exercise 2

ajay parihar

April 26, 2019

Exercise 2

The data set Houses shows the yearly number of family houses sold in Belgium (Houses) and the total price of transactions (Houseprices) from 1973 up to 2017. First, create the average house price by dividing the total price of transactions by the yearly number of family houses sold. This new variable will be analysed. Split the data in a training set up to 2010 and a test set from 2011 up to 2017. Use the training set for estimation of the methods/models, and use the test set for assessing the forecast accuracy. In each step of the exercise, discuss your results and explain your choices. Use additional tables and graphs wherever they clarify your answer.

```
library(fpp2)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: forecast
```

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unable to identify current timezone 'C':
## please set environment variable 'TZ'
```

```
## Loading required package: fma
```

```
## Loading required package: expsmooth
```

```
library(readxl)
library(portes)
```

```
## Loading required package: parallel
```

```
#install.packages("lmtest")
library(lmtest)
```

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

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

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

```
setwd("C:/Users/aparihar/Downloads")

# read the data
houses <- read_excel("DataSets.xlsx", sheet = "Houses")
houses$averageprice<-houses$Houseprices/houses$Houses
str(houses)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    45 obs. of  4 variables:
##  $ Date        : num  1973 1974 1975 1976 1977 ...
##  $ Houses      : num  53866 53129 51561 55978 58166 ...
##  $ Houseprices : num  7.81e+08 8.81e+08 9.68e+08 1.32e+09 1.66e+09 ...
##  $ averageprice: num  14498 16573 18765 23574 28522 ...
```

```
###create time series
rsvhouse <- ts(houses[,4], frequency = 1, start = 1973)
rsv1house <- window(rsvhouse, end=2009)
rsv2house <- window(rsvhouse, start=2010)
h=length(rsv2house)
h
```

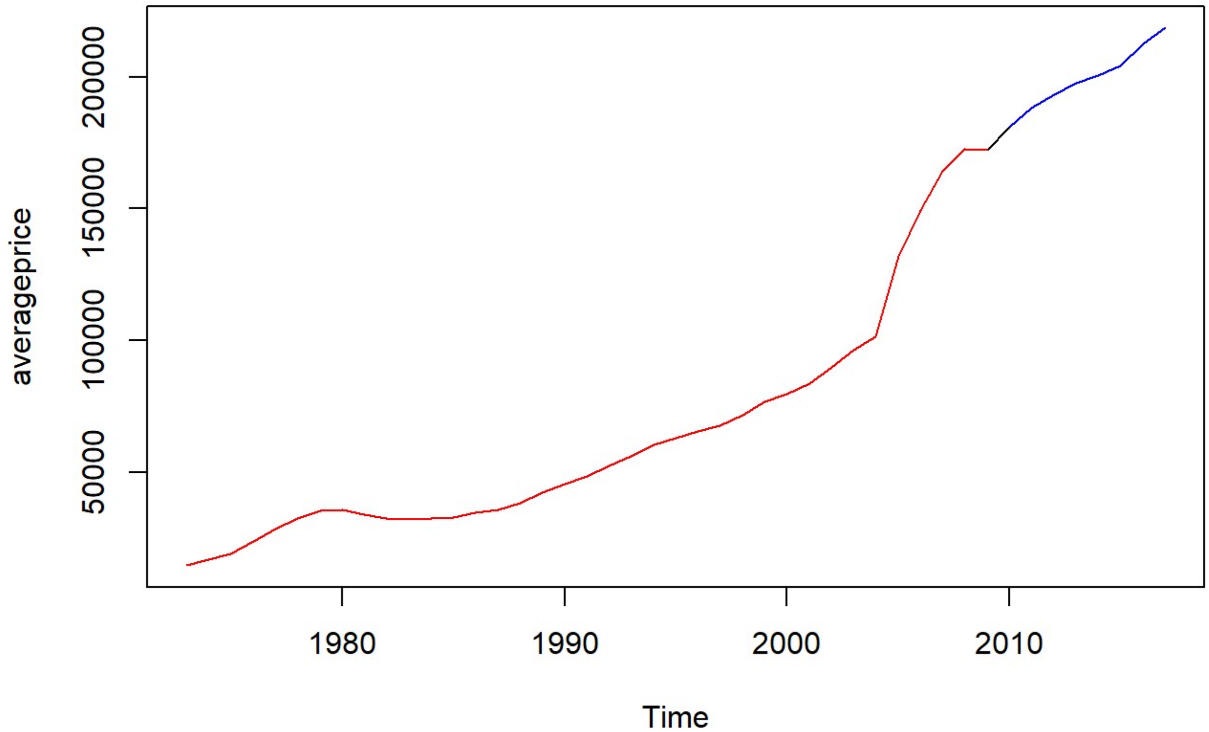
```
## [1] 8
```

exercise2.1

Explore the data using relevant graphs.

here we plot the average price graph for the time period, we don't see any seasonality in the time series

