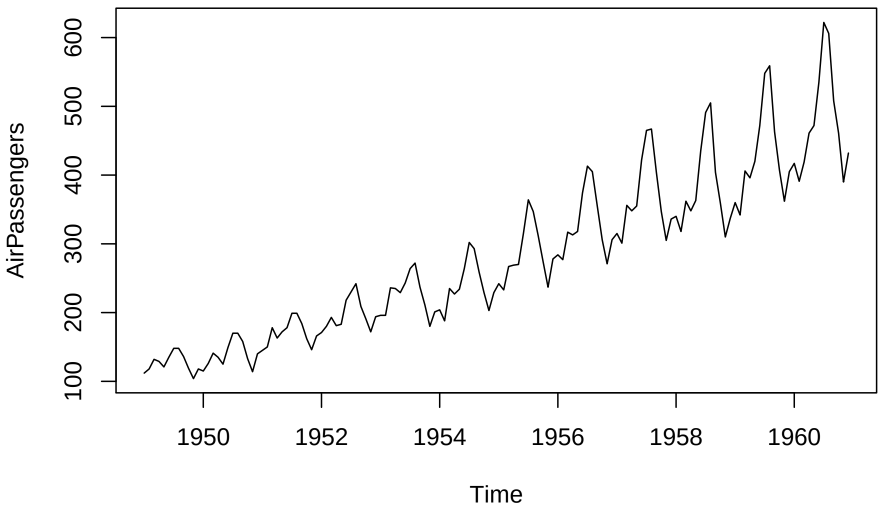
Air Passengers

Julien JACQUES

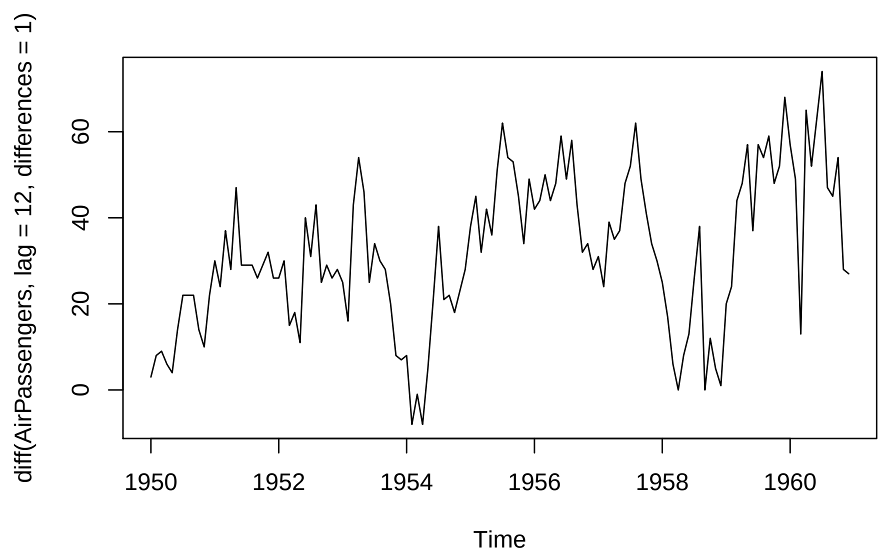
2/19/2020

We study the number of passengers per month (in thousands) in air transport, from 1949 to 1960. This time series is available on R (`AirPassengers`).

```
data("AirPassengers")
plot(AirPassengers)
```



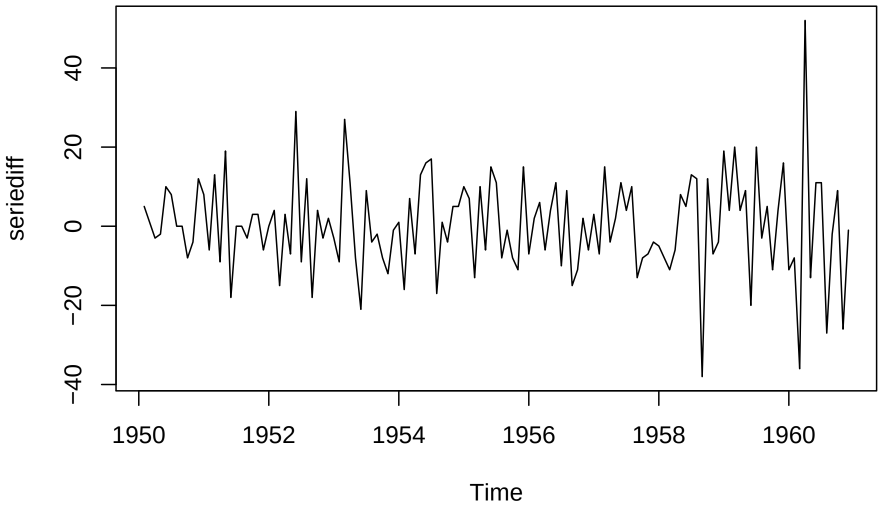

```
plot(diff(AirPassengers,lag = 12,differences = 1))
```



