Module 8 - Seasonal Models

MATH1318 Time Series Analysis

Prepared by: Dr. Haydar Demirhan based on the textbook by Cryer and Chan, Time Series Analysis with R, Springer, 2008.

Nomenclature

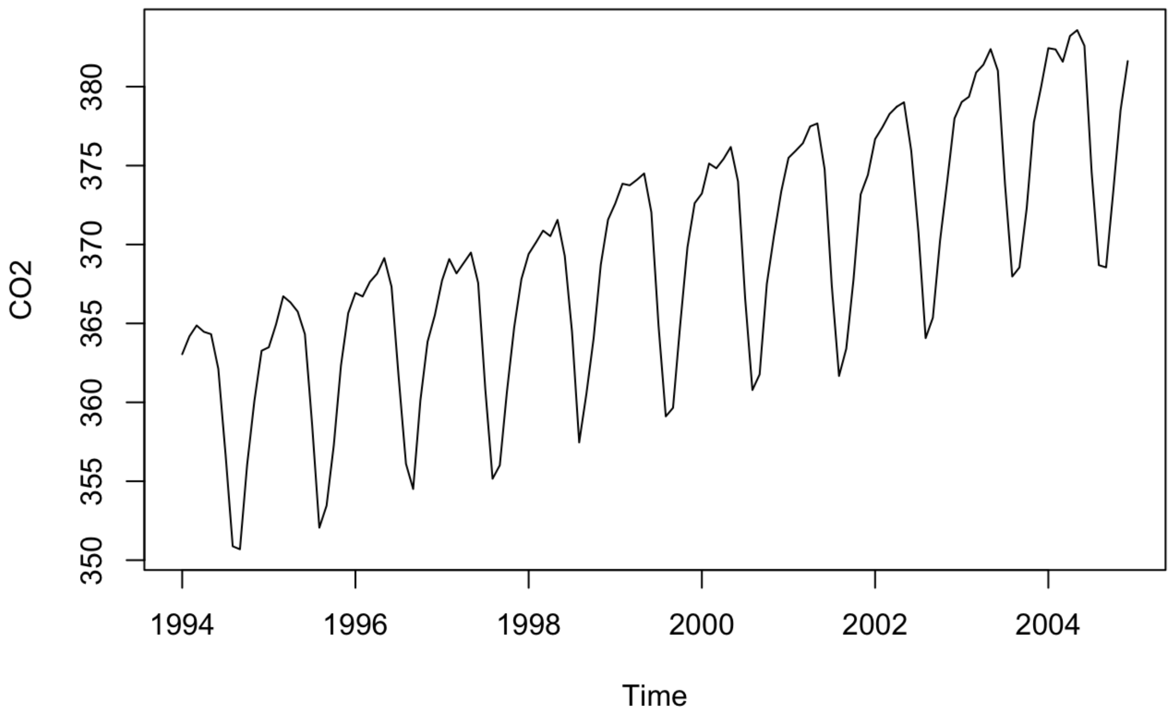
$\Theta_.$: Seasonal moving average parameter.
$\Phi_.$: Seasonal autoregressive parameter.
s: Seasonal period.
$\nabla_s^D Y_t$: the dth seasonal difference of period s.
S_t: Seasonal random walk series.
ϵ_t: White noise series.
ξ_t: White noise series.
$\ll$: Too much less than.
SARIMA(p,d,q)$\times$(P,D,Q)$_s$: Seasonal integrated autoregressive moving average process if reqular orders p, d, and q, and seasonal ordes P, D, and Q.