```
# Plot the data
plot(rsvhouse)
lines(rsv1house, col="red")
lines(rsv2house, col="blue")
```



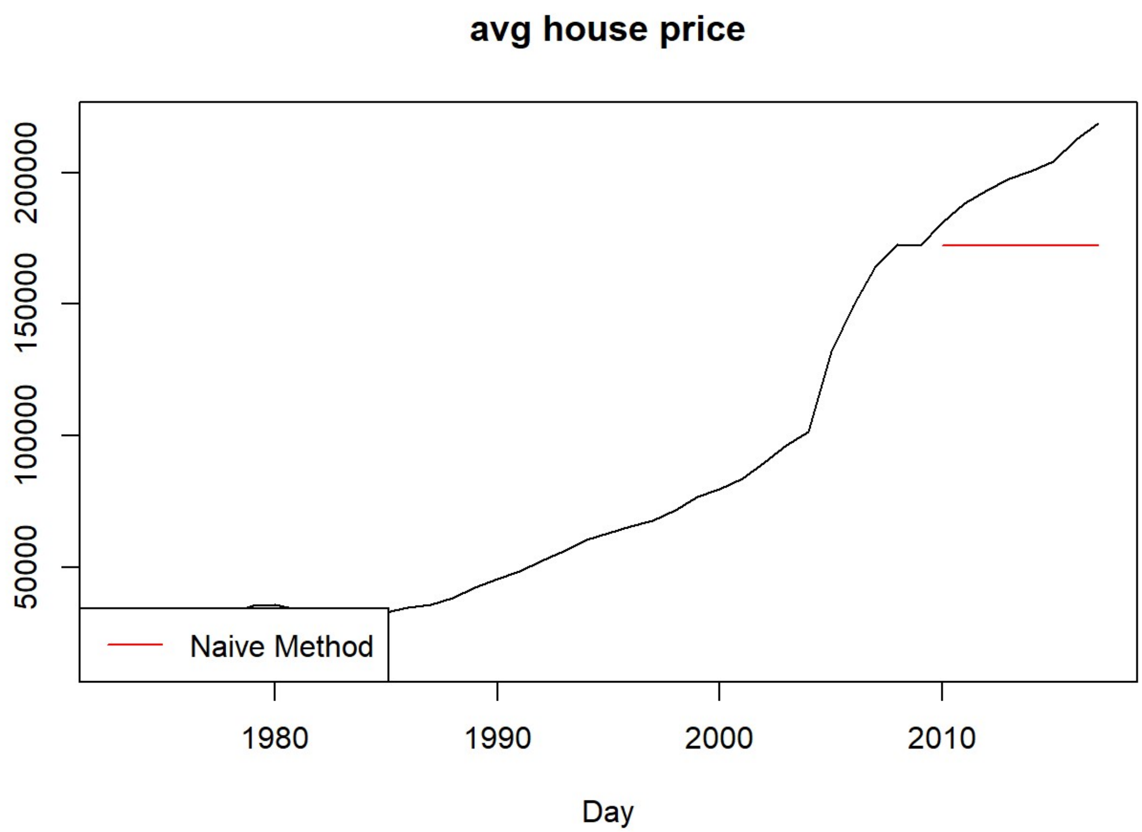
Exercise2.2

Create forecasts using the most appropriate naive method. Check the residual diagnostics and the forecast accuracy. We run the naive method, a plot is created. Looking at the residuls we don't see any white noise

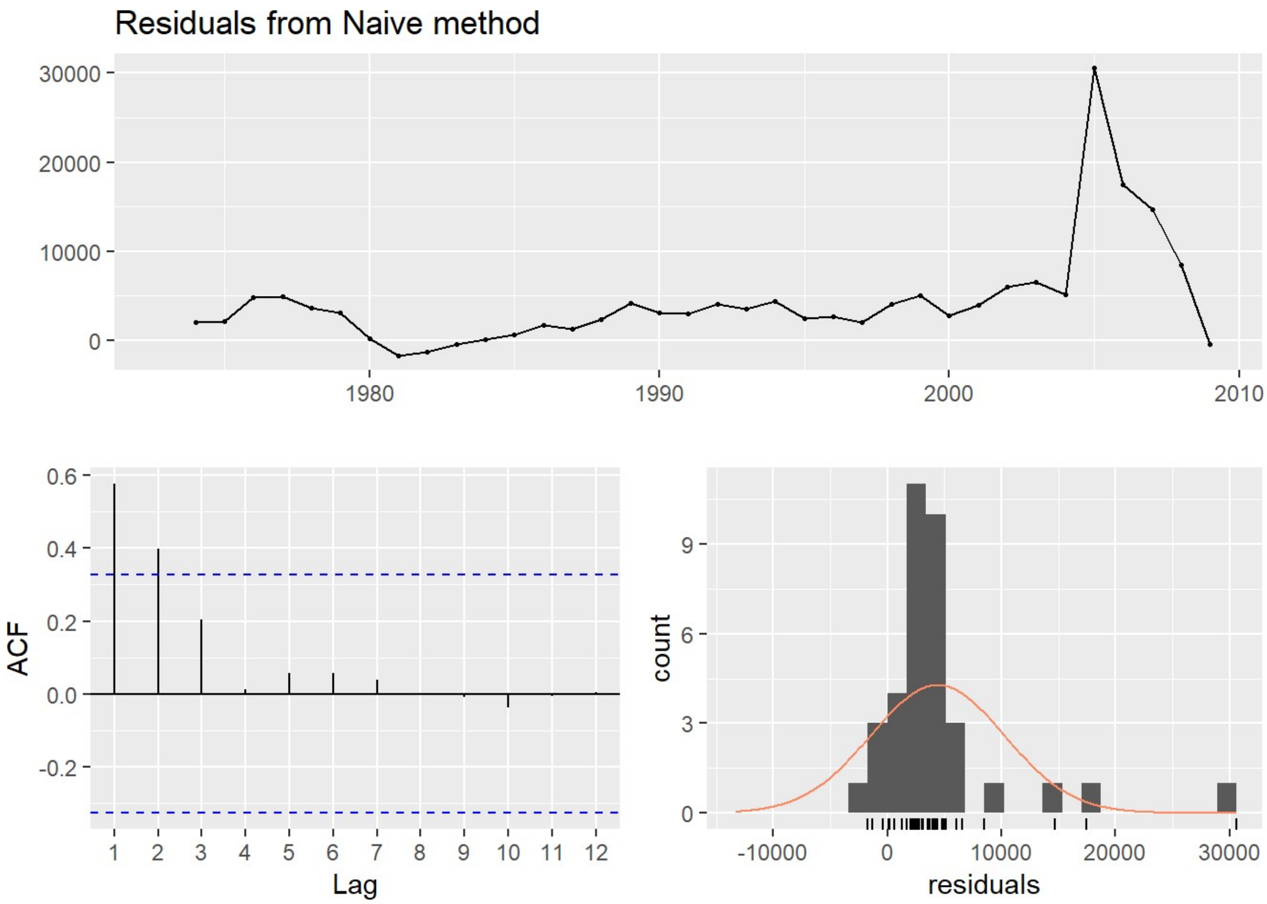
```
##here we run naive method
f2 <- naive(rsv1house, h=h)
accuracy(f2,rsv2house)
```

