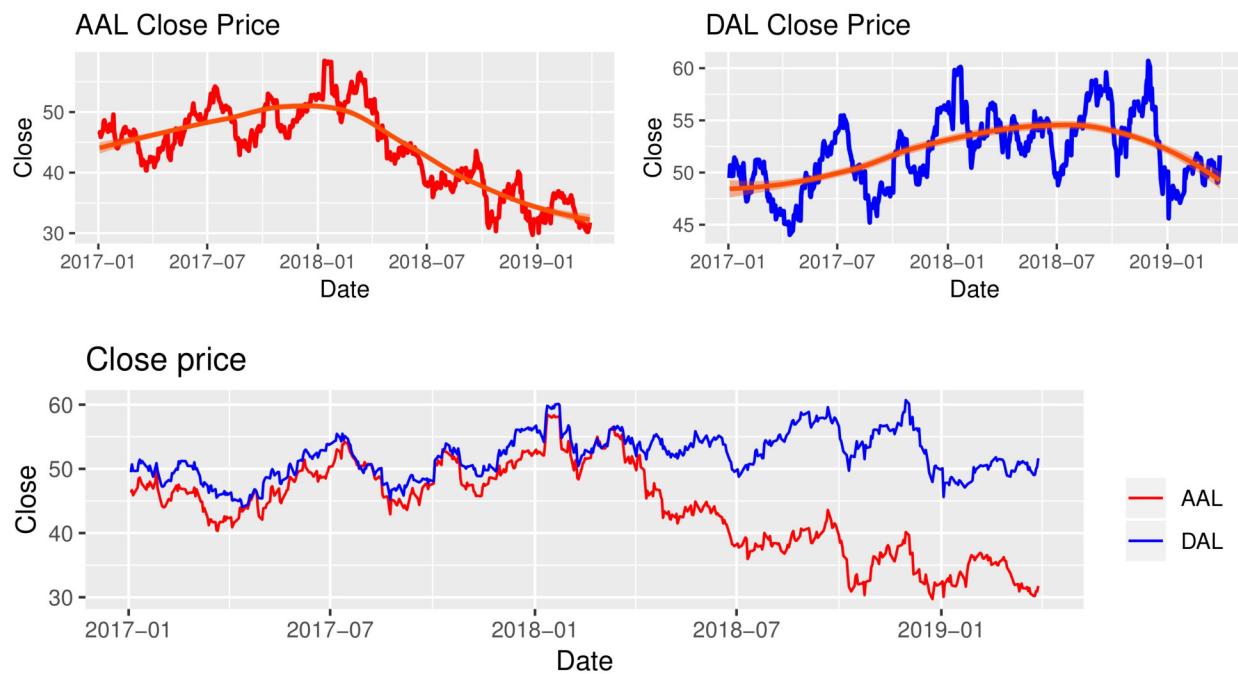
```

The observations begin on 2017-01-03 and end on 2019-03-29

The minimum closing price was 44.03 and the maximum was 60.71

The average is 51.95

Data Visualization



At first glance, we noticed that the values were quite similar until 2018, and since that year American Airlines began to fall, while Delta Airline remained near to its average.

Data Preparation

The first thing we must do with the data is to convert them into time series using the “`ts()`” function, but for this we need to have all the observations of each day, so we will complete the original data set with the missed days, with 818 observations finally remaining since January from 2017 to May 2019 (one per day). This will leave us the empty observations that we will have to complete with the interpolated values using the “`na.interp()`” function.

```
head(AAL, 8)
##           Date Close
## 1 2017-01-03 46.30
## 2 2017-01-04 46.70
## 3 2017-01-05 45.89
## 4 2017-01-06 46.21
## 5 2017-01-07     NA
## 6 2017-01-08     NA
## 7 2017-01-09 47.08
## 8 2017-01-10 48.48

head(AAL, 8)
## Time Series:
## Start = c(2017, 3)
## End   = c(2017, 10)
## Frequency = 365
## [1] 46.30 46.70 45.89 46.21 47.56 47.49 47.08 48.48
```

Note that once the data frame has been converted to a time series, some things changed with the “`head()`” function, first, we show the start and end dates of the 8 observations shown, starting on day 3 of 2017 and ending on day 10 of the same year. Note also that we did not mention the month, this is because the day makes mention of the day elapsed of the year, and as the first observations are from the month of January, these days coincide in this case.