Introduction

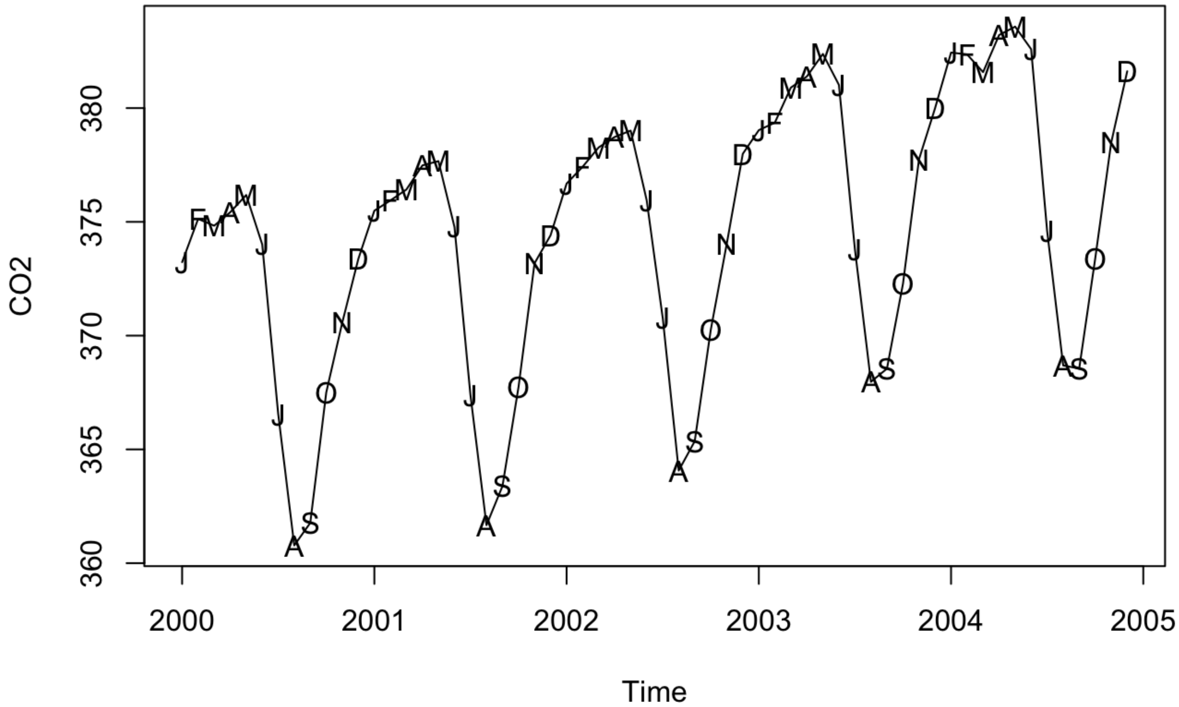
In this module, we will focus on modelling seasonal stochastic trends. For the time series used, particularly in business and economics, the assumption of any deterministic trend is quite suspect even though cyclical tendencies are very common in such series.

As an example, levels of carbon dioxide (CO_2) are monitored at several sites around the world to investigate atmospheric changes. One of the sites is at Alert, Northwest Territories, Canada, near the Arctic Circle. The following time series plot shows the monthly CO_2 levels from January 1994 through December 2004.

```
data(co2)
plot(co2,ylab='CO2')
```




```
plot(window(co2,start=c(2000,1)),ylab='CO2')
Month=c('J','F','M','A','M','J','J','A','S','O','N','D')
points(window(co2,start=c(2000,1)),pch=Month)
```



There is a strong upward trend but also a seasonality is present. As we see in the displays, carbon dioxide levels are higher during the winter months and much lower in the summer. Deterministic seasonal models such as seasonal means plus linear time trend or sums of cosine curves at various frequencies plus linear time trend could certainly be considered here. But we discover that such models do not explain the behavior of this time series. For this series and many others, it can be shown that the residuals from a seasonal means plus linear time trend model are highly autocorrelated at many lags. In contrast, we will see that the stochastic seasonal models work well for this series.

Seasonal ARIMA Models

We will start with stationary models and continue with nonstationary generalizations. In module, s will denote the known seasonal period; for monthly series $s = 12$ and for quarterly series $s = 4$.

Let's consider the time series generated from the model

$$Y_t = e_t - \Theta e_{t-12}.$$

Here $Cov(Y_t, Y_{t-1}) = 0$. But $Cov(Y_t, Y_{t-12}) = -\Theta\sigma_e^2$. This series is stationary and has nonzero autocorrelations only at lag 12. Thus, we can define a **seasonal MA(Q) model of order Q with seasonal period** s by

$$Y_t = e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s} - \cdots - \Theta_Q e_{t-Qs}.$$

The seasonal series defined in this way is always stationary and that the autocorrelation function will be nonzero only at the seasonal lags of $s, 2s, 3s, \dots, Qs$ The autocorrelation at lag k is

$$\rho_{ks} = \frac{-\Theta_k + \Theta_1\Theta_{k+1} + \Theta_2\Theta_{k+2} + \cdots + \Theta_{Q-k}\Theta_Q}{1 + \Theta_1^2 + \Theta_2^2 + \cdots + \Theta_Q^2}$$

for $k = 1, 2, \ldots, Q$.

