tsf_export

Kevork Sulahian

October 28, 2019

```r
library(readxl)
library(forecast)
```

```r
# library(readxl)

df <- read_xlsx("Export_for_TS.xlsx")
```

```
## New names:
## * `` -> ...2
```

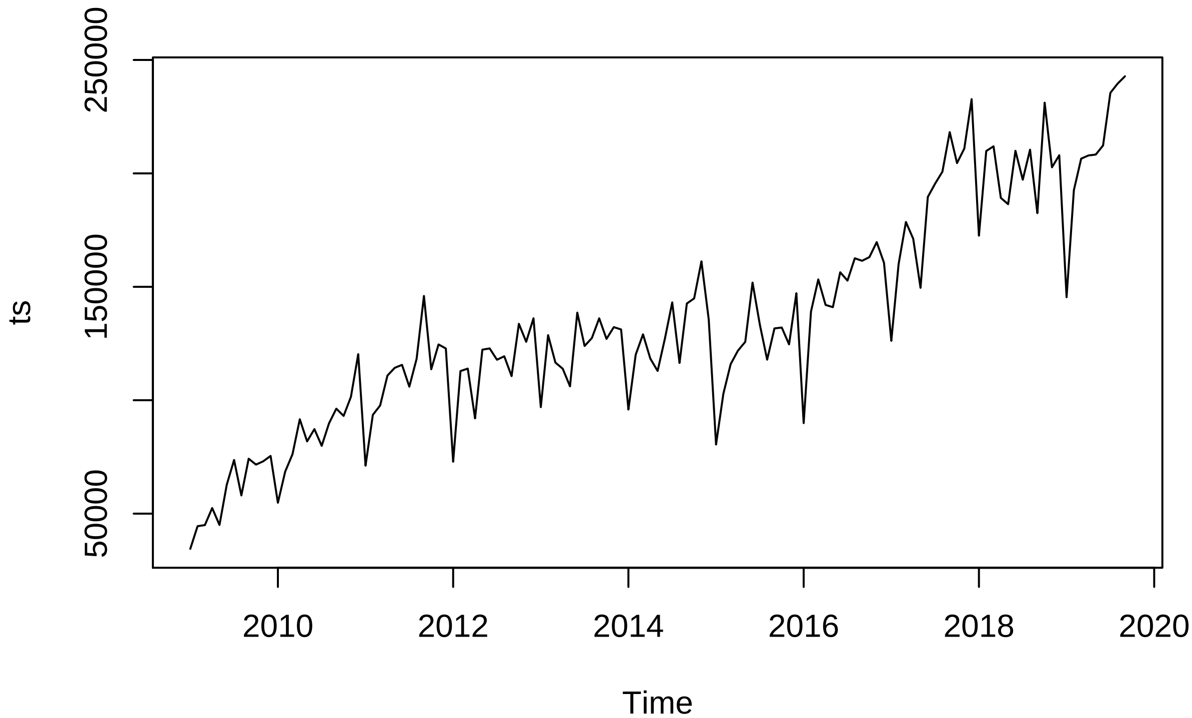
```r
df = df[1,]
df = df[-c(1,2)]

df2 = read_xlsx('export_19xlsx.xlsx')
```

```
## New names:
## * `` -> ...1
```

```r
df = t(df)
df2$...1 = c(paste0(2019,"-",1:9))
rownames(df2) =df2$...1
```

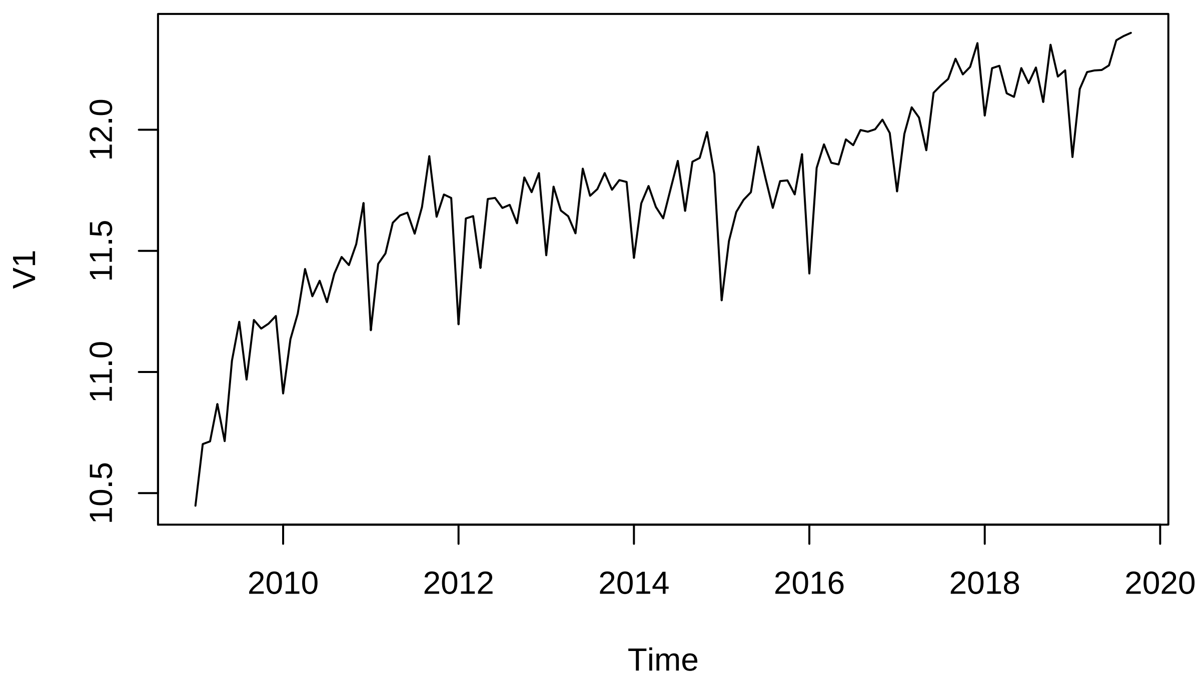
```
## Warning: Setting row names on a tibble is deprecated.
```

```r
df2$...1 = NULL
colnames(df2) = "V1"
df = as.data.frame(df)

df3 = rbind(df,df2)
ts = ts(df3,start=c(2009,1), frequency = c(12))
```



In this case, it appears that an additive model is not appropriate for describing this time series, since the size of the seasonal fluctuations and random fluctuations seem to increase with the level of the time series. Thus, we may need to transform the time series in order to get a transformed time series that can be described using an additive model. For example, we can transform the time series by calculating the natural log of the original data:

```
log_ts <- log(ts)
plot.ts(log_ts)
```



Decomposing Time Series

Decomposing a time series means separating it into its constituent components, which are usually a trend component and an irregular component, and if it is a seasonal time series, a seasonal component.

Decomposing Seasonal Data

A seasonal time series consists of a trend component, a seasonal component and an irregular component. Decomposing the time series means separating the time series into these three components: that is, estimating these three components.

