Identification of Seasonal Models

Matija Jakovac

2025-03-07

First series - AturMas: Number of men registered as unemployed in SEPE offices since January 1996

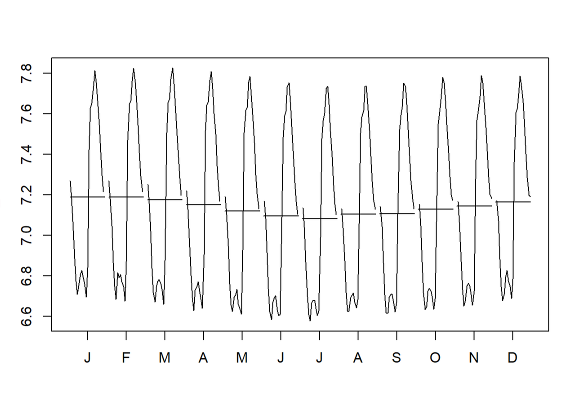
1. Load the file containing the series. Define the read data as an object of type ts (time series), specifying the origin and frequency of the series.

Time series start in 1996 January with yearly frequency (**s=12** months).

2. Create a graphical representation of the time series. Describe the most relevant aspects observed at first glance.

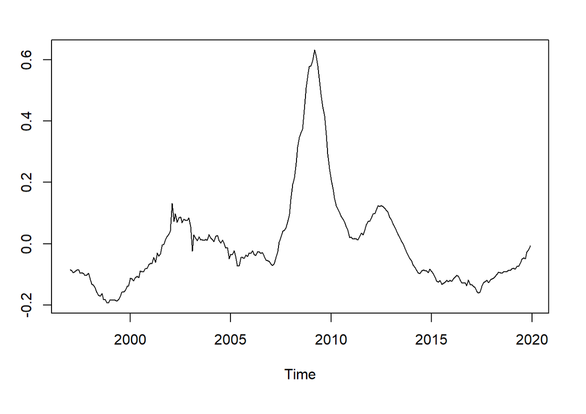
Above we can see two graphs, first one representing original time series, second one representing performed logarithmic operation on original time series.

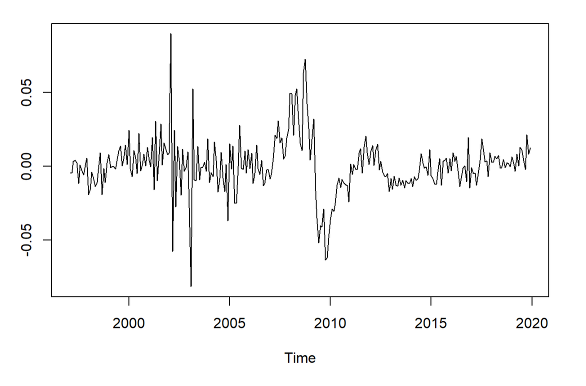
The graphs are very similar but the second one has lower values on y-axis because of performed logarithmic operation. They both show decrease in unemployment from 1996 to 2008 and then large increase from 2013 to 2018 which corresponds with 2008 global financial crisis. From then till the 2020 the economy is stabilizing and unemployment is decreasing. From those two graphs we can assume there is some seasonal pattern of regular up-and-down pattern.



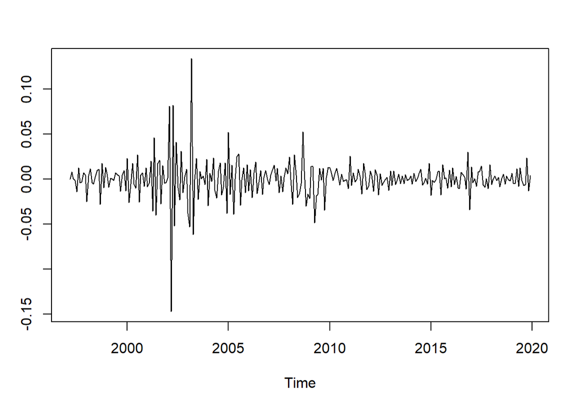
For confirmation we configure month plot in which we can clearly see there is a seasonal pattern where unemployment rises in specific months during the start of the year representing start of the year layoffs. Unemployment is lower during the summer probably corresponding to the various summer jobs.