It is useful to note that the seasonal MA(Q) model can also be viewed as a special case of a nonseasonal MA model of order $q = Qs$ but with all θ-values zero except at the seasonal lags $s, 2s, 3s, \ldots, Qs$

Seasonal autoregressive models can also be defined. Consider the following model:

$$Y_t = \Phi Y_{t-12} + e_t,$$

where $|\phi| < 1$, which ensures stationarity, and e_t is independent of $Y_{t-1}, Y_{t-2}, \ldots$. For this model, $\rho_k = 0$ and $\rho_k = \Phi\rho_{k-12}$ for $k \geq 1$.

A seasonal AR(P) model of order P and seasonal period s is given by

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \cdots + \Phi_P Y_{t-Ps} + e_t.$$

For this model, e_t is independent of $Y_{t-1}, Y_{t-2}, \ldots$ and roots of the characteristic function have to be greater than 1 in absolute value. Also, the seasonal AR model can be written as a special AR(p) model of order $p = Ps$ with nonzero Ψ-coefficients only at the seasonal lags $s, 2s, 3s, \ldots, Ps$

The autocorrelation function is nonzero only at lags $s, 2s, 3s, \ldots$, where it behaves like a combination of decaying exponentials and damped sine functions.

For the general seasonal AR(1) model,

$$\rho_{ks} = \Phi^k$$

for $k = 1, 2, \ldots$.

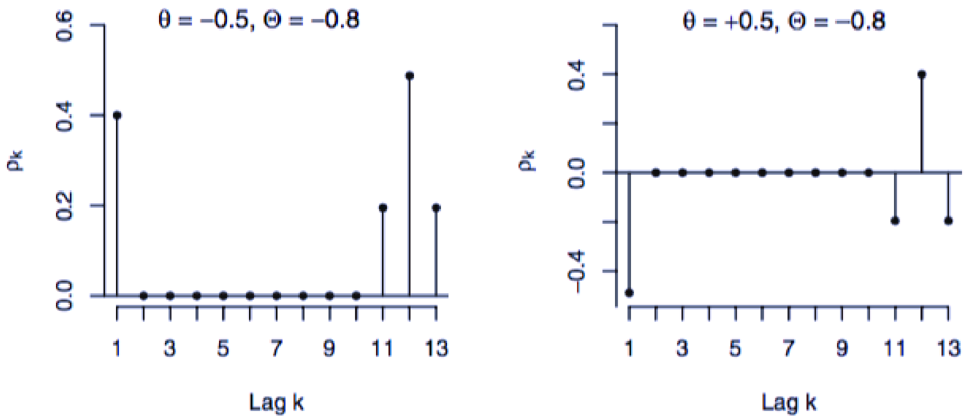
Multiplicative Seasonal ARMA Models

Multiplicative seasonal ARMA models contain autocorrelation for the seasonal lags but also for low lags of neighboring series values.

Consider an MA model with characteristic function $(1 - \theta x)(1 - \Theta x^{12}) = 1 - \theta x - \Theta x^{12} + \theta\Theta x^{13}$. Thus the corresponding time series satisfies

$$Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13}.$$

For this model, the autocorrelation function is nonzero only at lags 1, 11, 12, and 13 and shown below.



In general, we define a multiplicative seasonal ARMA(p,q)×(P,Q)s model with seasonal period s as a model with AR characteristic polynomial $\phi(x)\Phi(x)$ and MA characteristic polynomial $\theta(x)\Theta(x)$, where

$$\begin{aligned}\phi(x) &= 1-\phi_1 x-\phi_2 x^2-\cdots-\phi_p x^p\\ \Phi(x) &= 1-\Phi_1 x^s-\Phi_2 x^{2s}-\cdots-\Phi_p x^{Ps}\end{aligned}$$

and

$$\begin{aligned}\theta(x) &= 1-\theta_1 x-\theta_2 x^2-\cdots-\theta_q x^q\\ \Theta(x) &= 1-\Theta_1 x^s-\Theta_2 x^{2s}-\cdots-\Theta_Q x^{Qs}\end{aligned}.$$