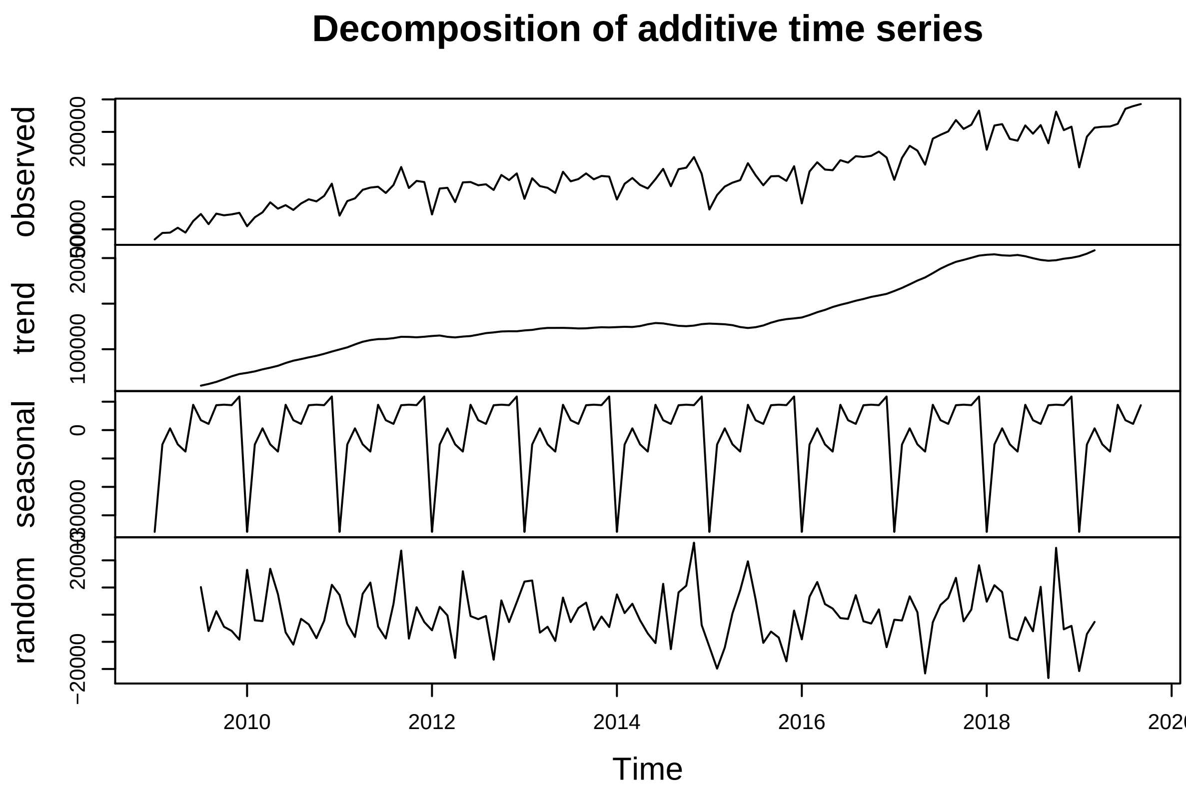
```
ts_components <- decompose(ts)
```

we can print out the estimated values of the seasonal component

```
ts_components$seasonal
```

```
##             Jan         Feb         Mar         Apr         May
## 2009 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2010 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2011 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2012 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2013 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
```

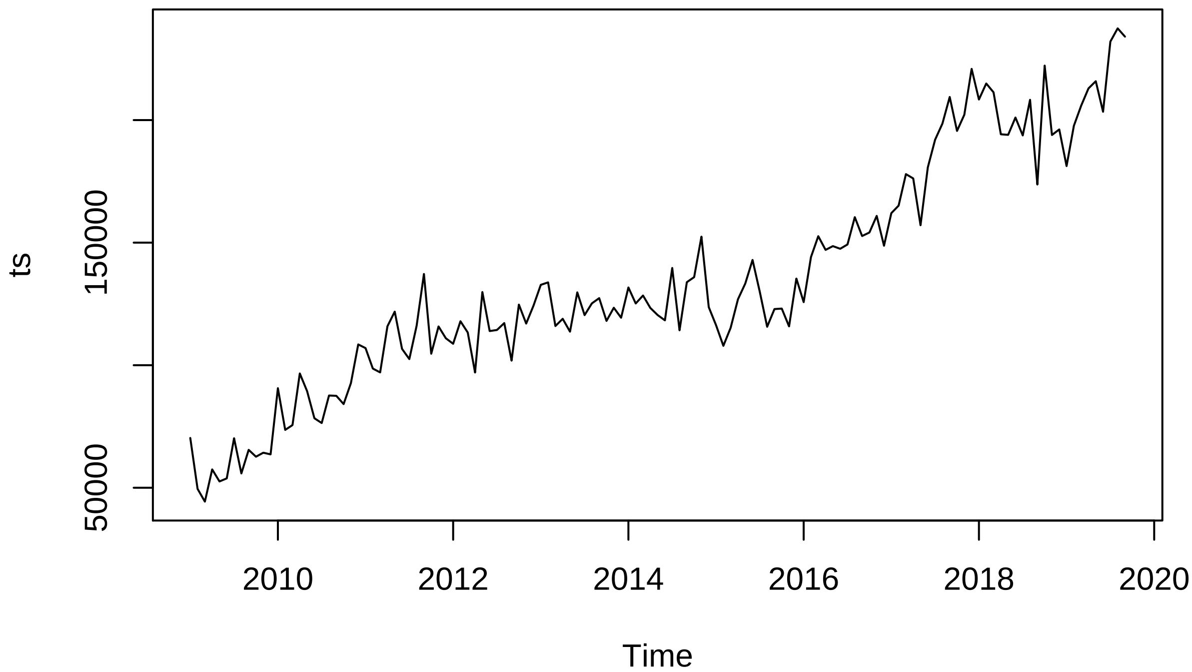
```
## 2014 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2015 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2016 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2017 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2018 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
## 2019 -35837.3733  -5078.4271    614.7571  -5025.8461  -7561.9849
##              Jun         Jul         Aug         Sep         Oct
## 2009   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2010   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2011   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2012   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2013   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2014   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2015   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2016   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2017   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2018   8910.0983   3476.2085   2185.4111   8757.6873   8951.2410
## 2019   8910.0983   3476.2085   2185.4111   8757.6873
##              Nov         Dec
## 2009   8780.2368  11827.9913
## 2010   8780.2368  11827.9913
## 2011   8780.2368  11827.9913
## 2012   8780.2368  11827.9913
## 2013   8780.2368  11827.9913
## 2014   8780.2368  11827.9913
## 2015   8780.2368  11827.9913
## 2016   8780.2368  11827.9913
## 2017   8780.2368  11827.9913
## 2018   8780.2368  11827.9913
## 2019
```



The plot above shows the original time series (top), the estimated trend component (second from top), the estimated seasonal component (third from top), and the estimated irregular component (bottom)

Seasonally Adjusting

```
ts_seasonall <- ts - ts_components$seasonal
```



Holt-Winters Exponential Smoothing

```
ts_forcaste <- HoltWinters(ts)
ts_forcaste
```

```
## Holt-Winters exponential smoothing with trend and additive seasonal component.
##
## Call:
## HoltWinters(x = ts)
##
## Smoothing parameters:
##  alpha: 0.3721975
##  beta : 0.008388046
##  gamma: 0.3879005
##
## Coefficients:
##             [,1]
## a   234236.22519
## b     1800.90011
## s1   10884.32142
## s2    2422.24425
## s3    6639.03969
## s4  -45793.46458
```