```
summary(AAL)
##      Min. 1st Qu. Median    Mean 3rd Qu.    Max.
## 29.72   38.27  44.90  43.99  49.37  60.23
```

The observations begin on 2017, 3 and end on 2019, 90 (remember that the days are elapsed of the year)

The average is 43.99

```
summary(DAL)
##      Min. 1st Qu. Median    Mean 3rd Qu.    Max.
## 44.03   49.71  51.99  52.08  54.50  62.09
```

The observations begin on 2017, 3 and end on 2019, 90 (remember that the days are elapsed of the year)

The average is 52.08

Analyzing Time Series

Classical Seasonal Decomposition by Moving Averages

Decompose a time series into seasonal, trend and irregular components using moving averages. Deals with additive or multiplicative seasonal component.

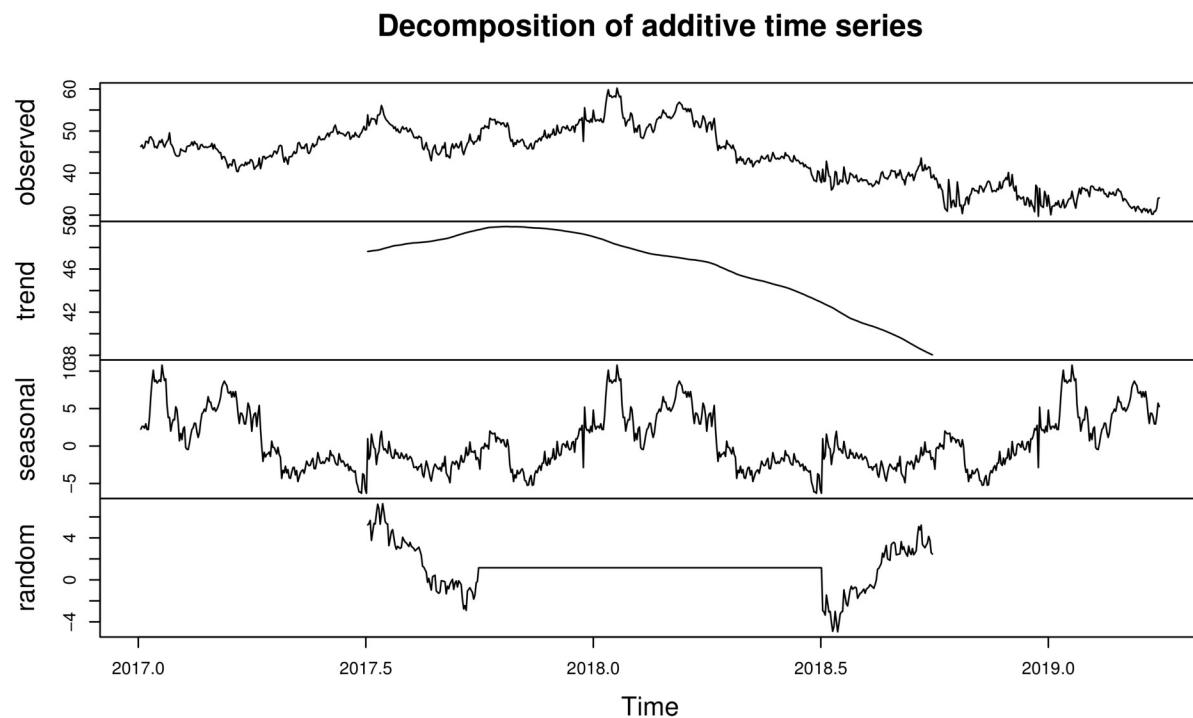
The additive model used is

$$Y[t] = T[t] + S[t] + e[t]$$

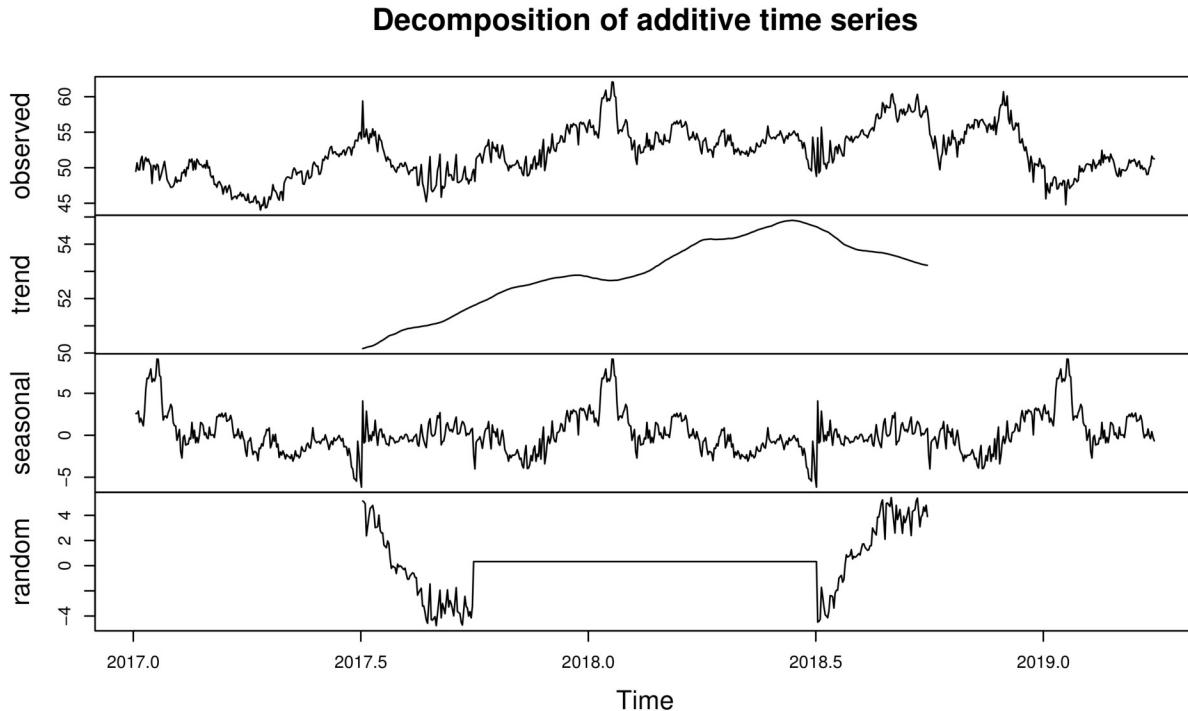
The multiplicative model used is:

$$Y[t] = T[t] * S[t] * e[t]$$