Suppose $P = q = 1$ and $p = Q = 0$ with $s = 12$. Then the model is

$$Y_t = \Theta Y_{t-12} + e_t - \theta e_{t-1}.$$

For this model, we have

$$\gamma_1 = \Phi\gamma_{11} - \theta\sigma_e^2$$

and

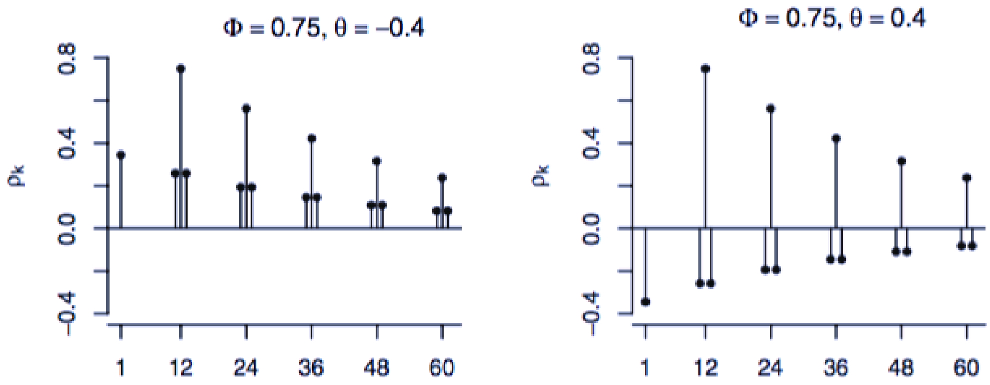
$$\gamma_k = \Phi\gamma_{k-12}.$$

Thus, we find

$$\begin{aligned}\gamma_0 &= \left[\frac{1+\theta^2}{1-\Phi^2}\right]\sigma_e^2\\ \rho_{12k} &= \Phi^k,\ \text{for } k \geq 1\\ \rho_{12k-1} &= \rho_{12k+1} = \left(-\frac{\theta}{1+\theta^2}\Phi^k\right)\end{aligned}$$

for $k = 0, 1, 2, \ldots$ and autocorrelations for all other lags equal to zero.

The following plots show the autocorrelation functions for two of these seasonal ARIMA processes with period 12.



The shape of these autocorrelations is somewhat typical of the sample autocorrelation functions for numerous seasonal time series.

Nonstationary Seasonal ARIMA Models

We will introduce **seasonal difference** to deal with nonstationary seasonal processes. The seasonal difference of period s for the series $\{Y_t\}$ is denoted $\nabla s Y_t$ and is defined as

$$\nabla_s Y_t = Y_t - Y_{t-s}.$$

For example, for monthly this way series we consider the changes from January to January, February to February, and so forth for successive years. Note that for a series of length n, the seasonal difference series will be of length $n - s$; that is, s data values are lost due to seasonal differencing.

As an example where seasonal differencing is appropriate, consider a process generated according to

$$Y_t = S_t + e_t$$

with

$$S_t = S_{t-s} + \epsilon_t$$

where $\{e_t\}$ and $\{\epsilon_t\}$ are independent white noise series. Notice that here S_t is seasonal random walk and if $\sigma_\epsilon \ll \sigma_e$, $\{S_t\}$ would model a slowly changing seasonal component. Because S_t is nonstationary, so does the original series Y_t. We get the following when we seasonally difference $\{Y_t\}$:

$$\begin{aligned}\nabla_s Y_t &= S_t - S_{t-s} + e_t - e_{t-s} \\ &= \epsilon_t + e_t - e_{t-s}.\end{aligned}$$

After the seasonal difference $\nabla_s Y_t$ is stationary and has the autocorrelation function of an MA(1)$_s$ model.

We can also write this model as a nonseasonal, slowly changing the stochastic trend. Consider

$$Y_t = M_t + S_t + e_t$$

with

$$S_t = S_{t-s} + \epsilon_t \text{ and } M_t = M_{t-1} + \xi_t$$

where $\{e_t\}$, $\{\epsilon_t\}$, and $\{\xi_t\}$ are independent white noise series. Taking both a seasonal difference and an ordinary nonseasonal difference we obtain that

$$\begin{aligned}\nabla\nabla_s Y_t &= \nabla(M_t - M_{t-t} + \epsilon_t + e_t - e_{t-s}) \\ &= (\xi_t + \epsilon_t + e_t) - (\epsilon_t + e_{t-1}) - (\xi_{t-s} + e_{t-s}) + e_{t-s-1}.\end{aligned}$$

This process is stationary and has nonzero autocorrelation only at lags $1, s-1, s$, and $s+1$, which agrees with the autocorrelation structure of the multiplicative seasonal model ARMA(0,1)×(0,1) with seasonal period s.