```
##                      ME      RMSE       MAE       MPE     MAPE     MASE
## Training set  4382.756  7267.137  4596.867  6.454523  7.04680 1.000000
## Test set     27326.193 29673.556 27326.193 13.400057 13.40006 5.944525
##                   ACF1 Theil's U
## Training set 0.5758397        NA
## Test set     0.5619291  5.365213
```

```
##lets plot the grpah
plot(rsvhouse,main="avg house price", ylab="",xlab="Day")
lines(f2$mean,col=2)
legend("bottomleft",lty=1,col=2,legend="Naive Method")
```




```
#now we check residuals
checkresiduals(f2)
```

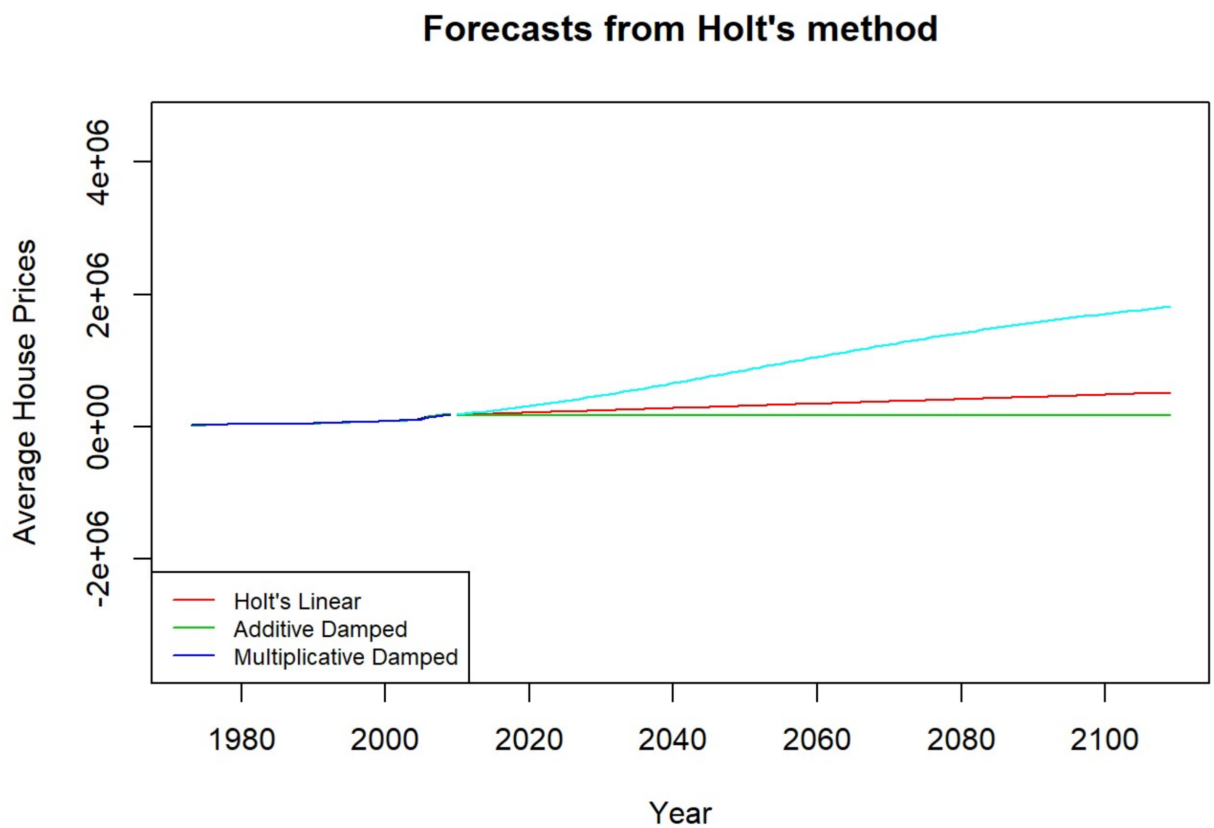