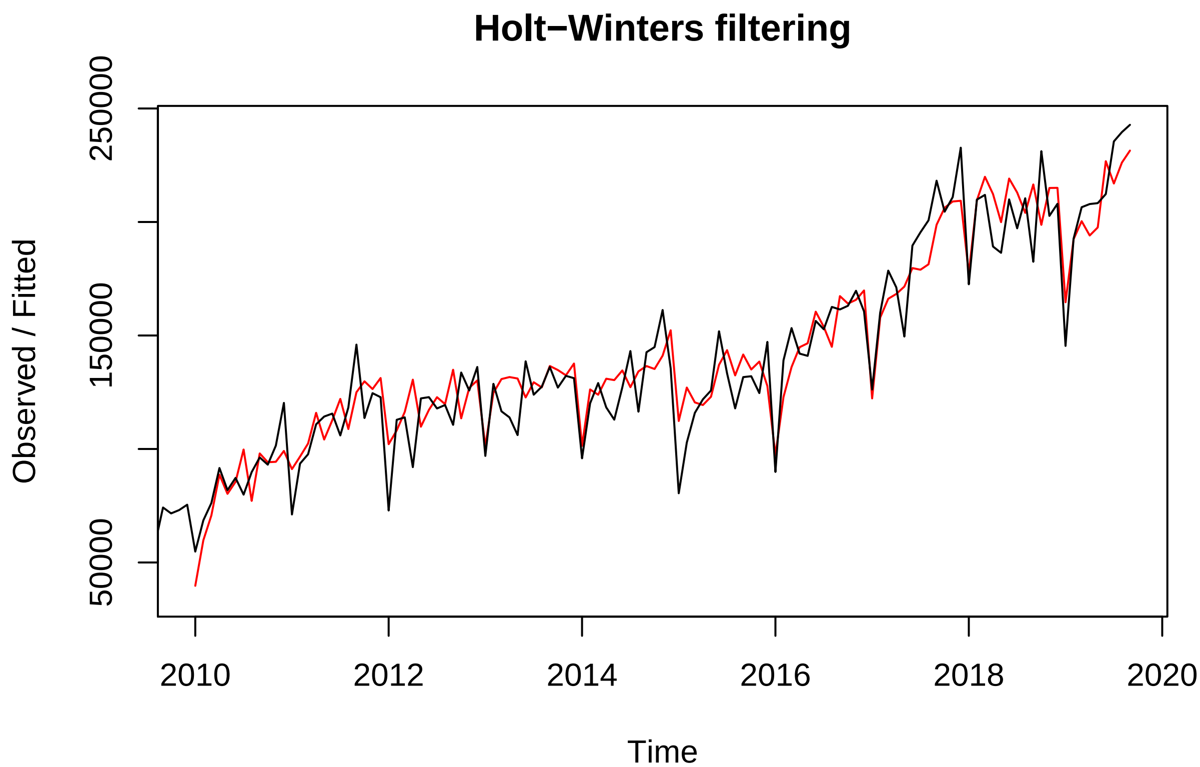
```
## s5   -7778.48158
## s6      -48.50151
## s7   -8403.94186
## s8  -12451.28939
## s9    4905.92488
## s10   6841.04904
## s11   6190.87742
## s12   4206.41639
```

```r
#
```

The value of alpha (0.35) is relatively low, indicating that the estimate of the level at the current time point is based upon both recent observations and some observations in the more distant past. The value of beta is 0.01, indicating that the estimate of the slope b of the trend component is updated but doesn't have much effect over the time series, and instead is set equal to its initial value. This makes good intuitive sense, as the level changes quite a bit over the time series, but the slope b of the trend component remains roughly the same. In contrast, the value of gamma (0.38) is high, indicating that the estimate of the seasonal component at the current time point is not just based upon very recent observations

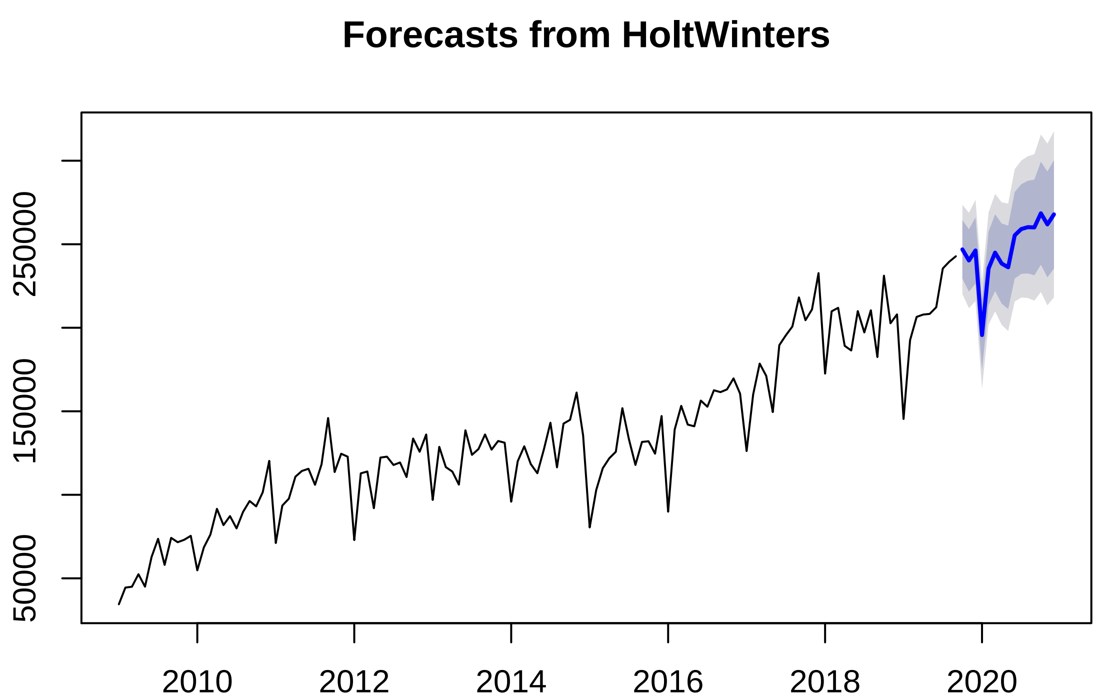
```r
ts_forcaste$SSE
```

```
## [1] 21744166076
```



```
ts_forcaste2 = forecast:::forecast.HoltWinters(ts_forcaste, h= 15)
(as.data.frame(ts_forcaste2))[1]
```

```
##          Point Forecast
## Oct 2019       246921.4
## Nov 2019       240260.3
## Dec 2019       246278.0
## Jan 2020       195646.4
## Feb 2020       235462.2
## Mar 2020       244993.1
## Apr 2020       238438.6
## May 2020       236192.1
## Jun 2020       255350.3
## Jul 2020       259086.3
## Aug 2020       260237.0
## Sep 2020       260053.4
## Oct 2020       268532.2
## Nov 2020       261871.1
## Dec 2020       267888.8
```



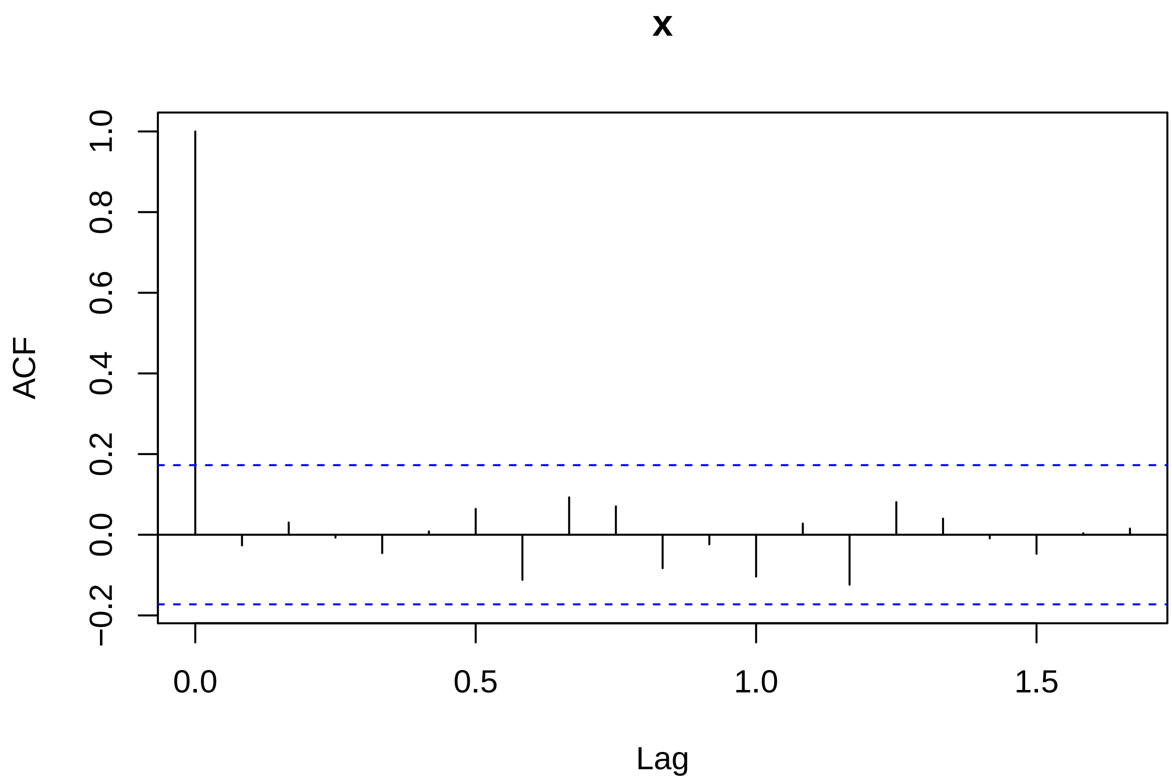
Growth

```
year_2019 <- window(ts, 2019)
year_2019_predict_HW <- (as.data.frame(ts_forcaste2))[1][c(1:3),]
sum_year_2019 = sum(c(year_2019,year_2019_predict_HW))
year_2020 = (as.data.frame(ts_forcaste2))[1][c(4:15),]
```

```r
growth_HW <- growth(sum(year_2020),sum_year_2019)
growth_HW
```

```
## [1] 0.1369446
```