A process $\{Y_t\}$ is said to be a multiplicative seasonal ARIMA model with nonseasonal (regular) orders p, d, and q, seasonal orders P, D, and Q, and seasonal period s if the differenced series

$$W_t = \nabla^d \nabla_s^D Y_t$$

satisfies an ARMA(p,q)×(P,Q)$_s$ model with seasonal period s. In this case, we say that $\{Y_t\}$ is an ARIMA(p,d,q)×(P,D,Q)$_s$ model with seasonal period s.

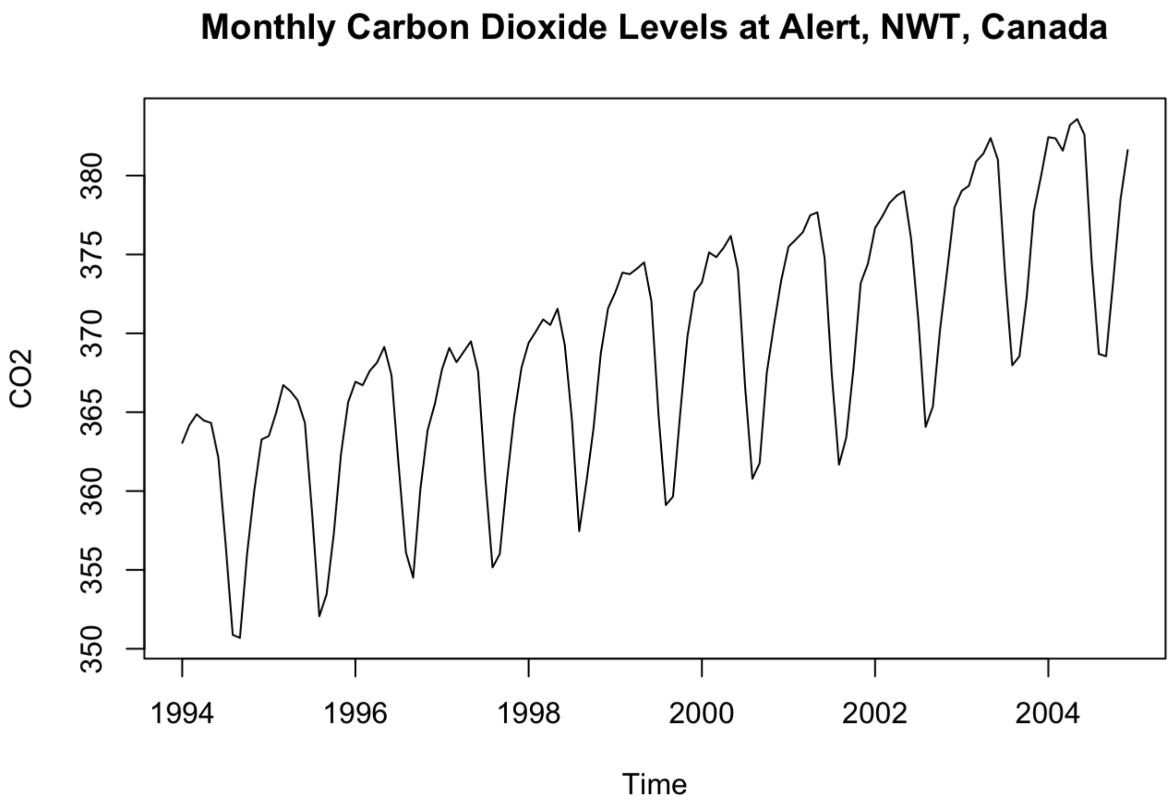
Model Specification, Fitting, and Checking

We will consider model specification, fitting and checking by paying special attention to seasonal components of the models.