3. Apply appropriate transformations to make the series stationary. Justify your decisions.





First plot is showing time series after performing seasonal differencing and we can clearly see that the series is not stationary and we must perform regular (non-seasonal) differencing. The second graph shows that and there we can see that the mean is constant and we can assume that the series is stationary. We see that the mean is not constant around 2008 when the financial crisis happened but all other years apart from that unusual activity are constant.



Here we can see the plot if we perform another regular difference which then happens to have a constant mean even during financial crisis.

```
## Variance of log-transformed original (d=0,D=0) series: 0.1619848
```

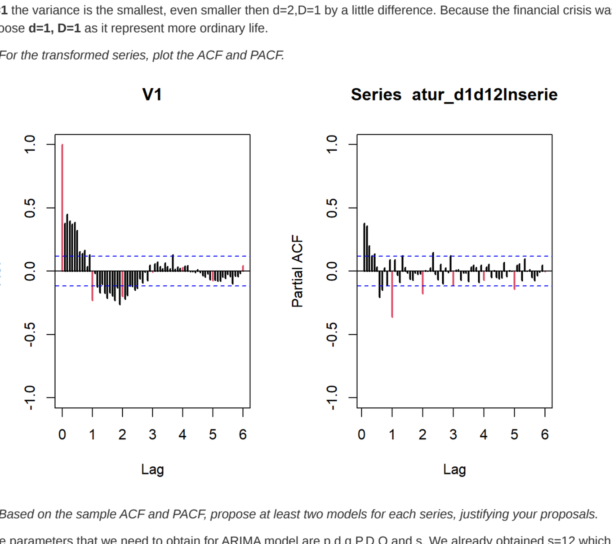
```
## Variance of after seasonal differencing (d=0,D=1) at lag=12: 0.02531255
```

```
## Variance of after first non-seasonal (d=1,D=1) differencing: 0.0003577561
```

```
## Variance of after additional non-seasonal (d=2,D=1) differencing: 0.0004449461
```

Another way to see which transformations we needed to perform to compare variances for different transformations. Here we can see that for **d=1, D=1** the variance is the smallest, even smaller then d=2,D=1 by a little difference. Because the financial crisis was something unordinary I will choose **d=1, D=1** as it represent more ordinary life.

4. For the transformed series, plot the ACF and PACF.



5. Based on the sample ACF and PACF, propose at least two models for each series, justifying your proposals.

The parameters that we need to obtain for ARIMA model are p,d,q,P,D,Q and s. We already obtained s=12 which is the frequency of seasonal part. Using transformations we obtained d=1 and D=1 which is previously justified above. Based on ACF and PACF we can obtain/propose remaining parameters

SEASONAL COMPONENT $ARMA(P,Q)_{12}$: We will focus only on the red-colored lags and apply the identification criteria used for standard lags.

ACF: We can see that last red-colored lag is at lag=2 that crosses blue line which represents confidence band so I can propose *MA(2)*.

PACF: I can propose *AR(5)* as the fifth red line(lag=5) is over the confidence band (blue line) by a little bit. We could also propose *AR(2)* as the second lag is over the band which gives fewer arguments for model.

REGULAR COMPONENT ARMA(p,q): We will focus only on the first 5-6 lags given the seasonality is 12. We must keep in mind that near the multiples of the seasonality (lags of order 12, 24, 36, etc.), there may be significant satellite lags that should not be considered for identification.

ACF: I can propose *MA(6)* because first 6 black lags are over the confidence band.

PACF: I can propose *AR(5)* because the fifth lag is over the band by only a little bit. Because of that if we want to have less parameters we can propose *AR(3)* as the third lag crosses the band greater then the fifth one.