We can investigate whether the predictive model can be improved upon by checking whether the in-sample forecast errors show non-zero autocorrelations at lags 1-20, by making a correlogram and carrying out the Ljung-Box test:

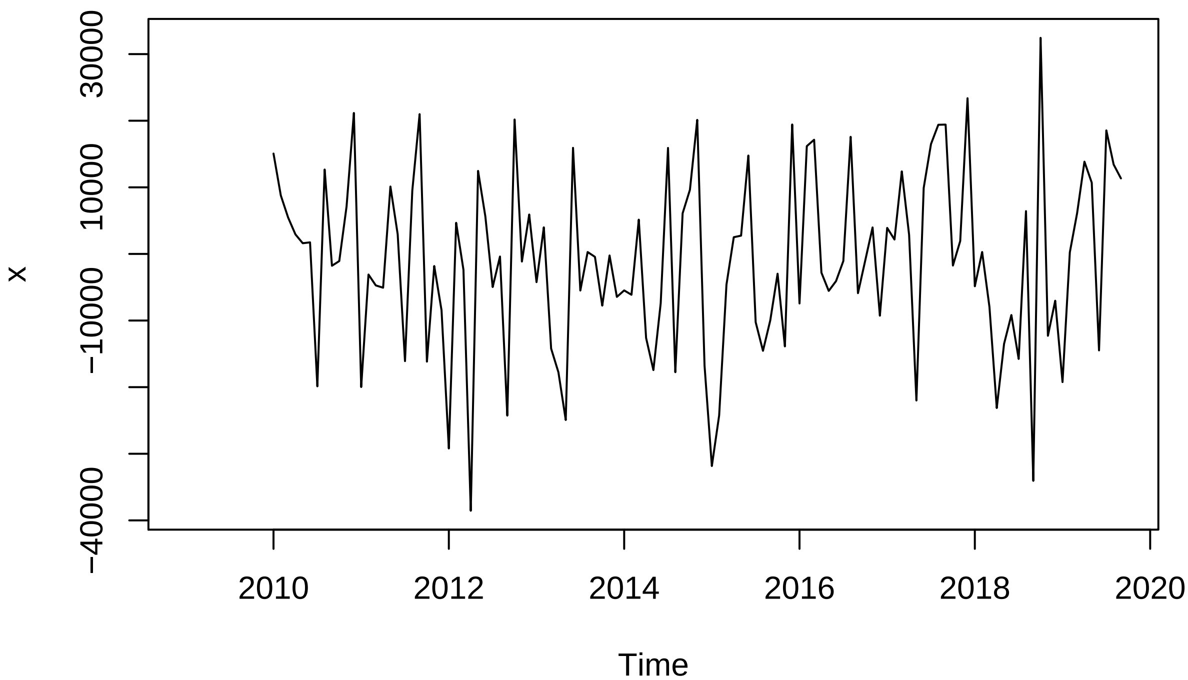



```
##
##  Box-Ljung test
##
## data:  ts_forcaste2$residuals
## X-squared = 10.37, df = 20, p-value = 0.961
```

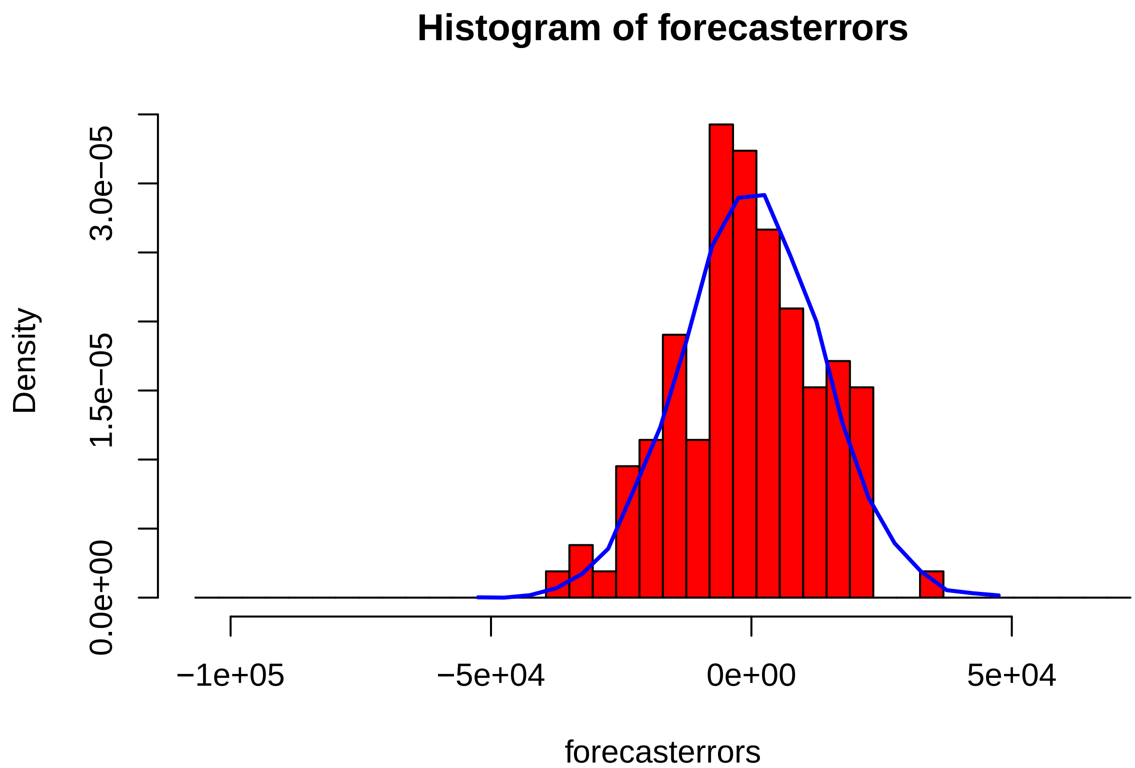
The correlogram shows that the autocorrelations for the in-sample forecast errors do not exceed the significance bounds for lags 1-20. Furthermore, the p-value for Ljung-Box test is 0.9, indicating that there is no evidence of non-zero autocorrelations at lags 1-20.

We can check whether the forecast errors have constant variance over time, and are normally distributed with mean zero, by making a time plot of the forecast errors and a histogram (with overlaid normal curve):

```r
plot.ts(ts_forcaste2$residuals)
```