Model Specification

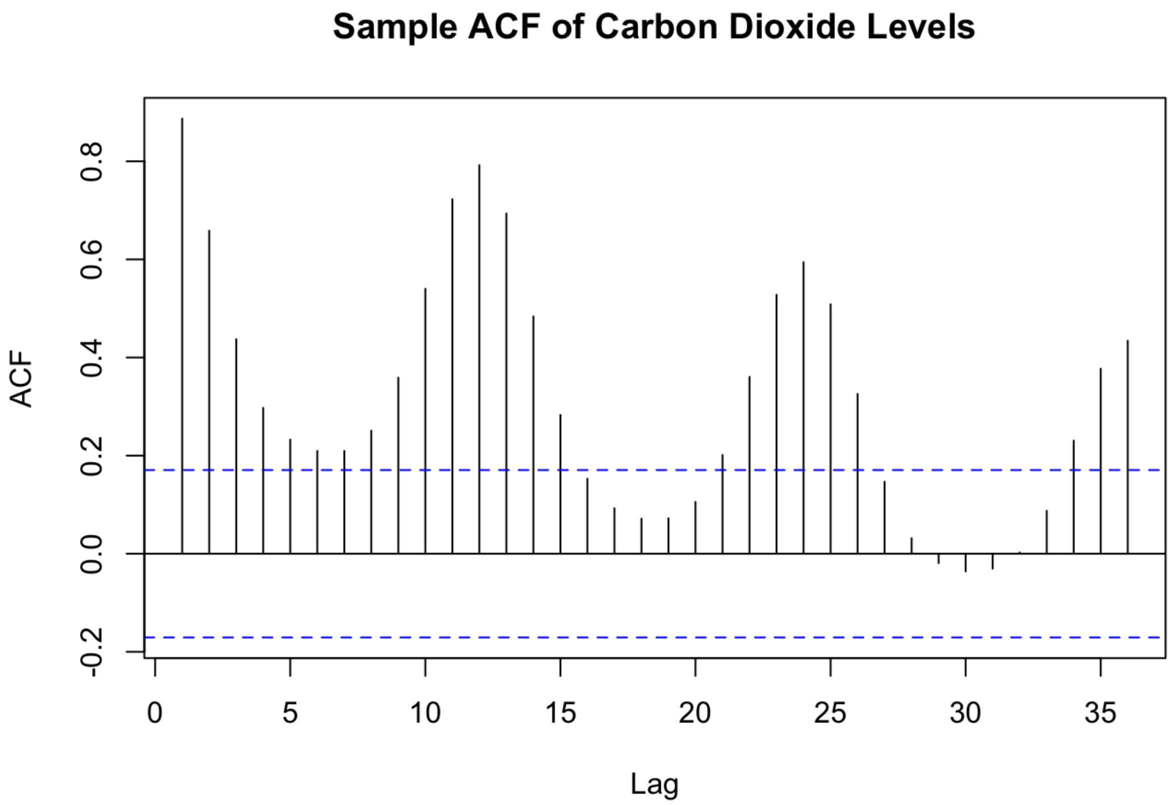
We follow the classical steps of model specification in time series. So, the first step is to have a look at the time series plot. For the monthly carbon dioxide levels in northern Canada series, we had the following time series plot.

```
data(co2)
plot(co2,ylab='CO2', main = "Monthly Carbon Dioxide Levels at Alert, NW
  T, Canada")
```