```
plot(AAL.decompose)
```



```
plot(DAL.decompose)
```

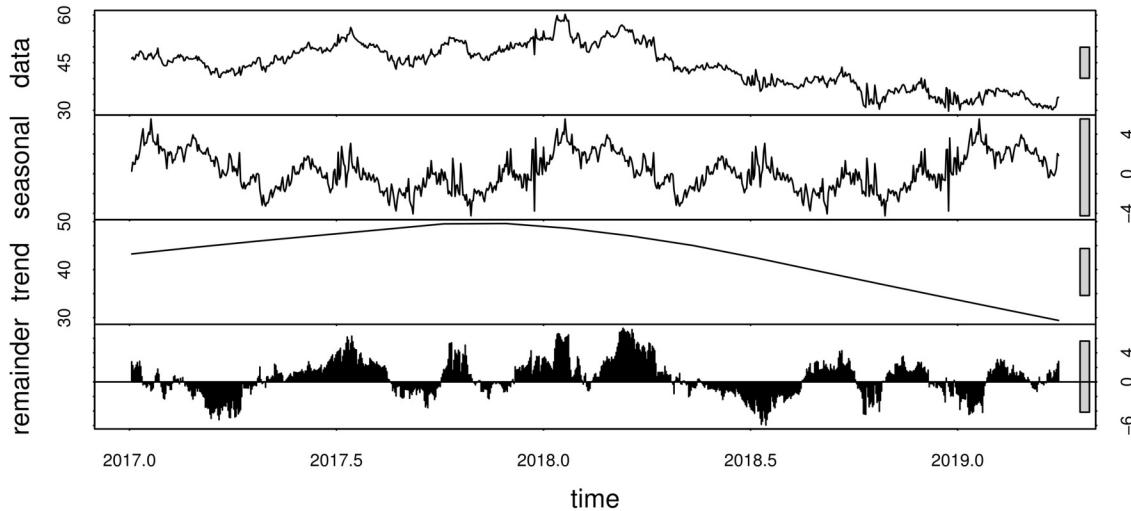


Seasonal Decomposition of Time Series by Loess

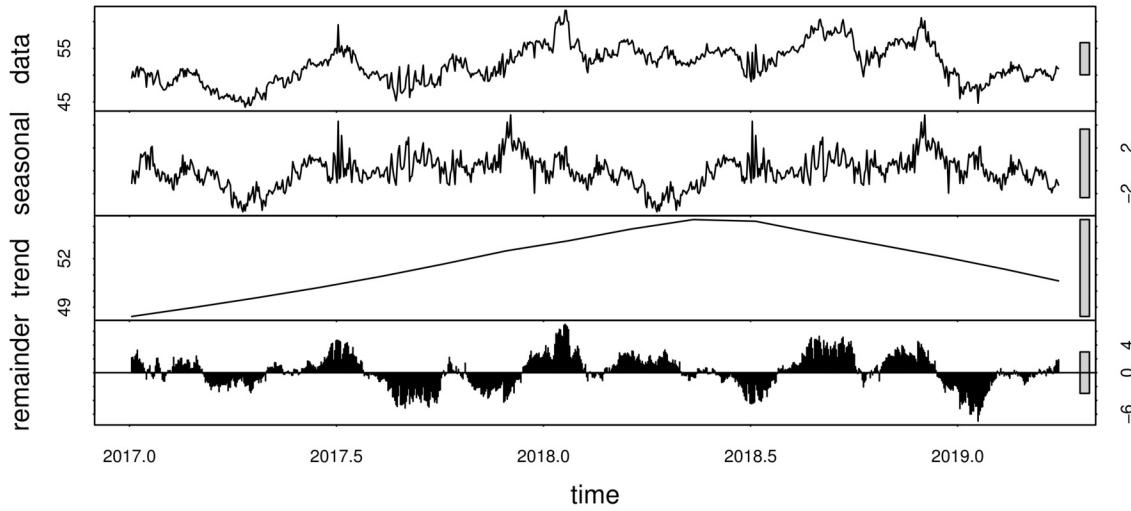
Decompose a time series into Seasonal, Trend and irregular components using loess, acronym STL.

The seasonal component is found by loess smoothing the seasonal sub-series (the series of all January values, ...); if s.window = “periodic” smoothing is effectively replaced by taking the mean. The seasonal values are removed, and the remainder smoothed to find the trend. The overall level is removed from the seasonal component and added to the trend component. This process is iterated a few times. The remainder component is the residuals from the seasonal plus trend fit.