Because p,q>2 we can also propose *ARMA(1,1)* because it has less parameters as well.

With that I can propose models with d=1, D=1 and regular comp. of AR(5), MA(6) or ARMA(1,1) and seasonal comp. of AR(5) and MA(2) of which I will take two models.

Model 1) $ARIMA(5,1,0)(0,1,2)_{12}$ where I took AR(6) for regular comp. and MA(2) for seasonal comp.

Model 2) $ARIMA(0,1,6)(5,1,0)_{12}$ where I took MA(6) for regular comp. and AR(5) for seasonal comp.

6. Estimate the proposed models and verify the significance of the coefficients, ensuring that the residuals have an ACF compatible with white noise. If any coefficient is not significant, remove it from the model.

Model 1) $ARIMA(5,1,0)(0,1,2)_{12}$

Firstly, we specify the transformed stationary series (W_t) to obtain mean estimation.

```
##
## Call:
## arima(x = atur_d1d12lnserie, order = c(5, 0, 0), seasonal = list(order = c(0,
##     0, 2), period = 12))
##
## Coefficients:
##          ar1     ar2     ar3     ar4     ar5     sma1     sma2  intercept
##       0.2068  0.1888  0.1253  0.1559  0.1251  -0.5478  -0.1293     0.0002
## s.e.  0.0618  0.0612  0.0612  0.0612  0.0597   0.0641   0.0686     0.0015
##
## sigma^2 estimated as 0.0001899:  log likelihood = 784.66,  aic = -1551.33
```

By looking at intercept we can see that u=0.0002 while S_u=0.0015. We can perform t-test do prove if mean is significant or not. For t-test we can following hypothesis:

H_0: u_wt = 0

H_1: u_wt != 0

t = u/S_u

abs(t) > 2 => H1, abs(t) <= 2 => H0

In this case abs(t)=0.13 which means we keep H0 therefore mean is not significant and we can re-estimate model with log(X_t)

```
##
## Call:
## arima(x = atur_lnserie, order = c(5, 1, 0), seasonal = list(order = c(0, 1,
##     2), period = 12))
##
## Coefficients:
##          ar1     ar2     ar3     ar4     ar5     sma1     sma2
##       0.2070  0.1889  0.1252  0.1561  0.1251  -0.5480  -0.1294
## s.e.  0.0618  0.0611  0.0612  0.0612  0.0597   0.0641   0.0686
##
## sigma^2 estimated as 0.0001899:  log likelihood = 784.67,  aic = -1553.34
```

On this model we can also perform t-test for every coefficient to see if they are significant or not. We can see that maybe sma2 coeff. is not significant as absolute t-value for it is near 2.

```
##
## Call:
## arima(x = atur_lnserie, order = c(5, 1, 0), seasonal = list(order = c(0, 1,
##     1), period = 12))
##
## Coefficients:
##          ar1     ar2     ar3     ar4     ar5     sma1
##       0.2201  0.1842  0.1316  0.1517  0.1235  -0.5861
## s.e.  0.0612  0.0612  0.0611  0.0612  0.0597   0.0650
##
## sigma^2 estimated as 0.0001934:  log likelihood = 782.85,  aic = -1551.71
```

Removing that sma2 coeff. we get model with greater AIC which means that previous model was better but this one has less parameters which is always good for model but I will strictly follow AIC metric and will chose the previous model $ARIMA(5,1,0)(0,1,2)_{12}$ *with AIC=-1553.34*.