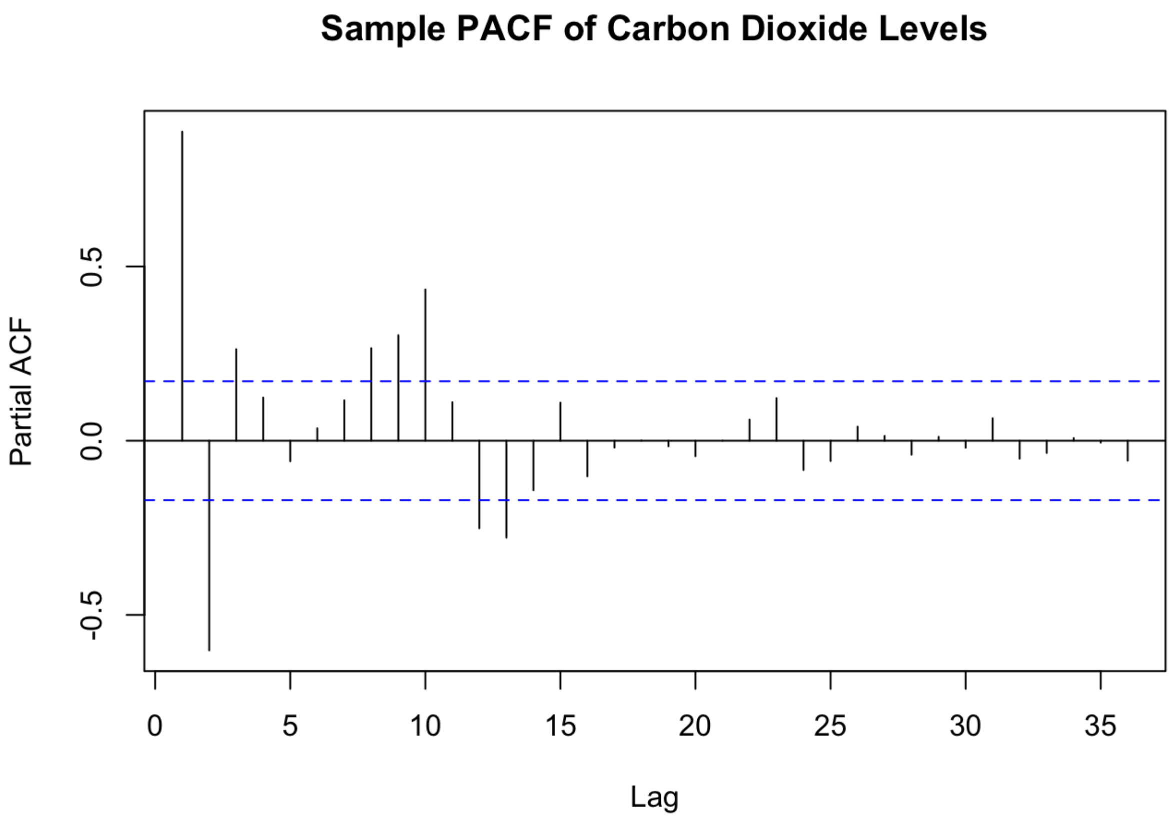


We observe a nonstationary model due to the upward trend. Then, we plot the ACF and PACF as below.

```
acf(as.vector(co2),lag.max=36, main = "Sample ACF of Carbon Dioxide Lev
  els")
```



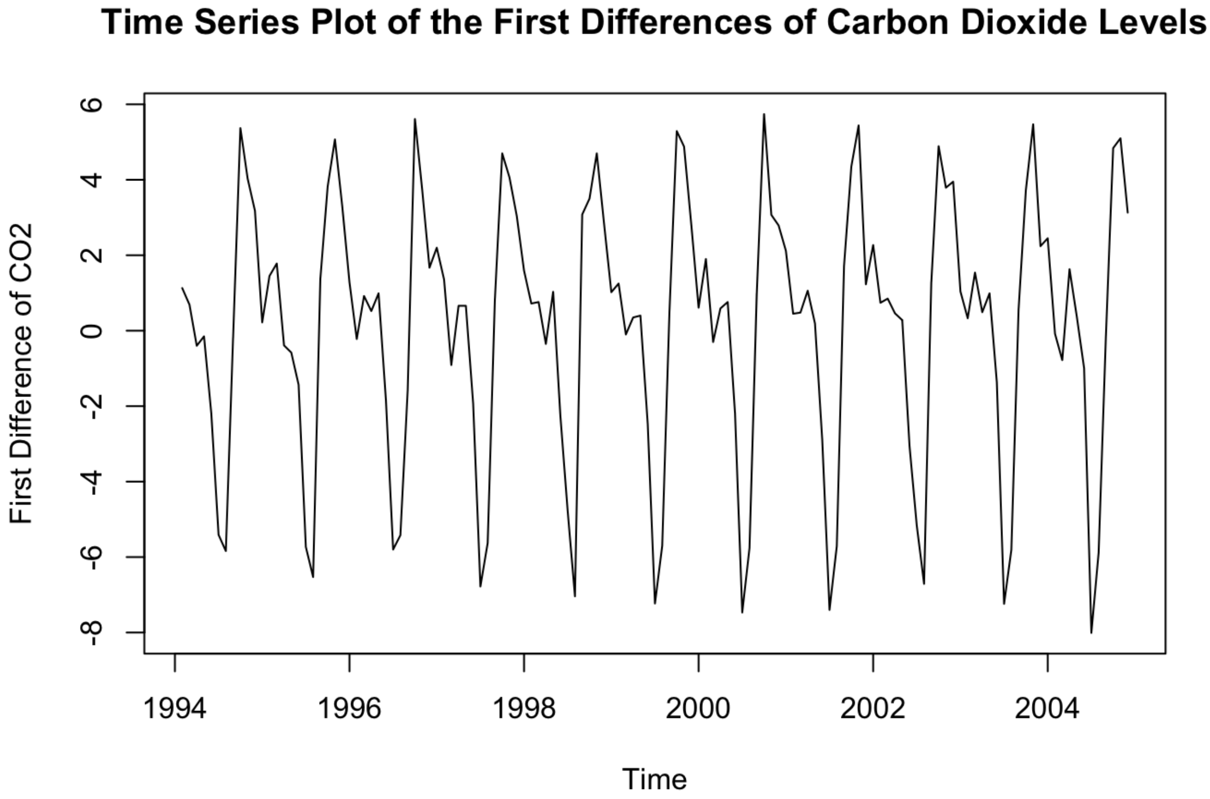
```
pacf(as.vector(co2),lag.max=36, main = "Sample PACF of Carbon Dioxide L
  evels")
```