It seems that there is still a slightly growing trend...

Let's apply differencing once again

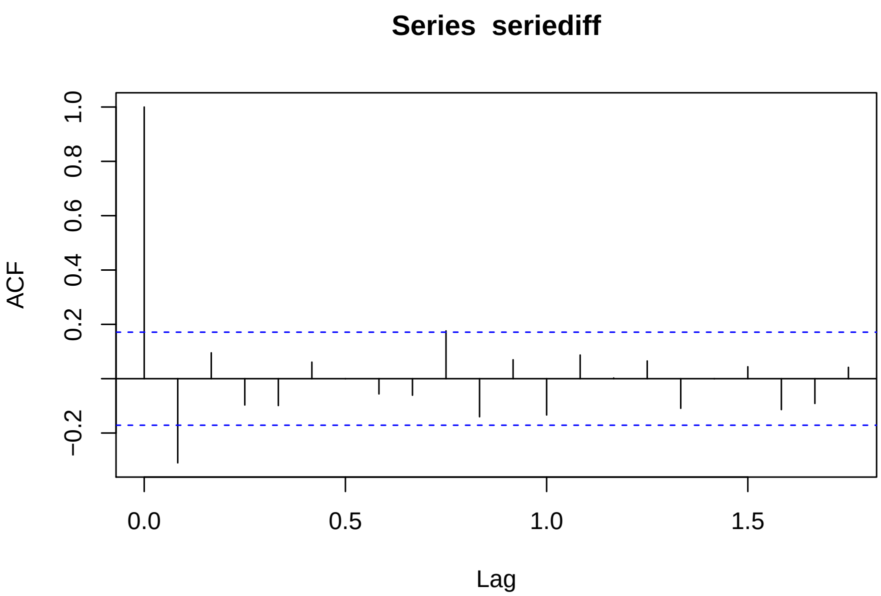
```
seriediff=diff(diff(AirPassengers,lag = 12))
plot(seriediff)
```



It seems visually stationary…

We can have a look to the correlogram and

```
acf(seriediff)
```



and test if the residual series is a white noise

```
Box.test(seriediff,lag=10,type="Ljung-Box")$p.value
```

```
## [1] 0.004553247
```

This is not the case: there is some autocorrelations to modelize…

Forecasting with Exponential Smoothing

We split into train and test sets:

```
library(forecast)
AP_train=window(AirPassengers,start=c(1949,1),end=c(1958,12))
AP_test=window(AirPassengers,start=c(1959,1),end=c(1960,12))
```

The series admits a multiplicative seasonal pattern and a trend, so we can use an multiplicative seasonal HW smoothing or an additive on with a Box-Cox transformation