```
##
##  Ljung-Box test
##
## data:  Residuals from Naive method
## Q* = 21.428, df = 7.4, p-value = 0.004166
##
## Model df: 0.   Total lags used: 7.4
```

Exercise 2.3

Q)Generate forecasts using the relevant exponential smoothing methods. Check the residual diagnostics and the forecast accuracy A) h2 damped is the best

```
h1 <- holt(rsv1house, h=100)

h2 <- holt(rsv1house, h=100, damped=TRUE)

h3 <- holt(rsv1house, h=100, exponential=TRUE, damped=TRUE)

plot(h1, type="l", ylab="Average House Prices",

xlab="Year", fcol="white", shadecols="white")
lines(fitted(h1), col=2)

lines(fitted(h2), col=3)

lines(fitted(h3), col=4)



lines(h1$mean, col=2, type="l")

lines(h2$mean, col=3, type="l")

lines(h3$mean, col=5, type="l")

legend("bottomleft", lty=1, col=c(2,3,4,5),c("Holt's Linear", "Additive Damped", "Mult
iplicative Damped"), cex=0.75)
```




```
accuracy(h1, rsv2house)[,c(2,3,5,6)]
```

```
##                   RMSE       MAE     MAPE      MASE
## Training set  5003.296  2442.777 3.581799 0.5314005
## Test set     12687.380 12043.597 5.942417 2.6199576
```

```
accuracy(h2,rsv2house)[,c(2,3,5,6)]
```

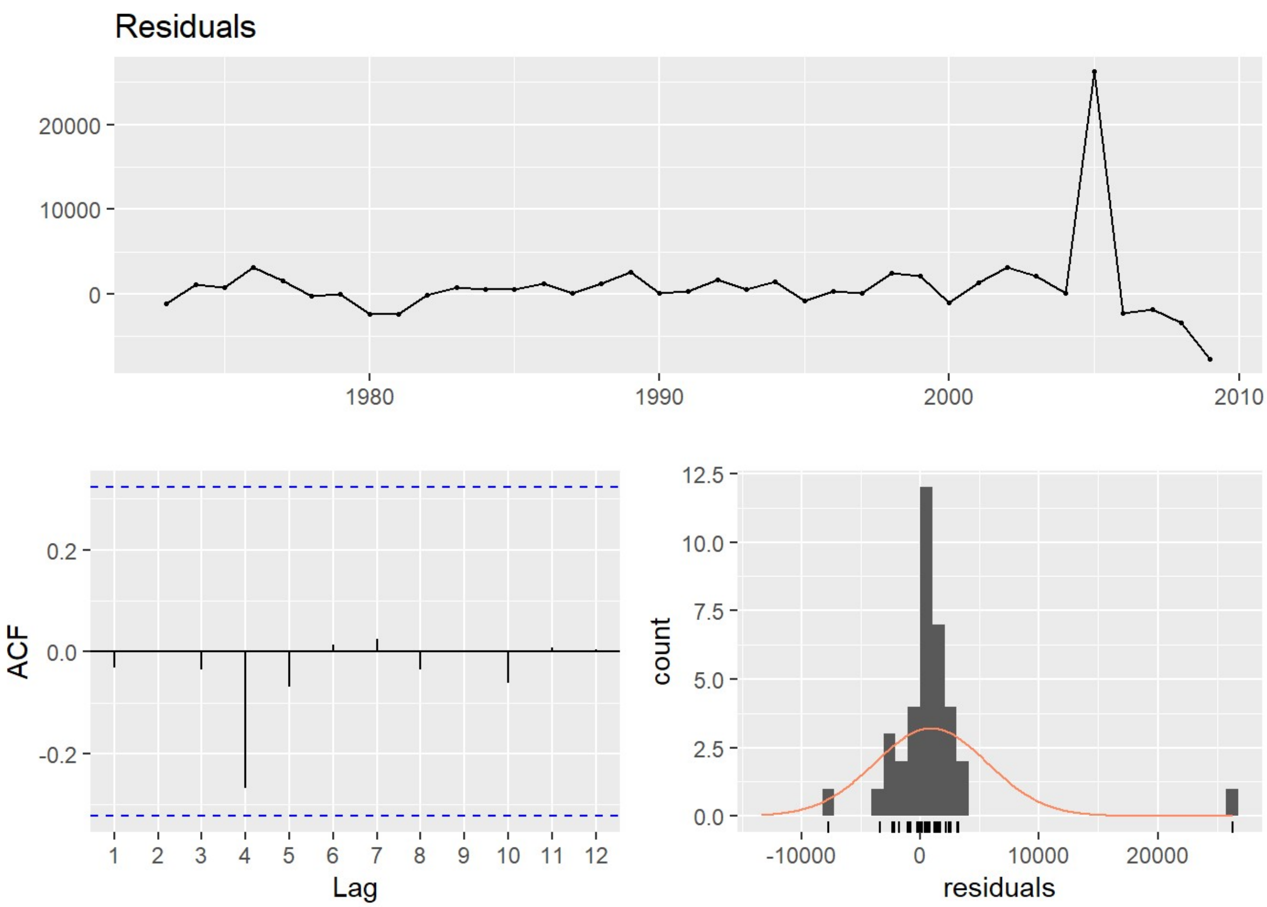
```
##                   RMSE       MAE      MAPE      MASE
## Training set  4772.581  2142.455  3.383765 0.4660685
## Test set     28869.254 26462.948 12.966957 5.7567355
```