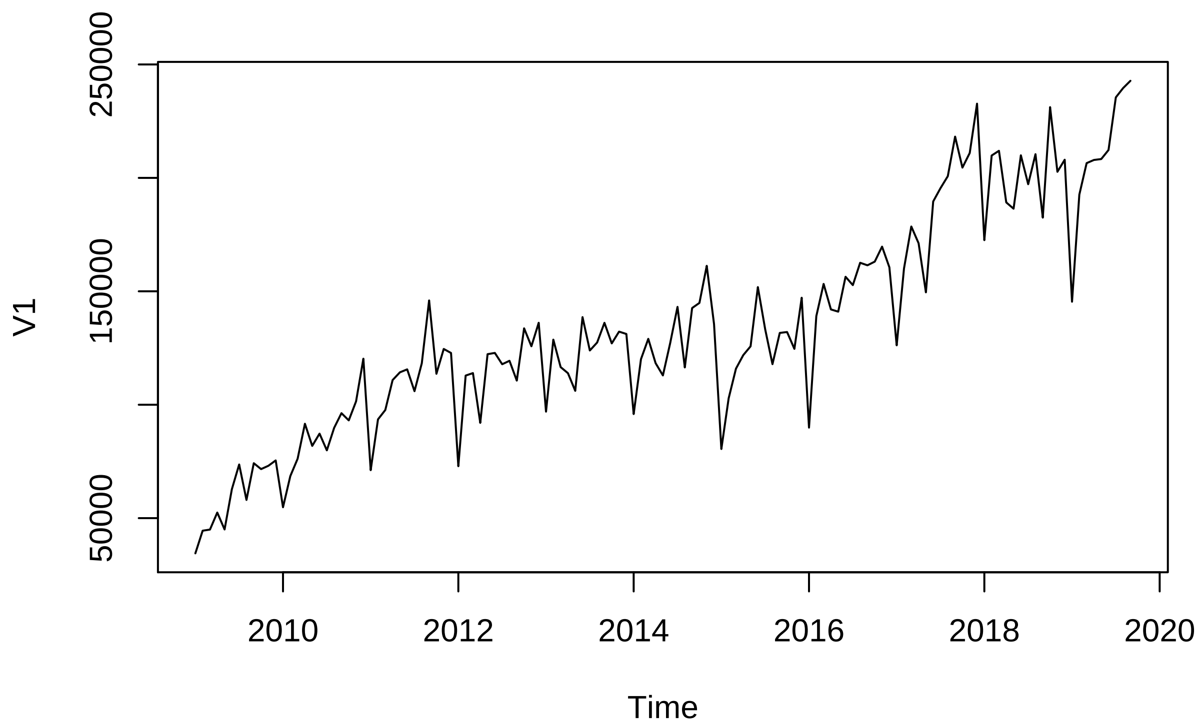

```
plotForecastErrors(ts_forcaste2$residuals)
```



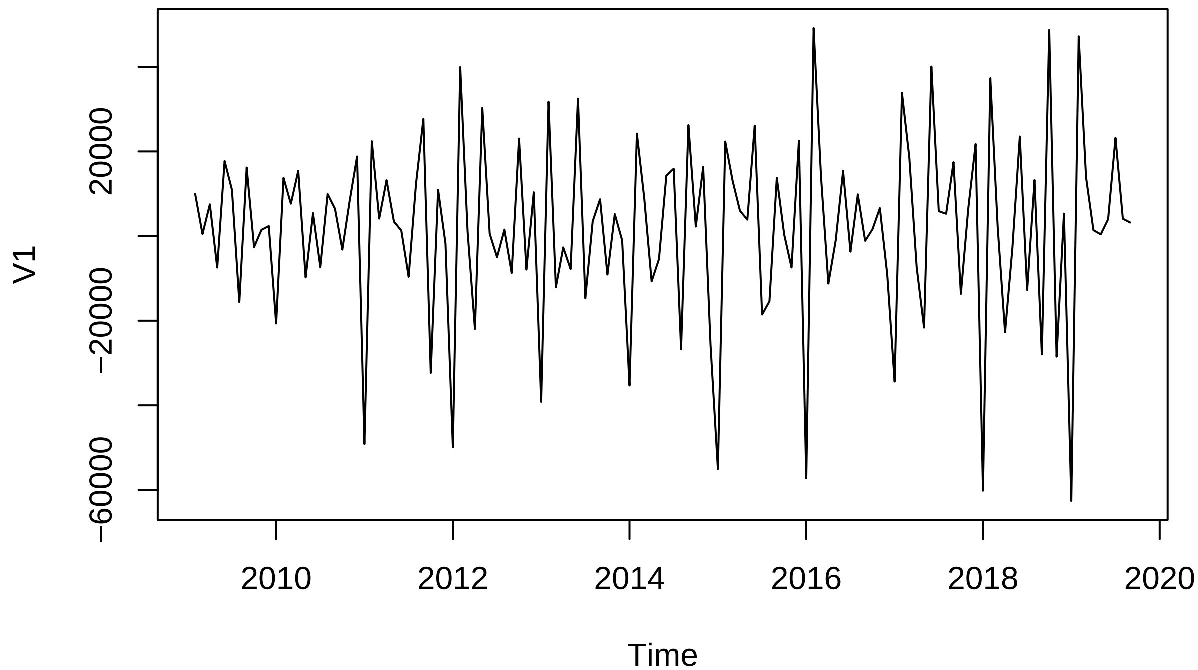
From the time plot, it appears plausible that the forecast errors have constant variance over time. From the histogram of forecast errors, it seems plausible that the forecast errors are normally distributed with mean zero.

Thus,there is little evidence of autocorrelation at lags 1-20 for the forecast errors, and the forecast errors appear to be normally distributed with mean zero and constant variance over time. This suggests that Holt-Winters exponential smoothing provides an adequate predictive model of the log of total productivity, which probably cannot be improved upon. Furthermore, the assumptions upon which the prediction intervals were based are probably valid.

```
plot.ts(ts)
```

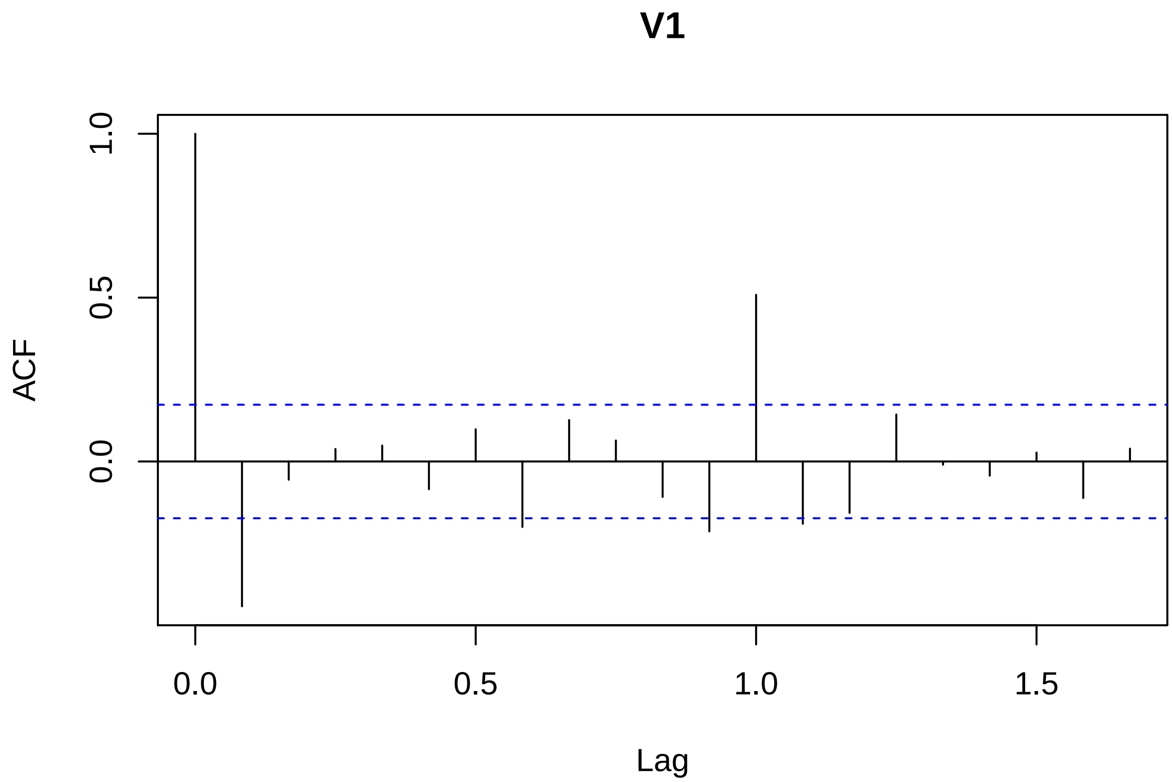



```r
ts_diff1 <-  diff(ts, differences = 1)

plot.ts(ts_diff1)
```



The time series of differences (above) does appear to be stationary in mean and variance, as the level of the series stays roughly constant over time, and the variance of the series appears roughly constant over time

```
acf(ts_diff1, lag.max=20)             # plot a correlogram
```