```
plot(AAL.stl)
```



```
plot(DAL.stl)
```



In both decompositions “decompose()” and “stl()”, a seasonality is appreciated for both AAL and DAL, increasing the values at the beginning of the year and presenting high volatility in the middle of the year. We also see that the AAL trend is downward, while DAL decreases to a lesser extent

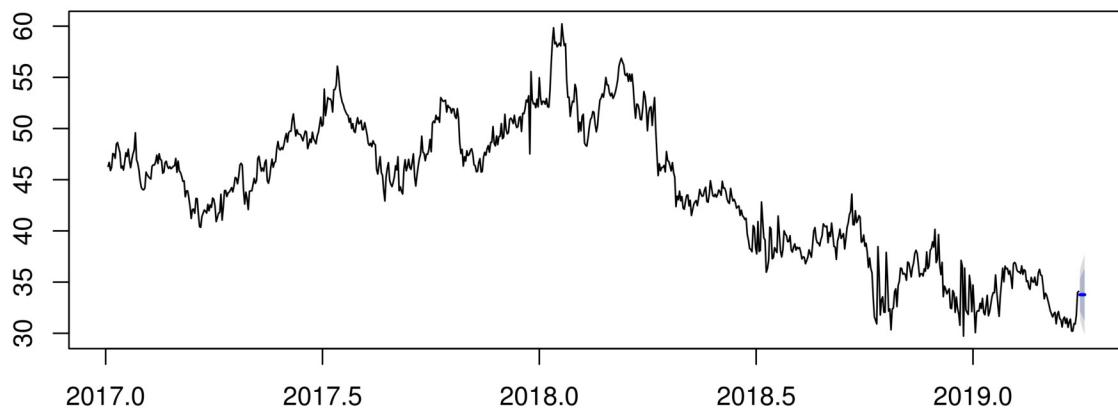
Forecasting

Exponential Smoothing Forecasts

Returns forecasts and other information for exponential smoothing forecasts applied to a time series.

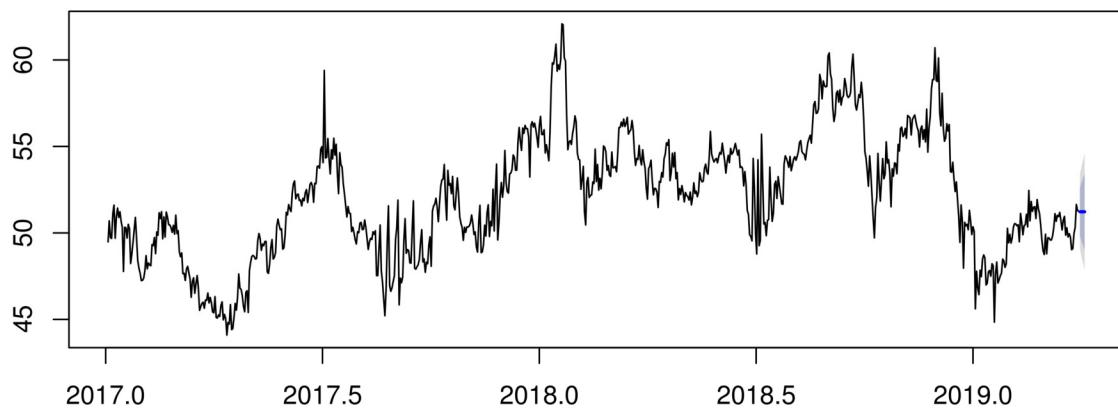
```
plot(AAL.ses)
```

Forecasts from Simple exponential smoothing