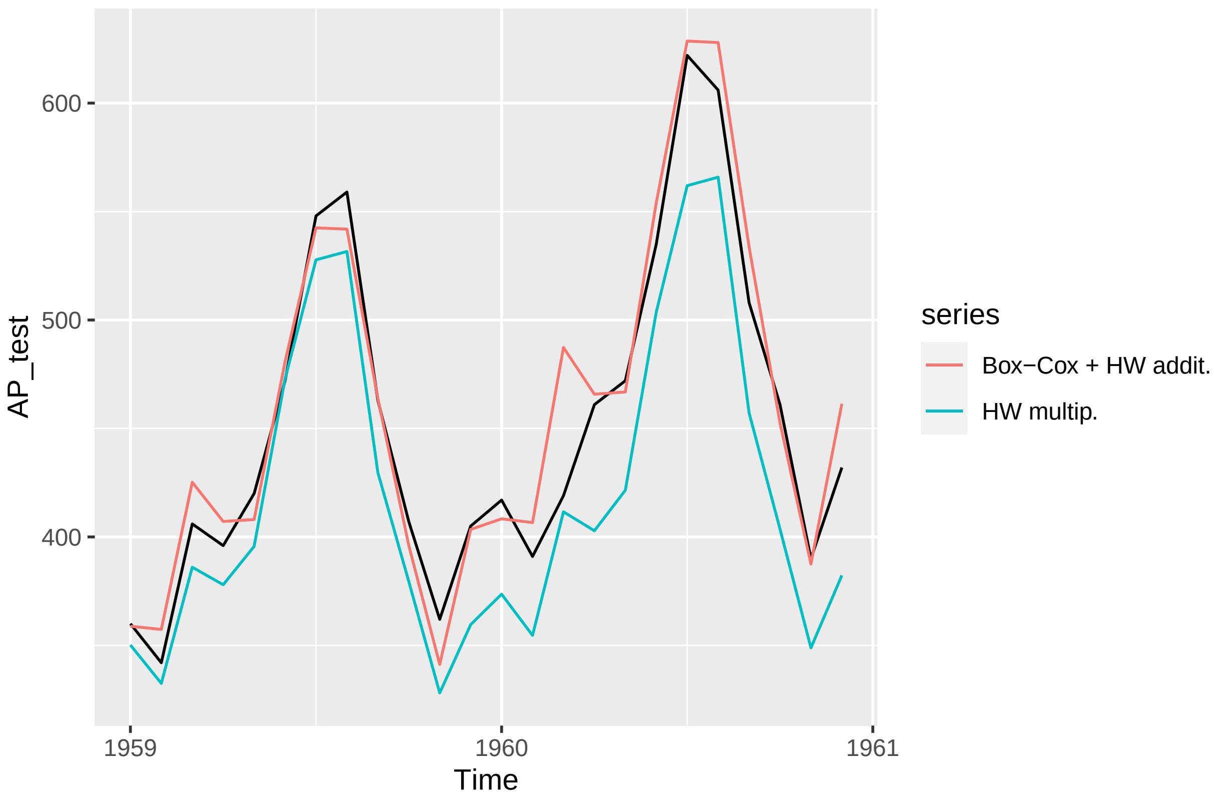
```
fit1=hw(AP_train,seasonal='multiplicative',h=24)
fit2=hw(AP_train,seasonal='additive',lambda="auto",h=24)
print(sqrt(mean((fit1$mean-AP_test)^2)))
```

```
## [1] 37.1814
```

```
print(sqrt(mean((fit2$mean-AP_test)^2)))
```

```
## [1] 19.75243
```

```
autoplot(AP_test)+autolayer(fit1$mean,series="HW multip.")+
  autolayer(fit2$mean,series="Box-Cox + HW addit.")
```



The second solution with the Box-Cox transformation is clearly better.