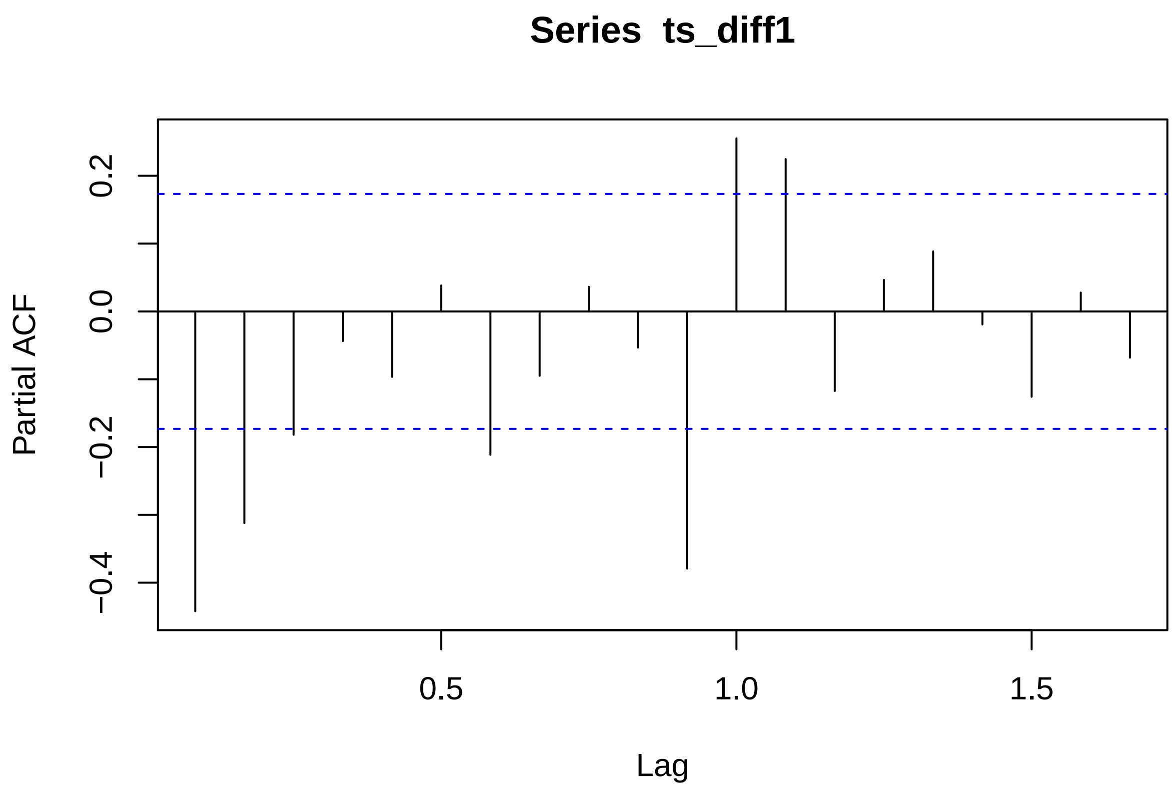


We see from the correlogram that the autocorrelation exceeds the significance bound 3 times but all the others do not exceed

```
acf(ts_diff1, lag.max=20, plot=FALSE) # get the autocorrelation values
```

```
##
## Autocorrelations of series 'ts_diff1', by lag
##
## 0.0000 0.0833 0.1667 0.2500 0.3333 0.4167 0.5000 0.5833 0.6667 0.7500
##  1.000 -0.442 -0.056  0.038  0.049 -0.085  0.098 -0.200  0.126  0.064
## 0.8333 0.9167 1.0000 1.0833 1.1667 1.2500 1.3333 1.4167 1.5000 1.5833
## -0.108 -0.213  0.508 -0.190 -0.157  0.143 -0.010 -0.043  0.027 -0.111
## 1.6667
##  0.039
```

```
pacf(ts_diff1, lag.max=20)            # plot a partial correlogram
```




```
pacf(ts_diff1, lag.max=20, plot=FALSE) # get the partial autocorrelation values
```

```
##
## Partial autocorrelations of series 'ts_diff1', by lag
##
## 0.0833 0.1667 0.2500 0.3333 0.4167 0.5000 0.5833 0.6667 0.7500 0.8333
## -0.442 -0.312 -0.182 -0.044 -0.097  0.038 -0.211 -0.095  0.036 -0.053
## 0.9167 1.0000 1.0833 1.1667 1.2500 1.3333 1.4167 1.5000 1.5833 1.6667
## -0.379  0.255  0.225 -0.117  0.047  0.089 -0.019 -0.126  0.028 -0.068
```

Arima, 0,1,0

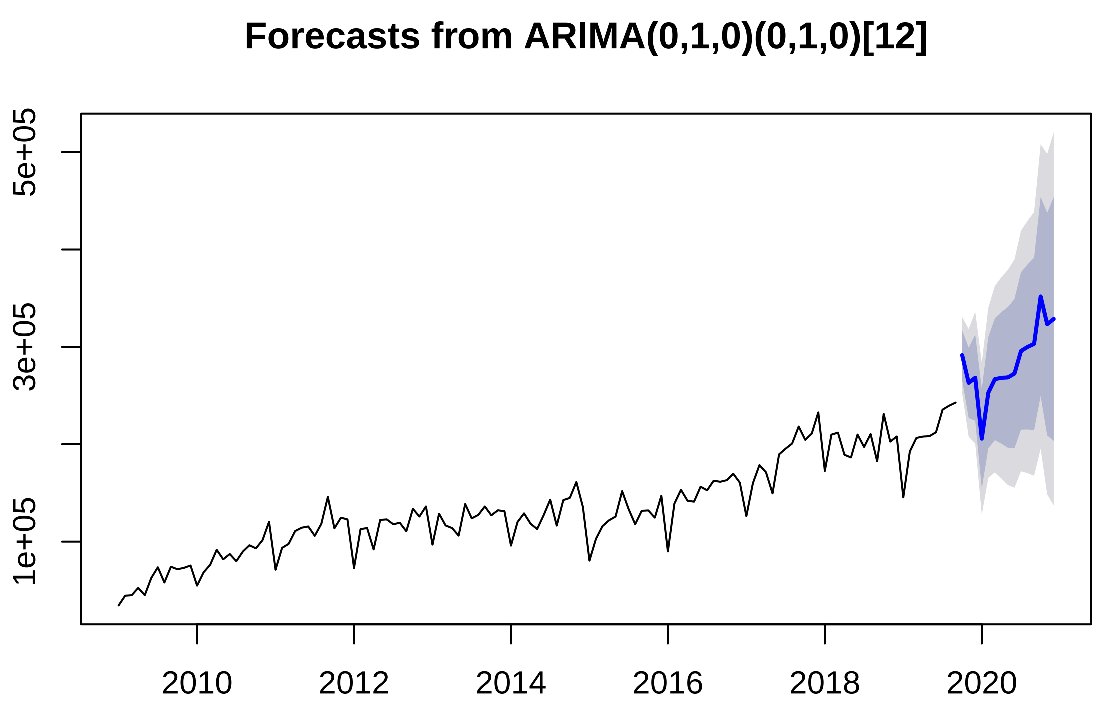
```
ts_arima = Arima(ts, order=c(0,1,0),seasonal = list(order = c(0,1,0)))
ts_arima
```

```
## Series: ts
## ARIMA(0,1,0)(0,1,0)[12]
##
## sigma^2 estimated as 397639845:  log likelihood=-1313.06
## AIC=2628.12   AICc=2628.15   BIC=2630.87
```

```
ts_arima_forecast = forecast(ts_arima,h = 15)
ts_arima_forecast
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Oct 2019       291503.5 265948.2 317058.8 252420.1 330587.0
## Nov 2019       263011.9 226871.2 299152.5 207739.5 318284.2
## Dec 2019       268333.3 224070.2 312596.4 200638.8 336027.9
## Jan 2020       205722.4 154611.8 256833.0 127555.5 283889.3
## Feb 2020       252922.4 195779.0 310065.8 165529.1 340315.7
## Mar 2020       266822.4 204225.0 329419.9 171087.9 362556.9
## Apr 2020       268222.4 200609.4 335835.4 164817.3 371627.5
## May 2020       268622.4 196341.1 340903.7 158077.7 379167.1
## Jun 2020       272622.4 195956.5 349288.3 155372.0 389872.8
## Jul 2020       295822.4 215009.4 376635.4 172229.6 419415.2
## Aug 2020       299922.4 215165.1 384679.8 170297.2 429547.6
## Sep 2020       303122.4 214596.2 391648.6 167733.3 438511.5
## Oct 2020       351825.9 249604.7 454047.1 195492.1 508159.8
## Nov 2020       323334.3 209047.5 437621.1 148547.7 498120.8
## Dec 2020       328655.7 203460.8 453850.6 137186.6 520124.8
```

```
forecast:::plot.forecast(ts_arima_forecast)
```