```
accuracy(h3, rsv2house)[,c(2,3,5,6)]
```

```
##                   RMSE       MAE      MAPE      MASE
## Training set  5173.595  3063.102  5.530347 0.6663457
## Test set     29497.602 24418.665 11.805170 5.3120233
```

```
res <- residuals(h2)

checkresiduals(res)
```

```
## Warning in modeldf.default(object): Could not find appropriate degrees of
## freedom for this model.
```



Exercise 2.4

Generate forecasts using ETS. First select the appropriate model(s) yourself and discuss their performance. Compare these models with the results of the automated ETS procedure. Check the residual diagnostics and the forecast accuracy for the various ETS models you've considered

A)Since the data is non seasonal we will use the below model, f1 and f6 are the best model

```
e1 <- ets(rsv1house,"AAN")
e2 <- ets(rsv1house,"MNN")
e3 <- ets(rsv1house,"ANN")
e4 <- ets(rsv1house,"MAN")
e5 <- ets(rsv1house,"MMN")
e6 <- ets(rsv1house, model = "AAN", damped = TRUE)
e7 <- ets(rsv1house, model = "MAN", damped = TRUE)
e8 <- ets(rsv1house, model = "MMN", damped = TRUE)
auto_ets = ets(rsv1house)

auto_ets = ets(rsv1house)
f1 = forecast(e1,h=h)
f2 = forecast(e2,h=h)
f3 = forecast(e3,h=h)
f4 = forecast(e4,h=h)
f5 = forecast(e5,h=h)
f6 = forecast(e6,h=h)
f7 = forecast(e7,h=h)
f8 = forecast(e8,h=h)

accuracy(f1,rsv2house)[,c(2,3,5,6)]
```

```
##                   RMSE       MAE      MAPE      MASE
## Training set  4772.581  2142.455  3.383765 0.4660685
## Test set     28869.254 26462.948 12.966957 5.7567355
```

```
accuracy(f2,rsv2house)[,c(2,3,5,6)]
```

```
##                   RMSE       MAE      MAPE     MASE
## Training set  7463.985  4814.675  9.213487 1.047382
## Test set     29673.519 27326.153 13.400037 5.944517
```

```
accuracy(f3,rsv2house)[,c(2,3,5,6)]
```

```
##                   RMSE       MAE      MAPE     MASE
## Training set  7463.985  4814.675  9.213487 1.047382
## Test set     29673.519 27326.153 13.400037 5.944517
```

```
accuracy(f4,rsv2house)[,c(2,3,5,6)]
```

```
##                   RMSE       MAE      MAPE      MASE
## Training set  5086.783  2444.953  3.346675 0.5318738
## Test set     26642.894 24613.080 12.076095 5.3543163
```

```
accuracy(f5,rsv2house)[,c(2,3,5,6)]
```

```
##                    RMSE       MAE      MAPE      MASE
## Training set  5060.616  2456.199  3.586214 0.5343203
## Test set     29380.598 26953.457 13.208980 5.8634406
```

```
accuracy(f6,rsv2house)[,c(2,3,5,6)]
```

```
##                    RMSE       MAE      MAPE      MASE
## Training set  4772.581  2142.455  3.383765 0.4660685
## Test set     28869.254 26462.948 12.966957 5.7567355
```

```
accuracy(f7,rsv2house)[,c(2,3,5,6)]
```

```
##                    RMSE       MAE      MAPE      MASE
## Training set  4931.137  2386.179  3.490045 0.5190882
## Test set     29640.550 27185.229 13.322043 5.9138603
```