Forecasting with ARIMA

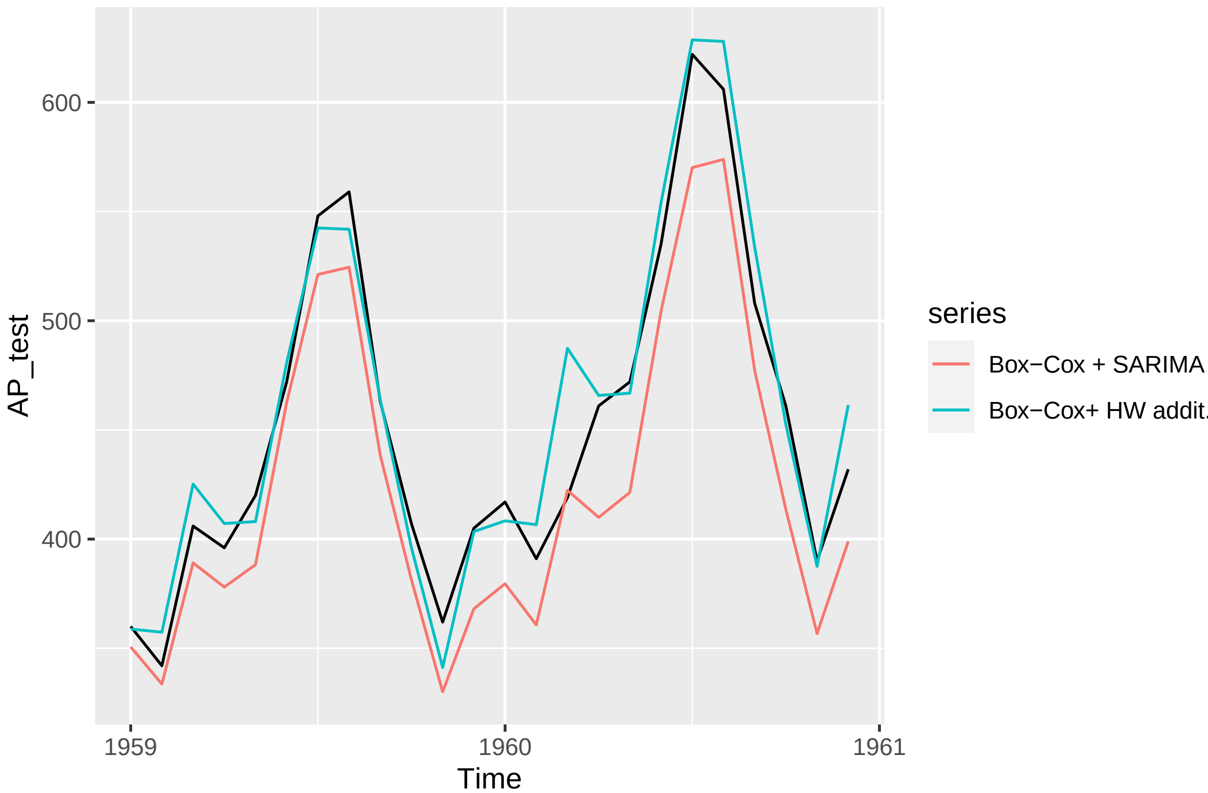
Automatically

We choose automaticaly an SARIMA model, using the Box-Cox transformation

```
fit3=auto.arima(AP_train,lambda = "auto")
prev=forecast(fit3,h=24)
print(sqrt(mean((prev$mean-AP_test)^2)))
```

```
## [1] 32.25238
```

```
autoplot(AP_test)+autolayer(prev$mean,series="Box-Cox + SARIMA")+
  autolayer(fit2$mean,series="Box-Cox+ HW addit.")
```