```
plot(DAL.ses)
```

Forecasts from Simple exponential smoothing

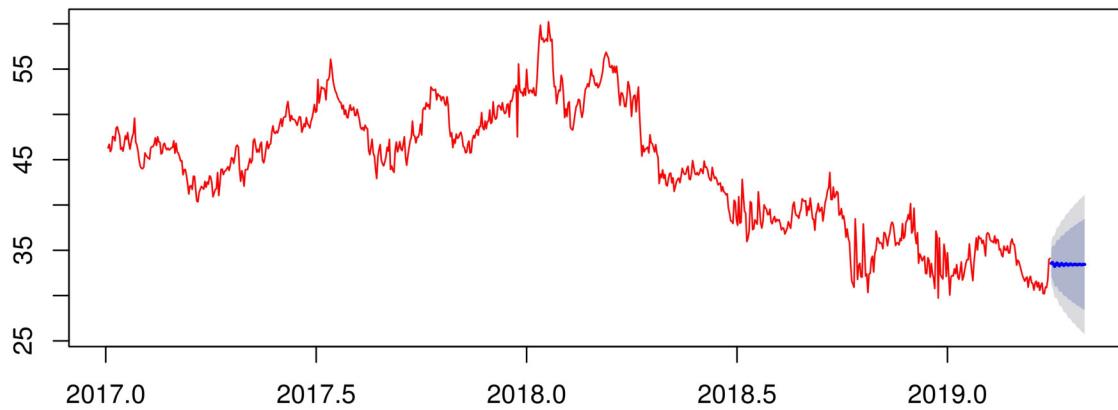


ARIMA Modelling of Time Series

Fit an ARIMA model to a univariate time series. Autoregressive integrated moving average (ARIMA)

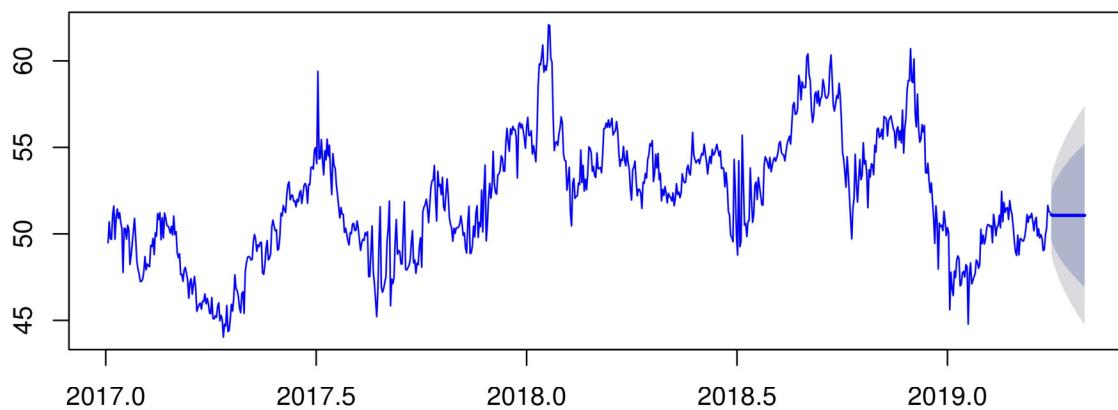
```
plot(AAL.fore.ari, col = "red")
```

Forecasts from ARIMA(3,1,5)