```
accuracy(f8,rsv2house)[,c(2,3,5,6)]
```

```
##                    RMSE       MAE      MAPE      MASE
## Training set  5060.616  2456.199  3.586214 0.5343203
## Test set     29380.598 26953.457 13.208980 5.8634406
```

Exercise 2.5

Generate forecasts using ARIMA. First select the appropriate model(s) yourself and discuss their performance. Compare these models with the results of the auto.arima procedure. Check the residual diagnostics and the forecast accuracy for the various ARIMA models you've considered.

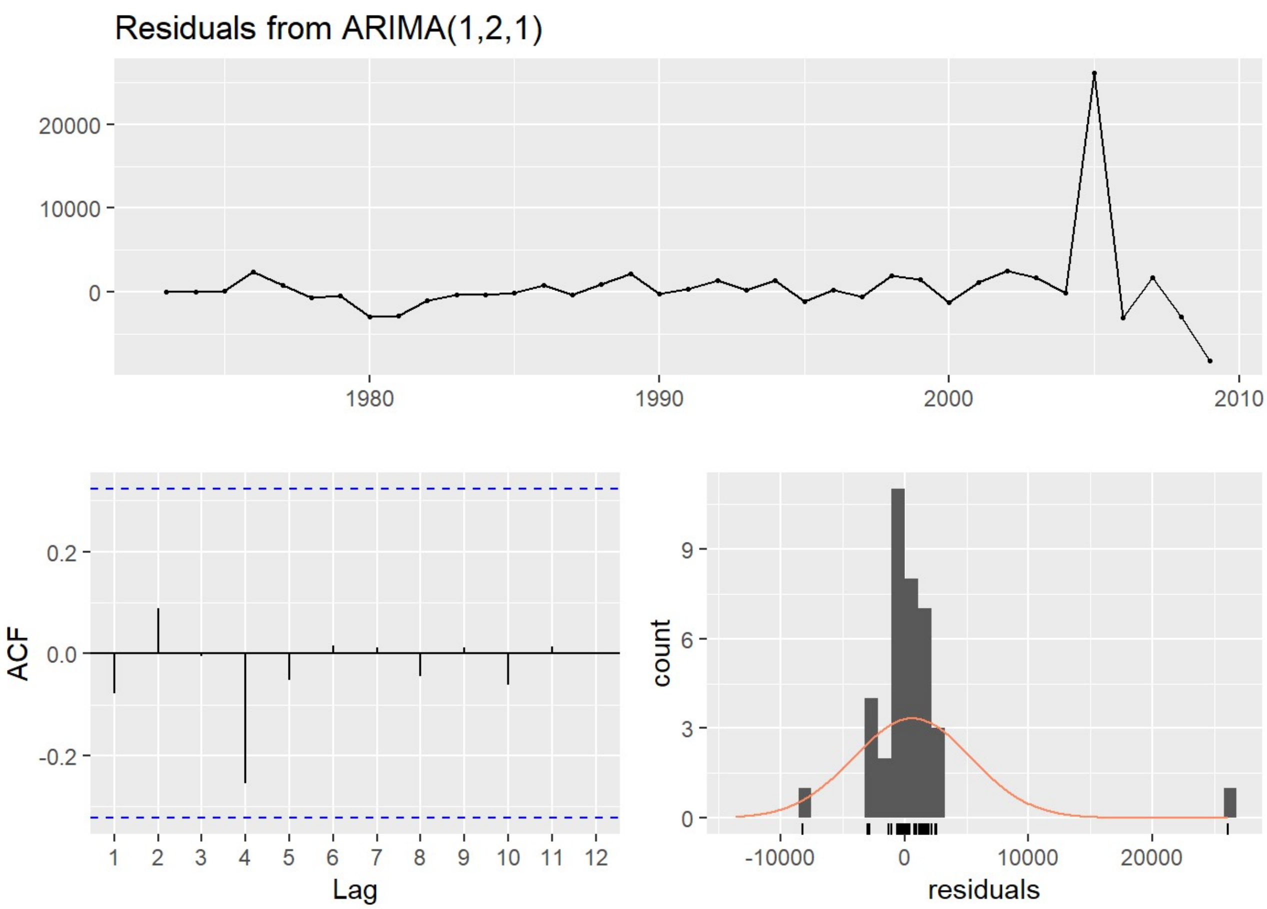
```
auto_arima <-auto.arima(rsv1house, seasonal=FALSE, allowdrift = TRUE)
summary(auto_arima)
```

```
## Series: rsv1house
## ARIMA(1,2,1)
##
## Coefficients:
##          ar1      ma1
##       0.5627  -0.9344
## s.e.  0.2356   0.1414
##
## sigma^2 estimated as 25068090:  log likelihood=-347.26
## AIC=700.51   AICc=701.29   BIC=705.18
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 592.3046 4728.427 1998.166 0.6352888 2.727353 0.4346799
##                     ACF1
## Training set -0.07803069
```

```
f1<-forecast(auto_arima,h=h)
accuracy(f1,rsv2house)[,c(2,3,5,6)]
```

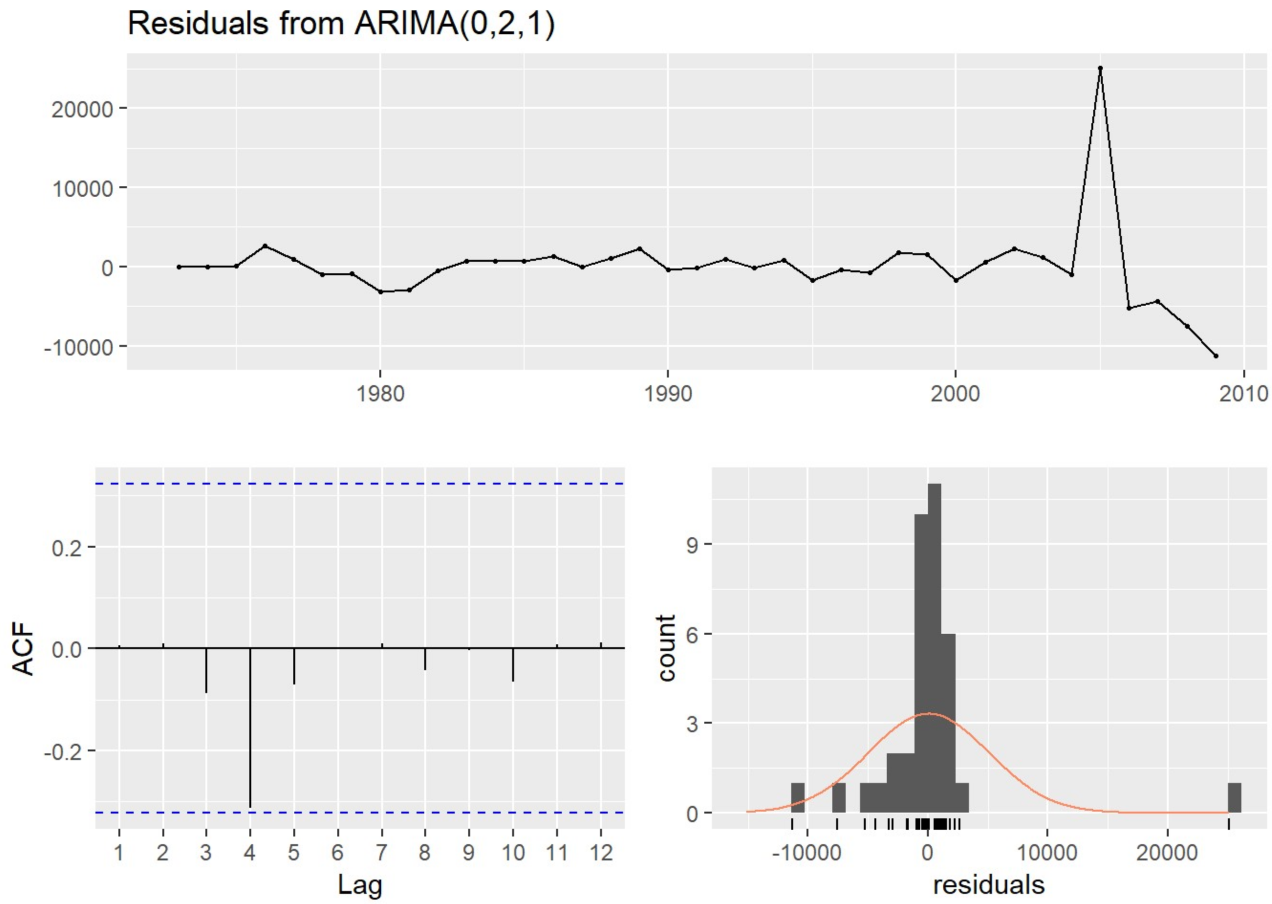
```
##                  RMSE      MAE     MAPE      MASE
## Training set 4728.427 1998.166 2.727353 0.4346799
## Test set     8669.620 8502.100 4.273283 1.8495422
```