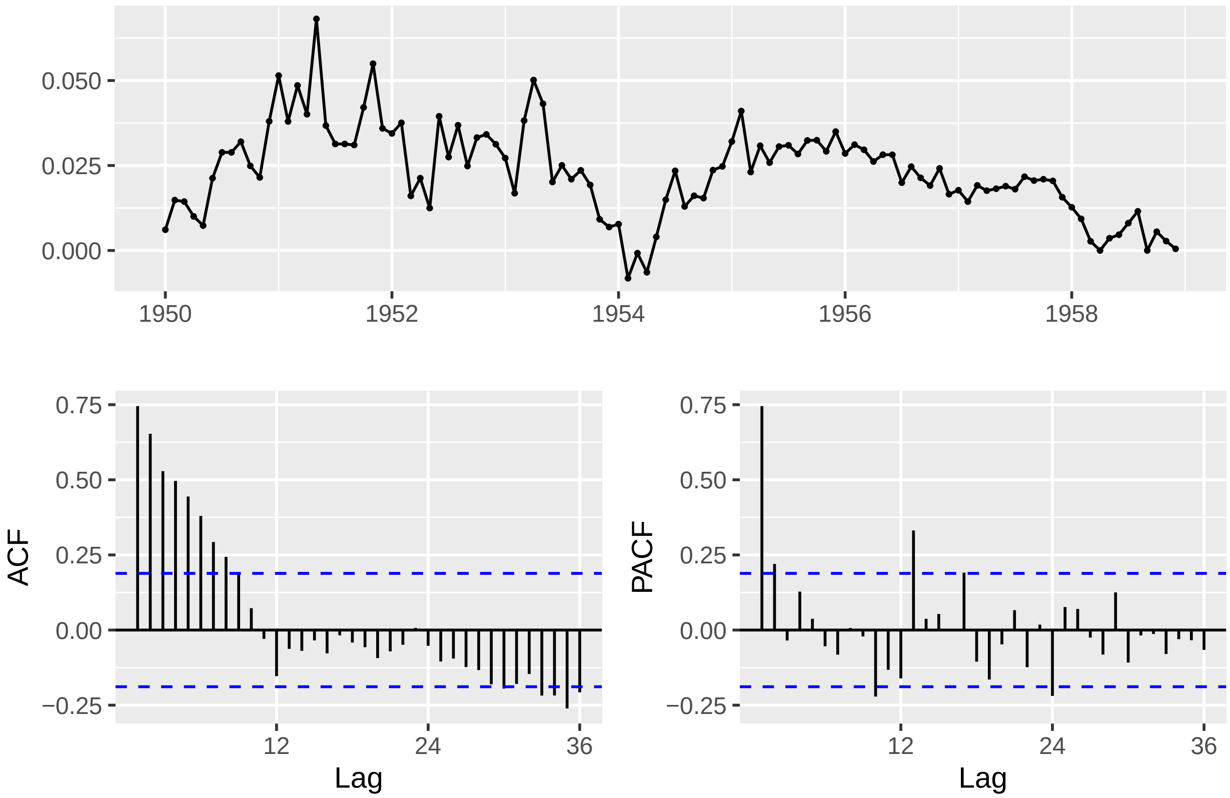


The results are bot better than with HW.

Manual choice of the order

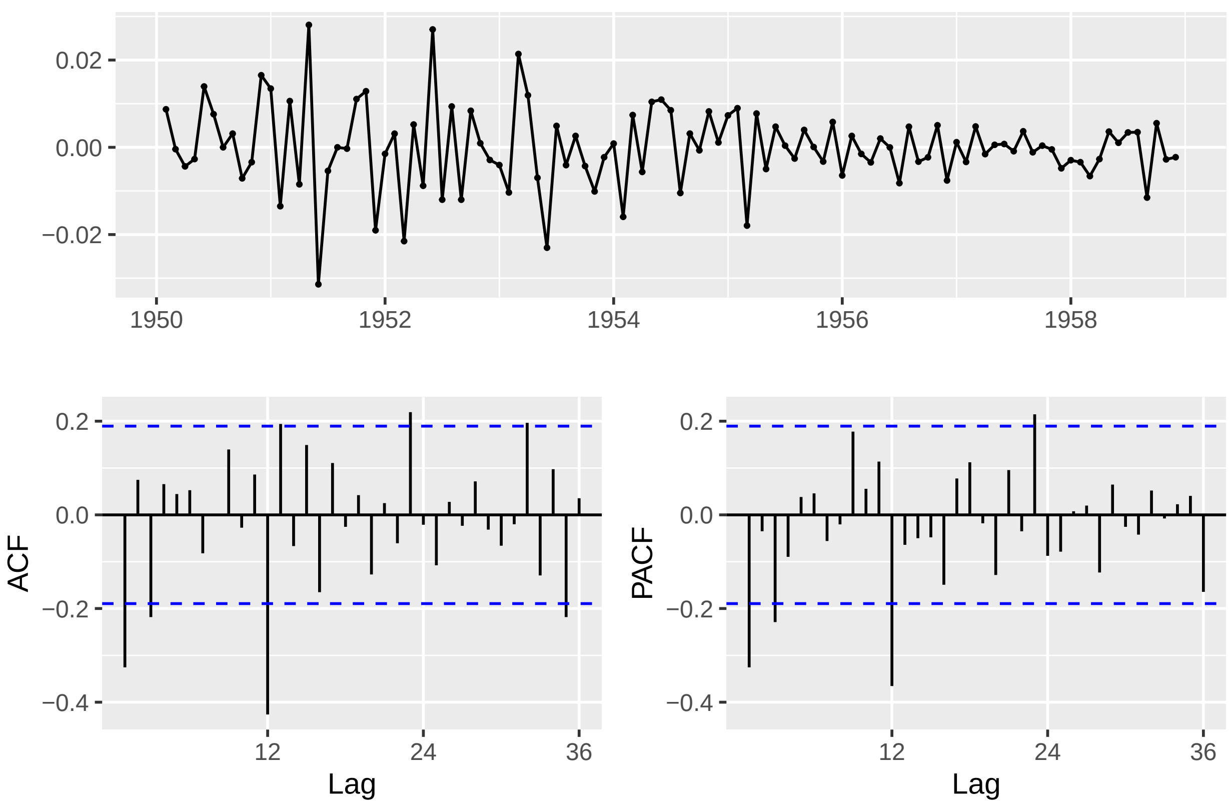
We can try to choose manually the order of a SARIMA.

```
lambda=BoxCox.lambda(AP_train)
AP_BC=BoxCox(AP_train,lambda)
tmp=diff(AP_BC,lag=12)
ggtsdisplay(tmp)
```