Model 2) $ARIMA(0,1,6)(5,1,0)_{12}$

Firstly, we specify the transformed stationary series (W_t) to obtain mean estimation.

```
##
## Call:
## arima(x = atur_d1d12lnserie, order = c(0, 0, 6), seasonal = list(order = c(5,
##     0, 0), period = 12))
##
## Coefficients:
##          ma1     ma2     ma3     ma4     ma5     ma6     sar1     sar2     sar3
##       0.2598  0.2944  0.2441  0.3474  0.2671  0.2761  -0.5254  -0.3458  -0.2473
## s.e.  0.0637  0.0602  0.0584  0.0696  0.0710  0.0662   0.0668   0.0708   0.0725
##          sar4     sar5  intercept
##       -0.1831  -0.1529      2e-04
## s.e.   0.0700   0.0600      1e-03
##
## sigma^2 estimated as 0.0001912:  log likelihood = 783.89,  aic = -1541.77
```

By looking at intercept we can see that u=0.0002 while S_u=0.001. We can perform t-test do prove if mean is significant or not. It has same hypothesis as before and t=0.2. In this case abs(t)=0.2 which means we keep H0 therefore mean is not significant and we can re-estimate model with log(X_t)

```
##
## Call:
## arima(x = atur_lnserie, order = c(0, 1, 6), seasonal = list(order = c(5, 1,
##     0), period = 12))
##
## Coefficients:
##          ma1     ma2     ma3     ma4     ma5     ma6     sar1     sar2     sar3
##       0.2600  0.2945  0.2442  0.3474  0.2672  0.2762  -0.5255  -0.3458  -0.2475
## s.e.  0.0637  0.0602  0.0584  0.0696  0.0710  0.0662   0.0668   0.0708   0.0726
##          sar4    sar5
##       -0.1833  -0.153
## s.e.   0.0700   0.060
##
## sigma^2 estimated as 0.0001912:  log likelihood = 783.88,  aic = -1543.76
```

On this model we can also perform t-test for every coefficient to see if they are significant or not. We can see compute that for every coeff. absolute t-value is greater then 2 so every one is significant. This model $ARIMA(0,1,6)(5,1,0)_{12}$ *has AIC=-1543.76*.

Previously, I decided to take d=1 instead of d=2. Here we cannot see the estimation of models with d=2 but when estimating them with corresponding p,q,P and Q they give greater AIC which means that the choice of d=1 was a good choice.

7. Indicate which model you would propose, using the AIC criterion.

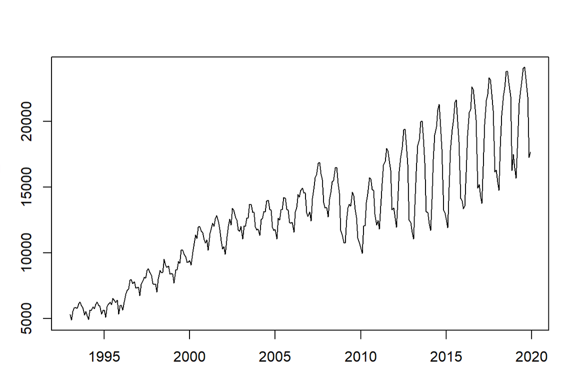
We proposed two models, model 1 - $ARIMA(5,1,0)(0,1,2)_{12}$ with corresponding AIC=-1553.34 and model 2 $ARIMA(0,1,6)(5,1,0)_{12}$ with corresponding AIC=-1543.76. We can conclude that model 1 has lower AIC which means model performs better and I would choose that model. In conclusion for AturMAS series we choose ARIMA(p,d,q)(P,D,Q)_s model with **p=5,d=1,q=0,P=0,D=1,Q=2 and s=12** - $ARIMA(5,1,0)(0,1,2)_{12}$

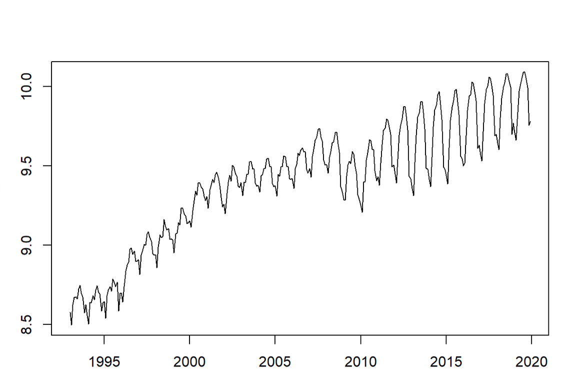
Second series - AeronausBCN: Number of monthly international flight aircraft at Barcelona-Prat (BCN) airport since January 1993.