```
checkresiduals(auto_arima)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,2,1)
## Q* = 3.5595, df = 5.4, p-value = 0.6664
##
## Model df: 2.   Total lags used: 7.4
```

```
best_model <- Arima(rsv1house, order=c(0,2,1))
checkresiduals(best_model)
```

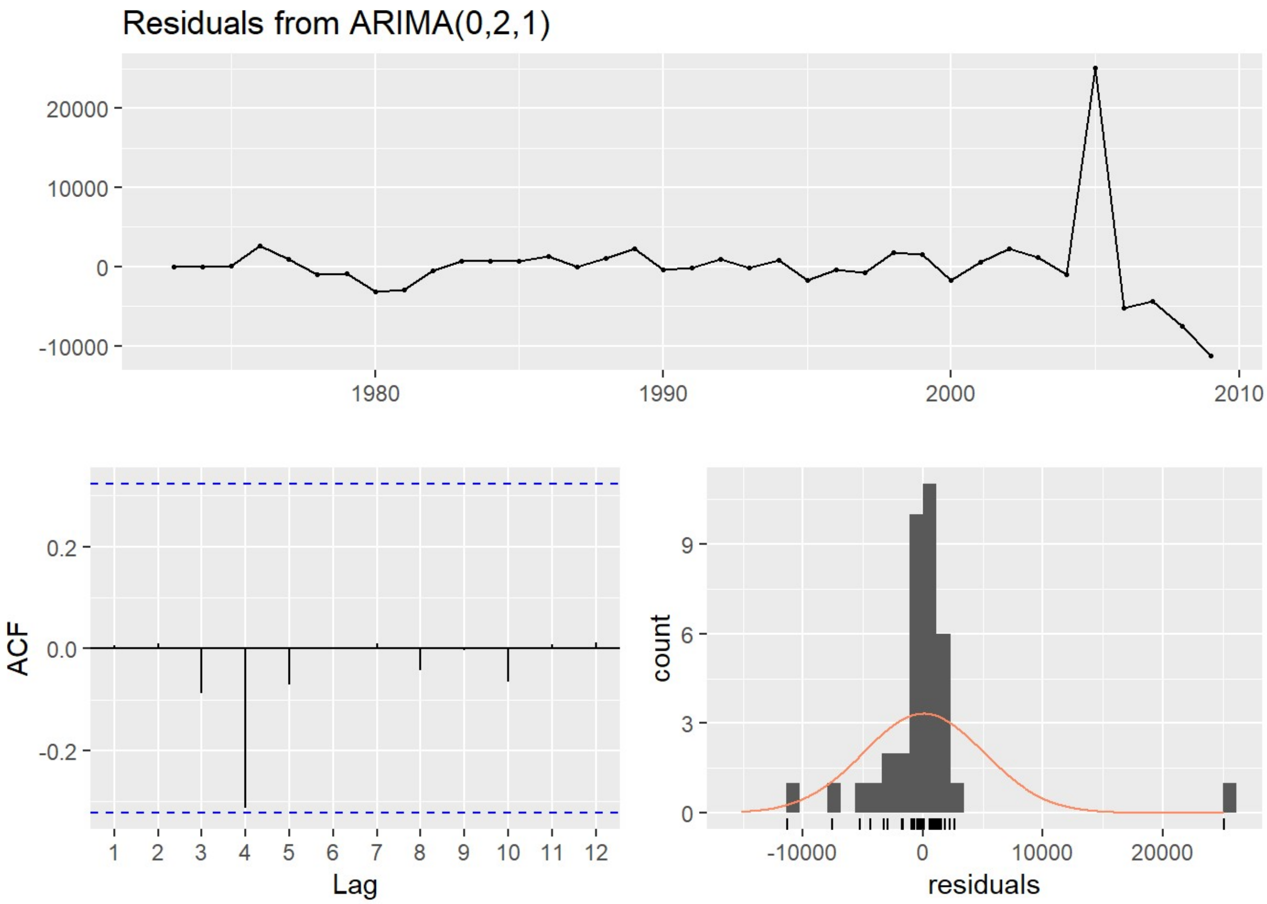



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,2,1)
## Q* = 4.8235, df = 6.4, p-value = 0.6147
##
## Model df: 1.   Total lags used: 7.4
```

```
f2<-forecast(best_model,h=h)
accuracy(f2,rsv2house)[,c(2,3,5,6)]
```