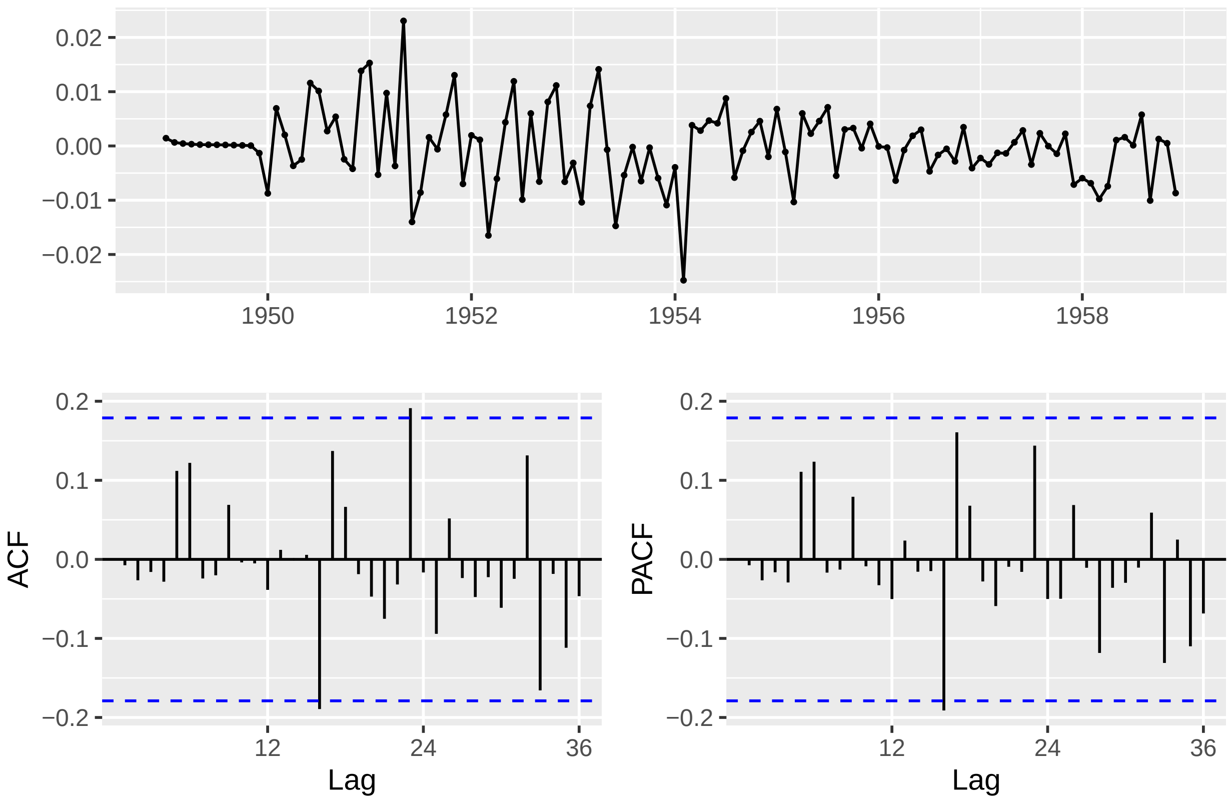
It seems that it remains a trend

```r
ggtsdisplay(diff(tmp))
```