1. Load the file containing the series. Define the read data as an object of type ts (time series), specifying the origin and frequency of the series.

Time series start in 1993 January with yearly frequency (**s=12** months). The end of this series I will put at 2019 because of COVID pandemic which was in 2020. With having pandemic we would include some parameter which normally would not be present.

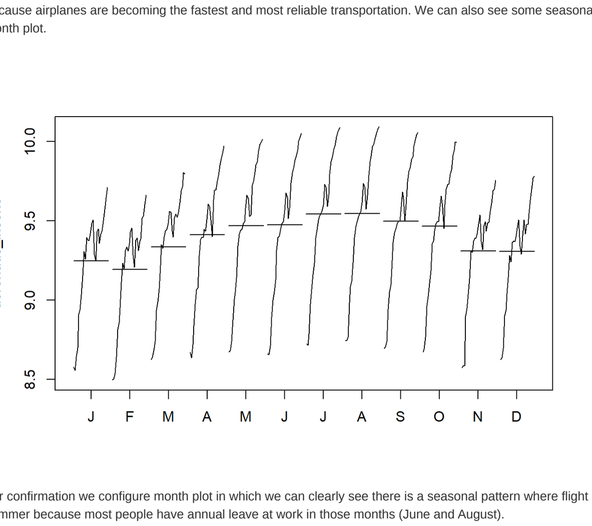
2. Create a graphical representation of the time series. Describe the most relevant aspects observed at first glance.





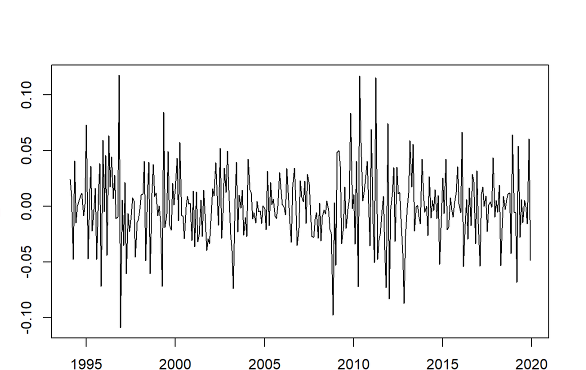
Above we can see two graphs, first one representing original time series, second one representing performed logarithmic operation on original time series.

The graphs are very similar but the second one has lower values on y-axis because of performed logarithmic operation. They both show decrease in flights around 2010 which could correspond to the global financial crisis in 2008. But the trend is that people are flying more which is expected because airplanes are becoming the fastest and most reliable transportation. We can also see some seasonal pattern which we will proof with month plot.



For confirmation we configure month plot in which we can clearly see there is a seasonal pattern where flight rise in specific months during the summer because most people have annual leave at work in those months (June and August).

3. Apply appropriate transformations to make the series stationary. Justify your decisions.



First plot is showing time series after performing seasonal differencing and we can clearly see that the series is not stationary and we must perform regular (non-seasonal) differencing. The second graph shows that and there we can see that the mean is constant and we can assume that the series is stationary.

```
## Variance of log-transformed original (d=0,D=0) series: 0.1515852
```