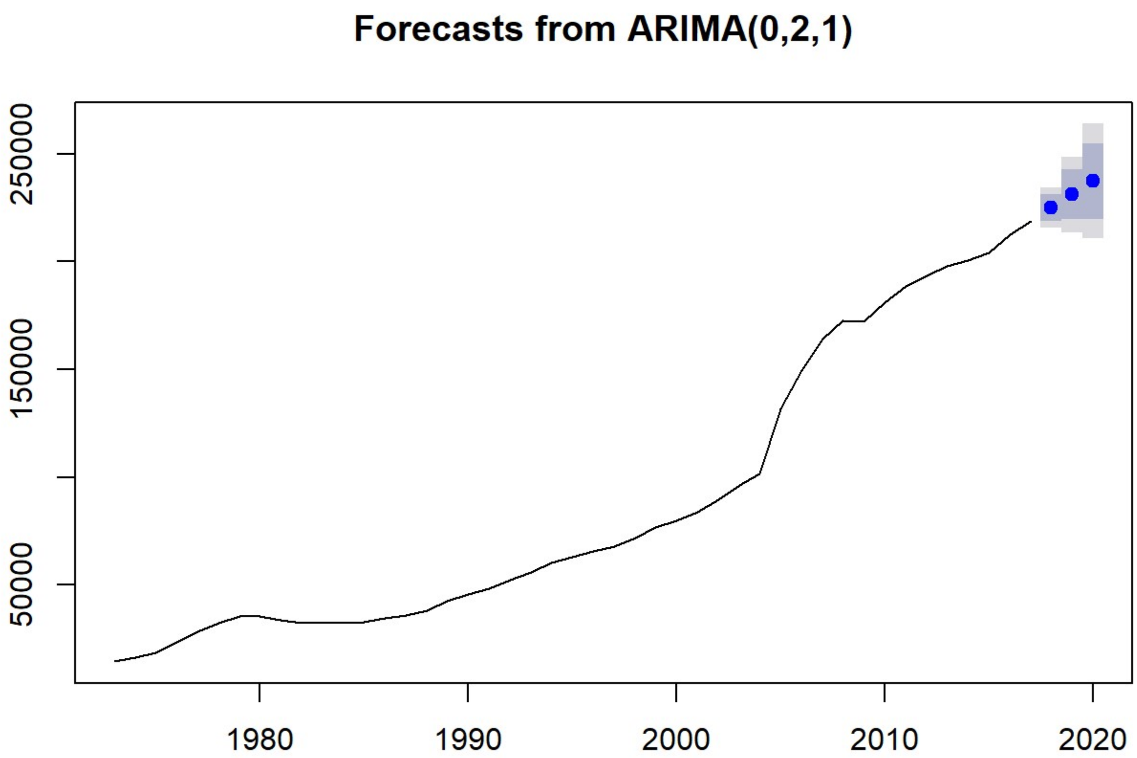
```
##                   RMSE       MAE     MAPE      MASE
## Training set  4994.146  2377.229 3.130373 0.5171411
## Test set     13871.456 13120.899 6.468137 2.8543134
```

```
checkresiduals(best_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,2,1)
## Q* = 4.8235, df = 6.4, p-value = 0.6147
##
## Model df: 1.   Total lags used: 7.4
```

```
main_dataset = Arima(rsvhouse, order = c(0,2,1))

for_main = forecast(main_dataset,h=3)
 plot(for_main)
```



Exercise 2.6

Compare the different models in terms of residual diagnostics, model fit, and forecast accuracy. Analyse your results and select your final model.

Answer:

best model based on partition was Exponential followed by Arima

Arima

```
              RMSE       MAE      MAPE      MASE
```

Training set 4994.146 2377.229 3.130373 0.5171411 Test set 13871.456 13120.899 6.468137 2.8543134

ETS RMSE MAE MAPE MASE Training set 5086.783 2444.953 3.346675 0.5318738 Test set 26642.894 24613.080 12.076095 5.3543163

Exponential RMSE MAE MAPE MASE Training set 5003.296 2442.777 3.581799 0.5314005 Test set 12687.380 12043.597 5.942417 2.6199576 -best

Naive ME RMSE MAE MPE MAPE MASE ACF1 Theil's U Training set 4382.756 7267.137 4596.867 6.454523 7.04680 1.000000 0.5758397 NA Test set 27326.193 29673.556 27326.193 13.400057 13.40006 5.944525 0.5619291 5.365213

Exercise 2.7

Finally, generate out of sample forecasts up to 2020, based on the complete time series. Discuss your results

```
plot(forecast(h1, h=10), include = 80)
```