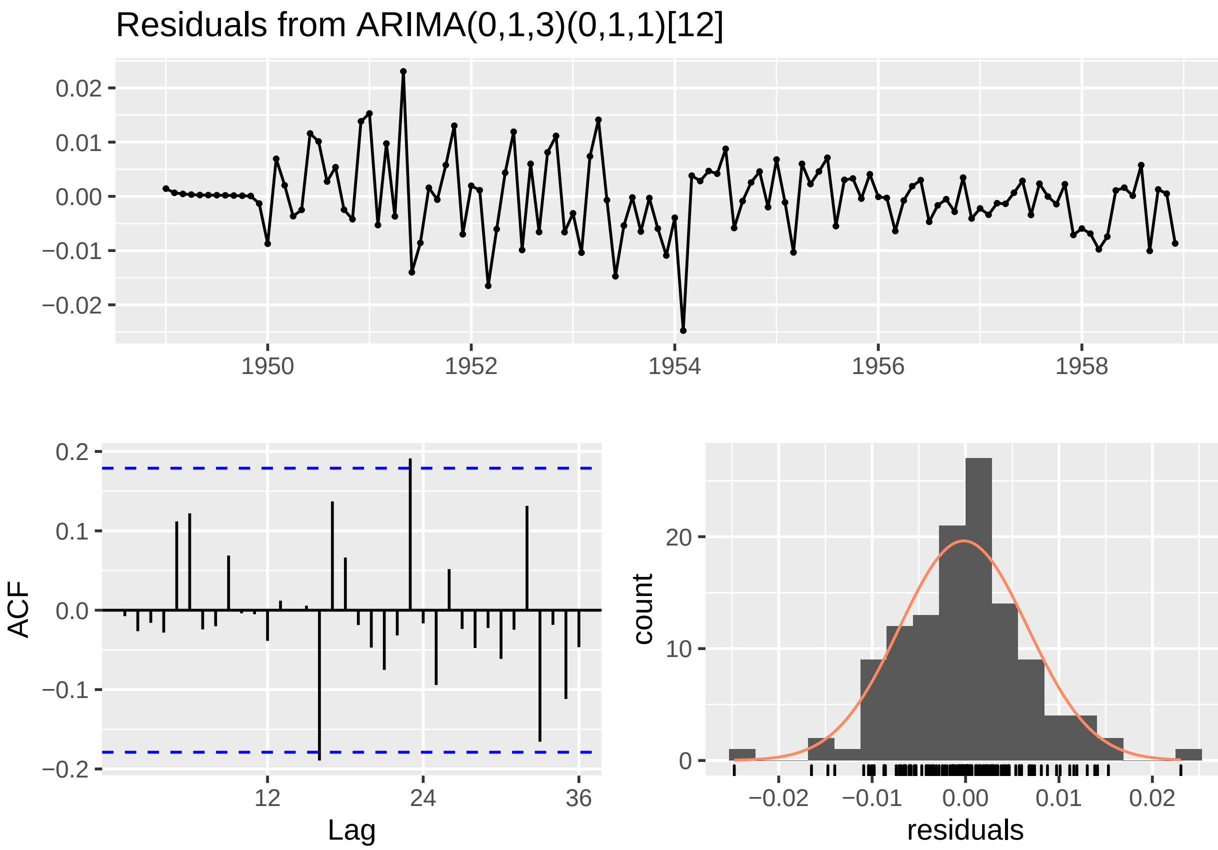
According to the ACF, we can think to a $SARIMA_{(0,1,3)(0,1,1)_{12}}$. The value of lambda will be estimated automatically

```
fit4=Arima(AP_train, order=c(0,1,3), seasonal=c(0,1,1),lambda = "auto")
fit4 %>% residuals() %>% ggtsdisplay()
```