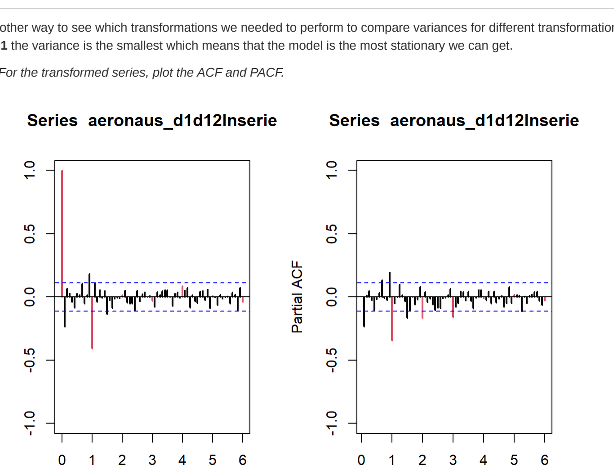
```
## Variance of after seasonal differencing (d=0,D=1) at lag=12: 0.005081572
```

```
## Variance of after first non-seasonal (d=1,D=1) differencing: 0.001083066
```

```
## Variance of after additional non-seasonal (d=2,D=1) differencing: 0.002673263
```

Another way to see which transformations we needed to perform to compare variances for different transformations. Here we can see that for **d=1, D=1** the variance is the smallest which means that the model is the most stationary we can get.

4. For the transformed series, plot the ACF and PACF.



5. Based on the sample ACF and PACF, propose at least two models for each series, justifying your proposals.

The parameters that we need to obtain for ARIMA model are p,d,q,P,D,Q and s. We already obtained s=12 which is the frequency of seasonal part. Using transformations we obtained d=1 and D=1 which is previously justified above. Based on ACF and PACF we can obtain/propose remaining parameters

SEASONAL COMPONENT $ARMA(P,Q)_{12}$: We will focus only on the red-colored lags and apply the identification criteria used for standard lags.

ACF: We can see that last red-colored lag is at lag=1 that crosses blue line which represents confidence band so I can propose *MA(1)*.

PACF: I can propose *AR(3)* as the third red line(lag=3) is over the confidence band (blue line).

We could not propose $ARMA(P,Q)_{12}$ for seasonal comp. because it would have more components the MA(1)

REGULAR COMPONENT ARMA(p,q): We will focus only on the first 5-6 lags given the seasonality is 12. We must keep in mind that near the multiples of the seasonality (lags of order 12, 24, 36, etc.), there may be significant satellite lags that should not be considered for identification.

ACF: I can propose *MA(1)* because only the first black lag is over the confidence band.

PACF: I can propose *AR(1)* because only the first lag is over the band. We can maybe propose *AR(5)* but the lag is not fully crossing the band and probably the coefficient and the ones before that one would be insignificant.

We could not propose ARMA(p,q) for regular comp. because it would have more components the MA(1) or AR(1)

With that I can propose models with d=1, D=1 and regular comp. of AR(1), MA(1) and seasonal comp. of AR(3) and MA(1) of which I will take two models.

Model 1) $ARIMA(1,1,0)(0,1,1)_{12}$ where I took AR(1) for regular comp. and MA(1) for seasonal comp.

Model 2) $ARIMA(0,1,1)(3,1,0)_{12}$ where I took MA(1) for regular comp. and AR(3) for seasonal comp.

6. Estimate the proposed models and verify the significance of the coefficients, ensuring that the residuals have an ACF compatible with white noise. If any coefficient is not significant, remove it from the model.

Model 1) $ARIMA(1,1,0)(0,1,1)_{12}$

Firstly, we specify the transformed stationary series (W_t) to obtain mean estimation.

```
##
## Call:
## arima(x = aeronaus_d1d12lnserie, order = c(1, 0, 0), seasonal = list(order = c(0,
##     0, 1), period = 12))
##
## Coefficients:
##           ar1     sma1  intercept
##       -0.1533  -0.4600      0e+00
## s.e.   0.0576   0.0485      8e-04
##
## sigma^2 estimated as 0.0008293:  log likelihood = 660.53,  aic = -1313.05
```

By looking at intercept we can see that u=0.0002 while S_u=0.0015. We can perform t-test do prove if mean is significant or not. For t-test we can following hypothesis:

H_0: u_wt = 0

H_1: u_wt != 0

t = u/S_u

abs(t) > 2 => H1, abs(t) <= 2 => H0

In this case abs(t)=0 which means we keep H0 therefore mean is not significant and we can re-estimate model with log(X_t).

```
##
## Call:
## arima(x = aeronaus_lnserie, order = c(1, 1, 0), seasonal = list(order = c(0,
##     1, 1), period = 12))
##
## Coefficients:
##           ar1     sma1
##       -0.1533  -0.4600
## s.e.   0.0576   0.0485
##
## sigma^2 estimated as 0.0008293:  log likelihood = 660.53,  aic = -1315.06
```

On this model we can also perform t-test for every coefficient to see if they are significant or not. We can see compute that for every coeff. absolute t-value is greater then 2 so every one is significant. This model $ARIMA(1,1,0)(0,1,1)_{12}$ *has AIC=-1315.06*.

Model 2) $ARIMA(0,1,1)(3,1,0)_{12}$

```
##
## Call:
## arima(x = aeronaus_d1d12lnserie, order = c(0, 0, 1), seasonal = list(order = c(3,
##     0, 0), period = 12))
##
## Coefficients:
##           ma1     sar1     sar2     sar3  intercept
##       -0.1280  -0.4996  -0.2305  -0.1292      0e+00
## s.e.   0.0538   0.0591   0.0646   0.0604      8e-04
##
## sigma^2 estimated as 0.0008206:  log likelihood = 661.86,  aic = -1311.72
```

By looking at intercept we can see that u=0 while S_u=0.0008. We can perform t-test do prove if mean is significant or not. It has same hypothesis as before and absolute t=0. In this case abs(t)=0 which means we keep H0 therefore mean is not significant and we can re-estimate model with log(X_t).

```
##
## Call:
## arima(x = aeronaus_lnserie, order = c(0, 1, 1), seasonal = list(order = c(3,
##     1, 0), period = 12))
##
## Coefficients:
##           ma1     sar1     sar2     sar3
##       -0.1280  -0.4996  -0.2305  -0.1293
## s.e.   0.0538   0.0591   0.0645   0.0604
##
## sigma^2 estimated as 0.0008205:  log likelihood = 661.87,  aic = -1313.74
```

On this model we can also perform t-test for every coefficient to see if they are significant or not. We can see compute that for every coeff. absolute t-value is greater then 2 so every one is significant. I will exclude sar3 coeff. and perform a estimation again because for that coeff. abs(t) value is near 2 but slightly greater.

```
##
## Call:
## arima(x = aeronaus_lnserie, order = c(0, 1, 1), seasonal = list(order = c(2,
##     1, 0), period = 12))
##
## Coefficients:
##           ma1     sar1     sar2
##       -0.1329  -0.4778  -0.1717
## s.e.   0.0535   0.0583   0.0585
##
## sigma^2 estimated as 0.0008342:  log likelihood = 659.6,  aic = -1311.2
```

We can see that the this model with AR(2) for seasonal comp. is not better than previous one as AIC is greater. Strictly following AIC metric I choose the model with AR(3) with 3 coefficients for seasonal component concluding $ARIMA(0,1,1)(3,1,0)_{12}$ *which has AIC=-1313.74.*

7. Indicate which model you would propose, using the AIC criterion.

We proposed two models, model 1 - $ARIMA(1,1,0)(0,1,1)_{12}$ with corresponding AIC=-1315.06 and model 2 $ARIMA(0,1,1)(3,1,0)_{12}$ with corresponding AIC=-1313.74. We can conclude that model 1 has lower AIC which means model performs better and I would choose that model. In conclusion for AeronausBCN series we choose ARIMA(p,d,q)(P,D,Q)_s model with **p=1,d=1,q=0,P=0,D=1,Q=1 and s=12** - $ARIMA(1,1,0)(0,1,1)_{12}$