Growth

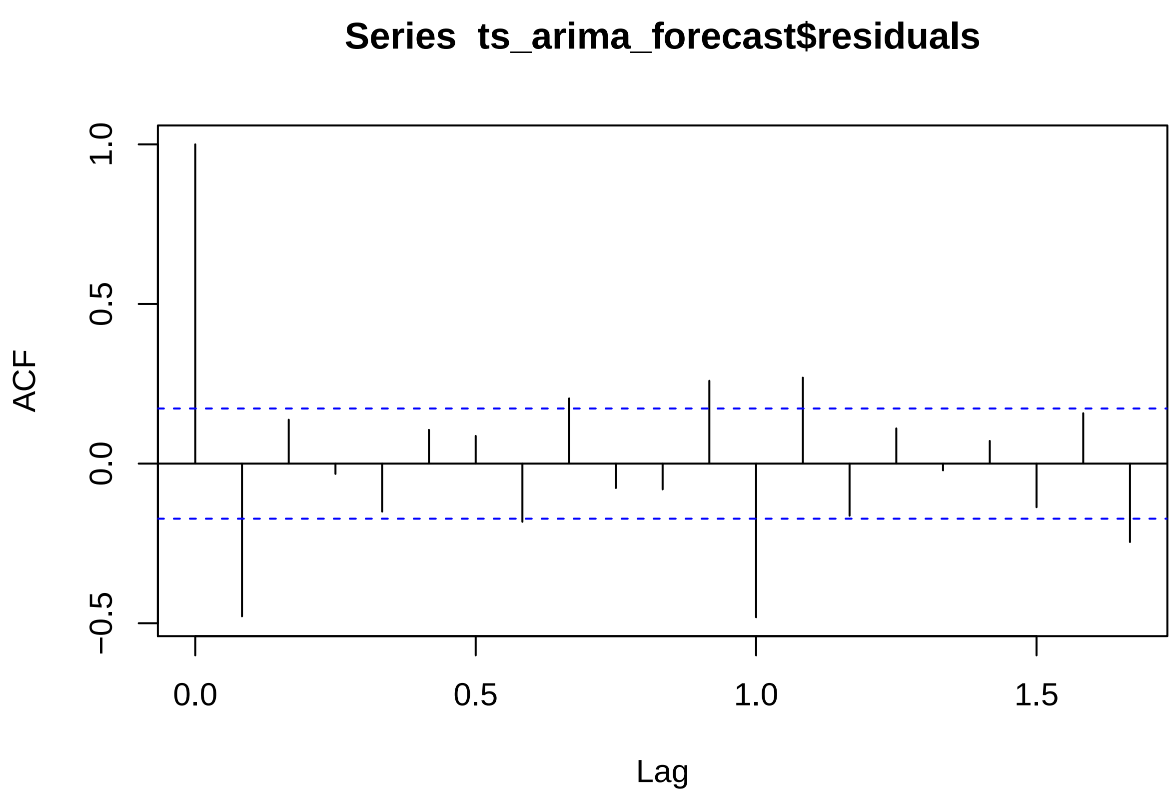
```r
this_year_predict_ARIMA <- (as.data.frame(ts_arima_forecast))[1]

# growth_ARIMA <- growth(sum(c(this_year,as.numeric(this_year_predict_ARIMA$`Point Forecast`))), sum(la
# growth_ARIMA


year_2019_predict_ARIMA <- (as.data.frame(ts_arima_forecast))[1][c(1:3),]
sum_year_2019 = sum(c(year_2019,year_2019_predict_ARIMA))
year_2020 = (as.data.frame(ts_arima_forecast))[1][c(4:15),]
growth_ARIMA <- growth(sum_year_2019, sum(year_2020))
-growth_ARIMA
```

```
## [1] 0.2105728
```

As in the case of exponential smoothing models, it is a good idea to investigate whether the forecast errors of an ARIMA model are normally distributed with mean zero and constant variance, and whether the are correlations between successive forecast errors.

For example, we can make a correlogram of the forecast errors for our ARIMA(0,1,1) model, and perform the Ljung-Box test for lags 1-20, by typing:

```r
acf(ts_arima_forecast$residuals, lag.max=20)
```

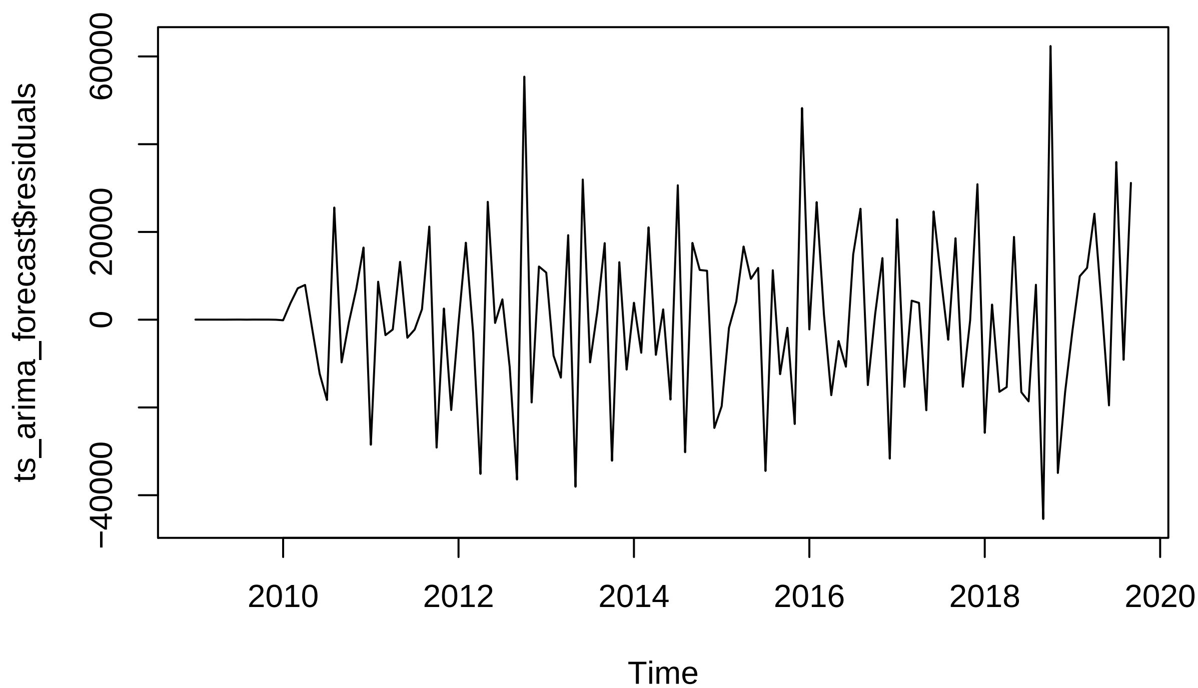


```r
Box.test(ts_arima_forecast$residuals, lag=20, type="Ljung-Box")
```