```
plot(DAL.fore.ari, col = "blue")
```

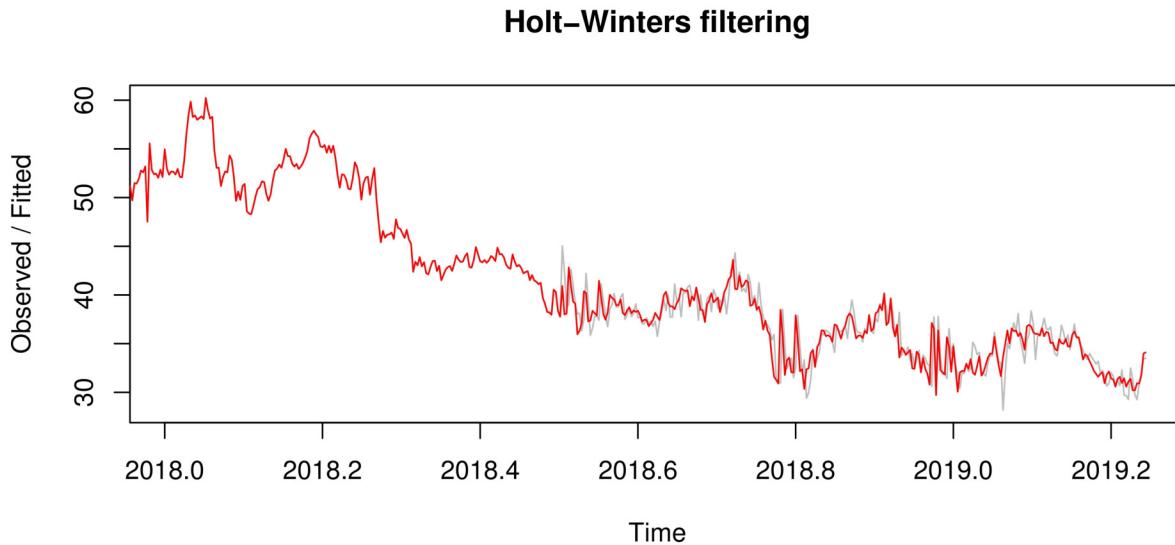
Forecasts from ARIMA(0,1,2)



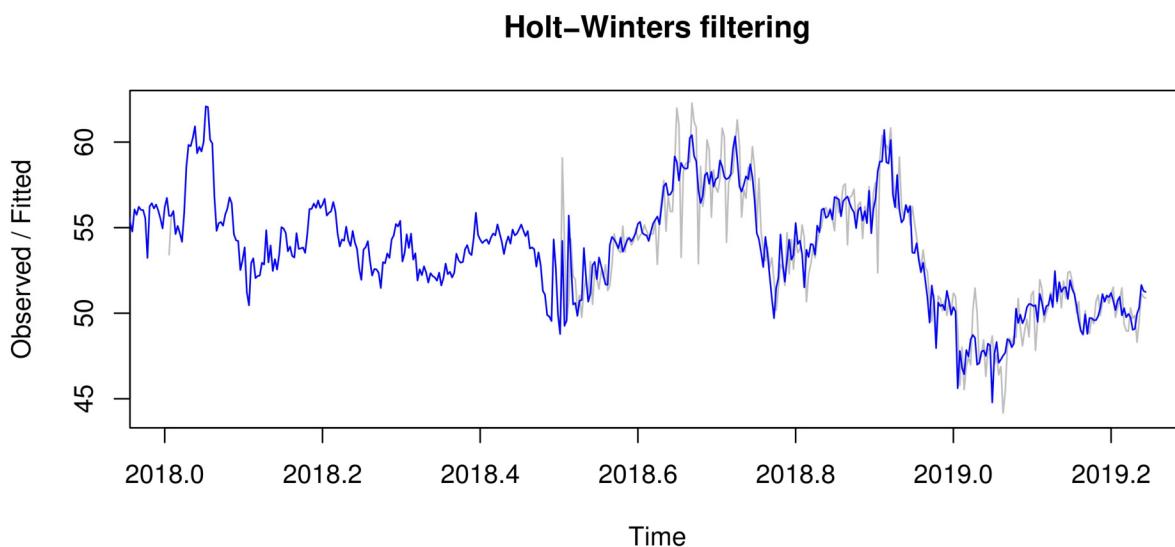
Forecasting using Holt-Winters objects

Returns forecasts and other information for univariate Holt-Winters time series models. This function calls predict.HoltWinters and constructs an object of class "forecast" from the results.