The residuals are almost ok… We can check:

```
checkresiduals(fit4)
```



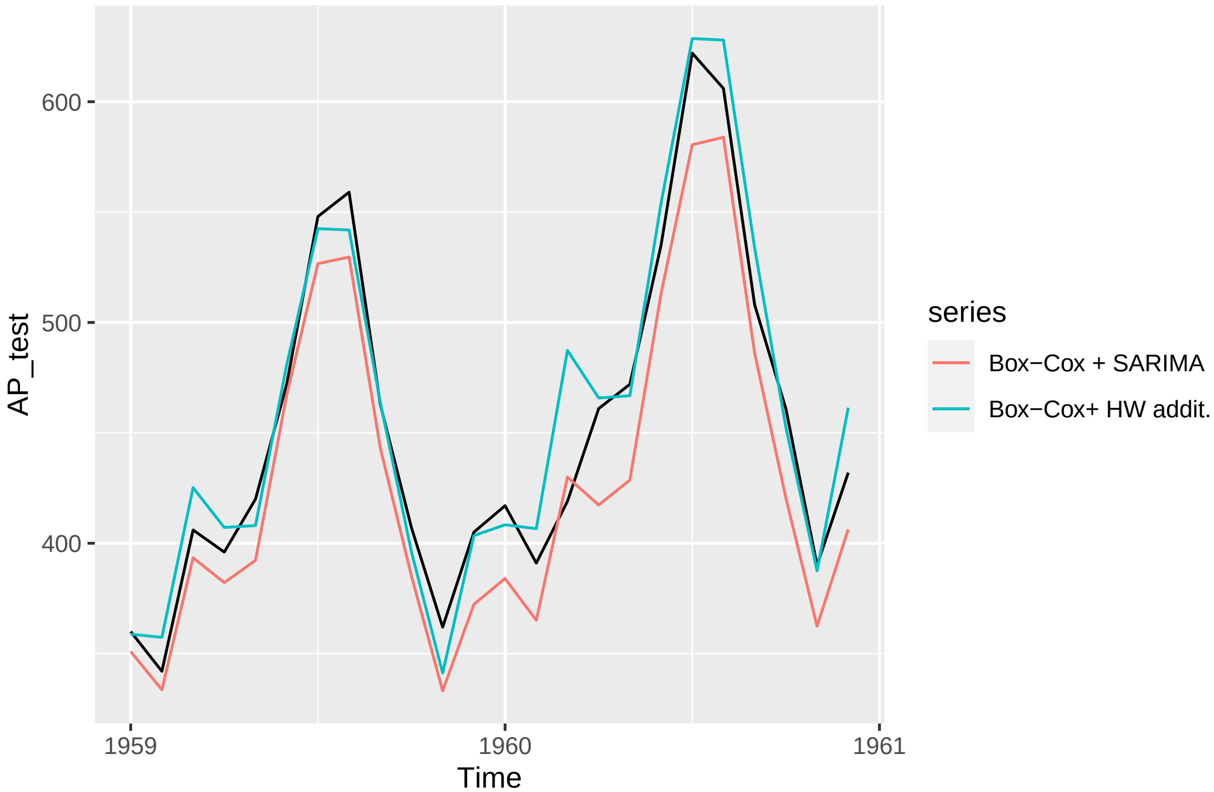

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,3)(0,1,1)[12]
## Q* = 19.992, df = 20, p-value = 0.4584
##
## Model df: 4.   Total lags used: 24
```

Let's forecast with this model

```
prev=forecast(fit4,h=24)
print(sqrt(mean((prev$mean-AP_test)^2)))
```

```
## [1] 26.83486
```

```
autoplot(AP_test)+autolayer(prev$mean,series="Box-Cox + SARIMA")+
  autolayer(fit2$mean,series="Box-Cox+ HW addit.")
```



We can also try another models, since some ACF are significant at lag 5, and PACF at lag 3 and 5:

```
fit5=Arima(AP_train, order=c(0,1,5), seasonal=c(0,1,1),lambda = "auto")
prev=forecast(fit5,h=24)
print(sqrt(mean((prev$mean-AP_test)^2)))
```

```
## [1] 26.70868
```

```
fit5=Arima(AP_train, order=c(3,1,0), seasonal=c(0,1,1),lambda = "auto")
prev=forecast(fit5,h=24)
print(sqrt(mean((prev$mean-AP_test)^2)))
```

```
## [1] 33.89911
```

```
fit5=Arima(AP_train, order=c(5,1,0), seasonal=c(0,1,1),lambda = "auto")
prev=forecast(fit5,h=24)
print(sqrt(mean((prev$mean-AP_test)^2)))
```

```
## [1] 25.58441
```