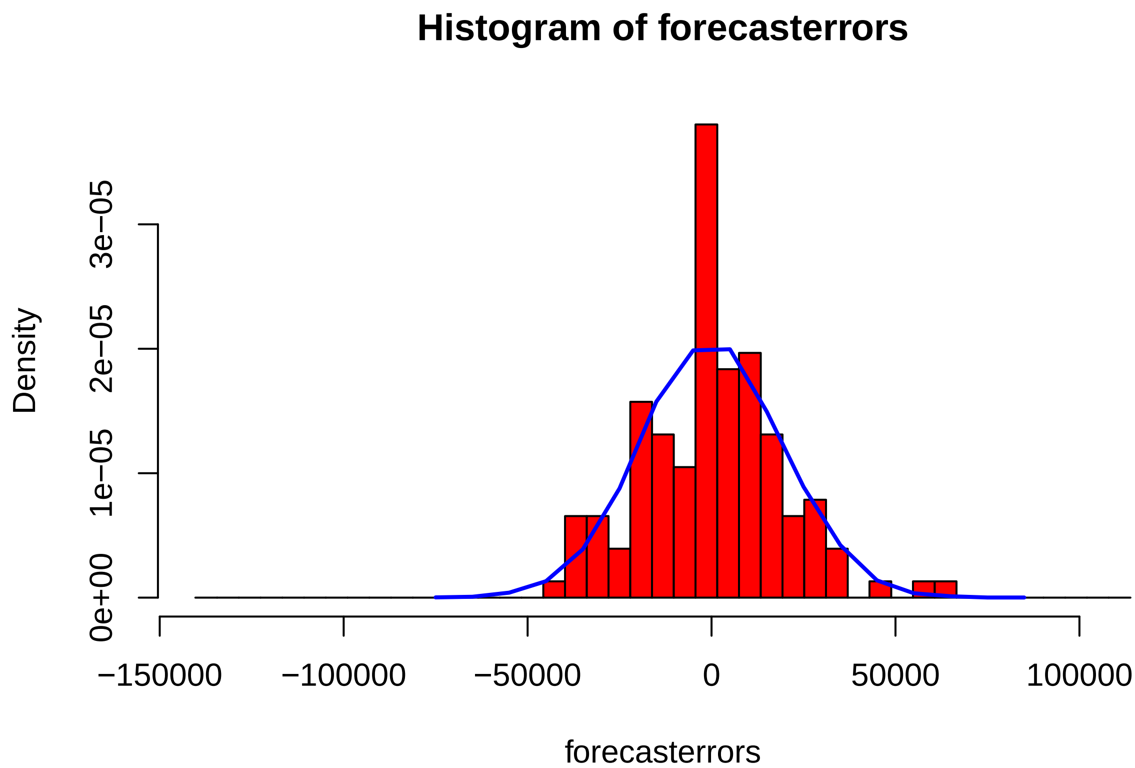
```
##
##  Box-Ljung test
##
## data:  ts_arima_forecast$residuals
## X-squared = 126.86, df = 20, p-value < 2.2e-16
```

p value too high to reject

```r
plot.ts(ts_arima_forecast$residuals)            # make time plot of forecast errors
```




```r
plotForecastErrors(ts_arima_forecast$residuals)
```

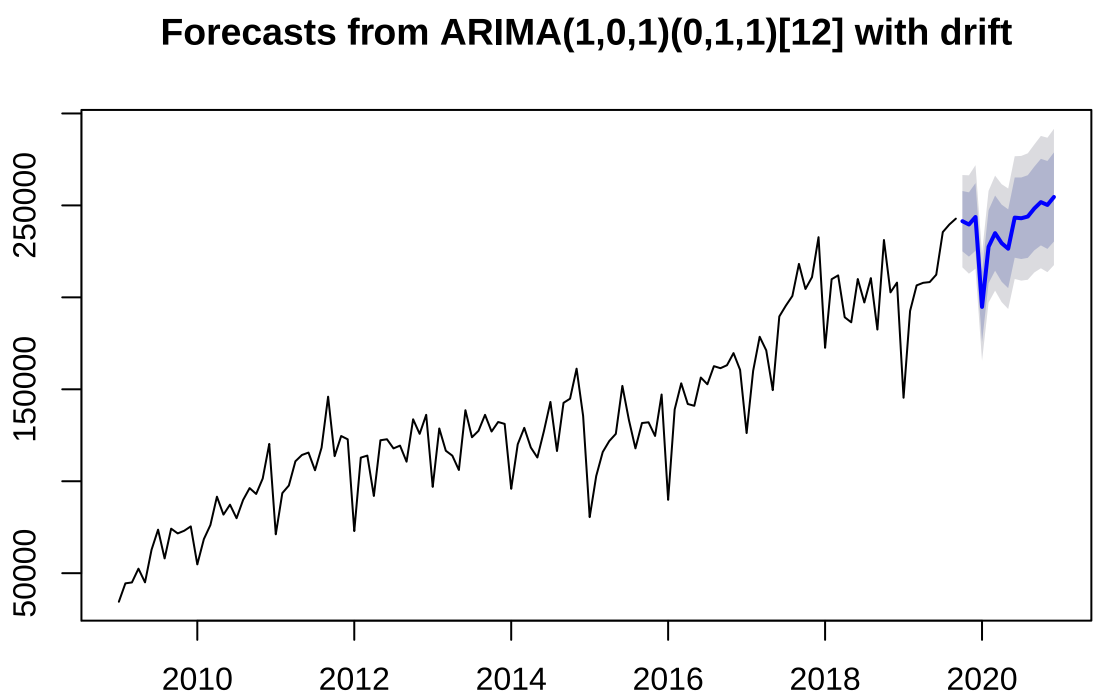


A model chosen automatically

```
fit3 <- auto.arima(ts)
fit3
```

```
## Series: ts
## ARIMA(1,0,1)(0,1,1)[12] with drift
##
## Coefficients:
##          ar1      ma1     sma1     drift
##       0.9482  -0.5785  -0.8459  1269.853
## s.e.  0.0409   0.0914   0.1186   203.636
##
## sigma^2 estimated as 162448032:  log likelihood=-1277.32
## AIC=2564.64   AICc=2565.18   BIC=2578.45
```

```
fit_forecast = forecast(fit3,h=15)
plot(fit_forecast)
```




```
# str(fit)
```

Growth

```
year_2019 <- window(ts, 2019)
year_2019_predict_HW <- (as.data.frame(ts_forcaste2))[1][c(1:3),]
sum_year_2019 = sum(c(year_2019,year_2019_predict_HW))
year_2020 = (as.data.frame(ts_forcaste2))[1][c(4:15),]
growth_HW <- growth(sum(year_2020),sum_year_2019)
growth_HW
```

```
## [1] 0.1369446
```

```
year_2019_predict_auto.arima <- (as.data.frame(fit_forecast))[1][c(1:3),]
year_2019_predict_auto.arima_95_low <- (as.data.frame(fit_forecast))[4][c(1:3),]
year_2019_predict_auto.arima_95_high <- (as.data.frame(fit_forecast))[5][c(1:3),]

sum_year_2019 = sum(c(year_2019,year_2019_predict_auto.arima))
sum_year_2019_low = sum(c(year_2019,year_2019_predict_auto.arima_95_low))
sum_year_2019_high = sum(c(year_2019,year_2019_predict_auto.arima_95_high))


year_2020_predict_auto.arima <- (as.data.frame(fit_forecast))[1][c(4:15),]
```

```r
year_2020_predict_auto.arima_95_low <- (as.data.frame(fit_forecast))[4][c(4:15),]
year_2020_predict_auto.arima_95_high <- (as.data.frame(fit_forecast))[5][c(4:15),]

growth_auto.arima <- growth(sum(year_2020_predict_auto.arima),sum_year_2019)
growth_auto.arima_95_low <- growth(sum(year_2020_predict_auto.arima_95_low),sum_year_2019_low)
growth_auto.arima_95_high <- growth(sum(year_2020_predict_auto.arima_95_high),sum_year_2019_high)

growth_auto.arima
```

```
## [1] 0.08897661
```

```r
growth_auto.arima_95_low
```

```
## [1] -0.03516189
```

```r
growth_auto.arima_95_high
```

```
## [1] 0.2057449
```