```
plot(AAL.hw, col = "red", col.predicted = "grey")
```



```
plot(DAL.hw, col = "blue", col.predicted = "grey")
```

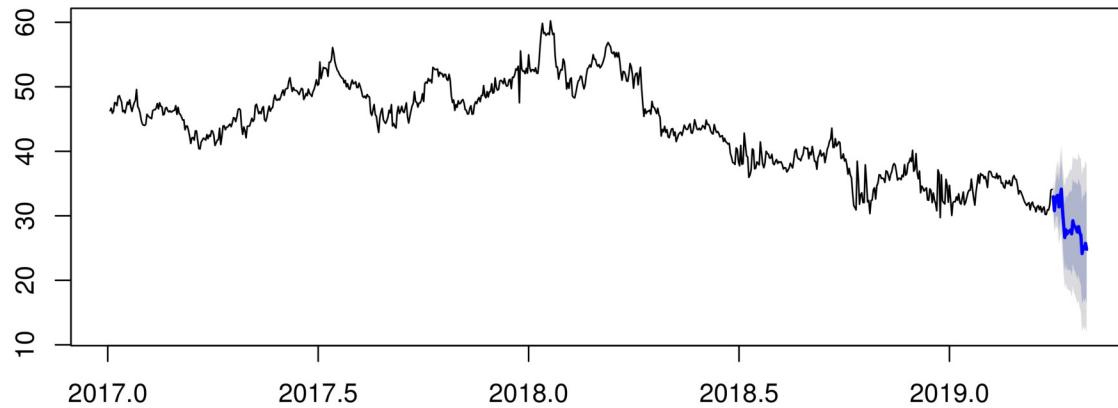


Results

After reviewing the previous methods, the best method was HoltWinters. Here we can clearly appreciate the continuity of the original curve

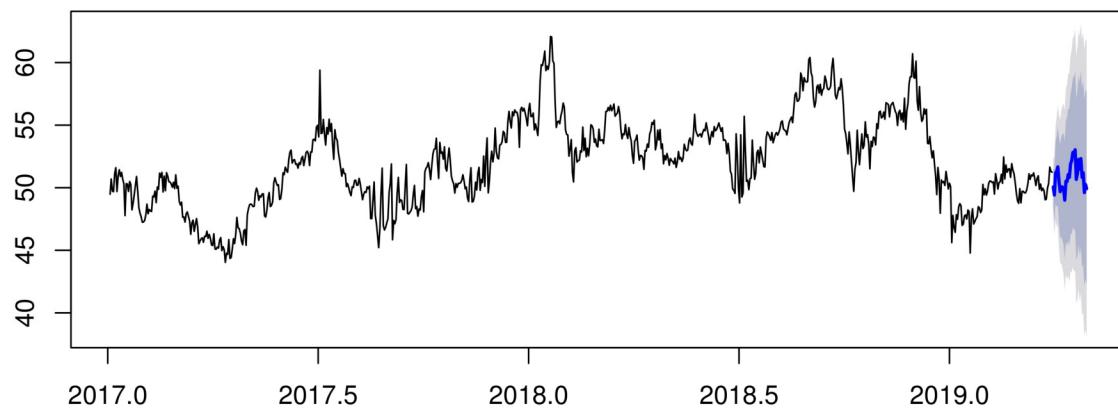
```
plot(AAL.fore)
```

Forecasts from HoltWinters



```
plot(DAL.fore)
```

Forecasts from HoltWinters

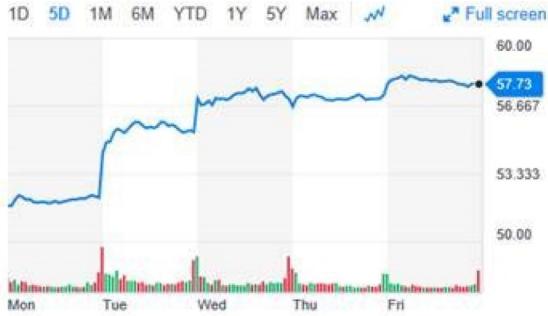


Now, according to the forecast, we can recommend buying shares of the DAL stock exchange instead of AAL, since the value of the first will remain in the future, while the second will be in constant decline.

Now we will do a check somewhat different from the traditional method of separating data in training and testing. We will check the forecast directly with the real values of this first week of April. Unfortunately at the time of making this report, the data was not yet available to be downloaded, but we could see it, so we will check against that image

```
DAL.fore$lower[1:5,]
##          80%      95%
## [1,] 48.60320 47.82940
## [2,] 47.38104 46.30762
## [3,] 48.70602 47.40000
## [4,] 48.47794 46.97488
## [5,] 48.50621 46.82912

DAL.fore$upper[1:5,]
##          80%      95%
## [1,] 51.52669 52.30048
## [2,] 51.43652 52.50994
## [3,] 53.64030 54.94632
## [4,] 54.15660 55.65966
## [5,] 54.84241 56.51950
```



We can verify that effectively the forecast approaches the upper limit of 95%.

Conclusions

We can conclude that it is not advisable to buy shares of American Airlines, since they are going down.

The shares of Delta Air Lines on the other hand remain close to their average, and rising at certain times, so our recommendation would be to invest in this company.

Try different methods of forecasting served us first to learn, and secondly to choose the one that best suited what we were looking for and to make a decision based on the method of our choice.

Note: the incorporation of almost all the code within the report was intentional with the purpose of showing the complete procedure of analysis of this work.