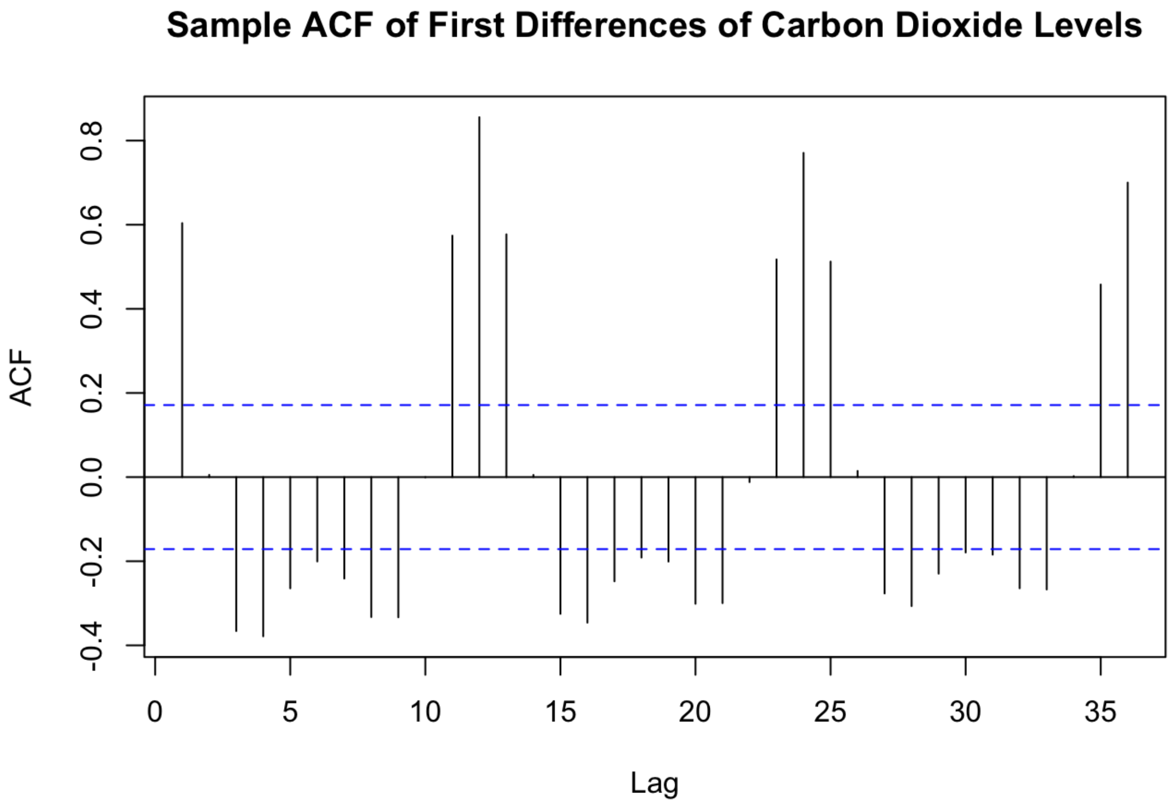
Because we have strong correlations at lags 12, 24, 36, and so on, we consider the existence of seasonal autocorrelation relationships. There is also a substantial correlation that needs to be considered.

Then, we examine time series, ACF, and PACF plots of the CO_2 levels after taking the first difference.

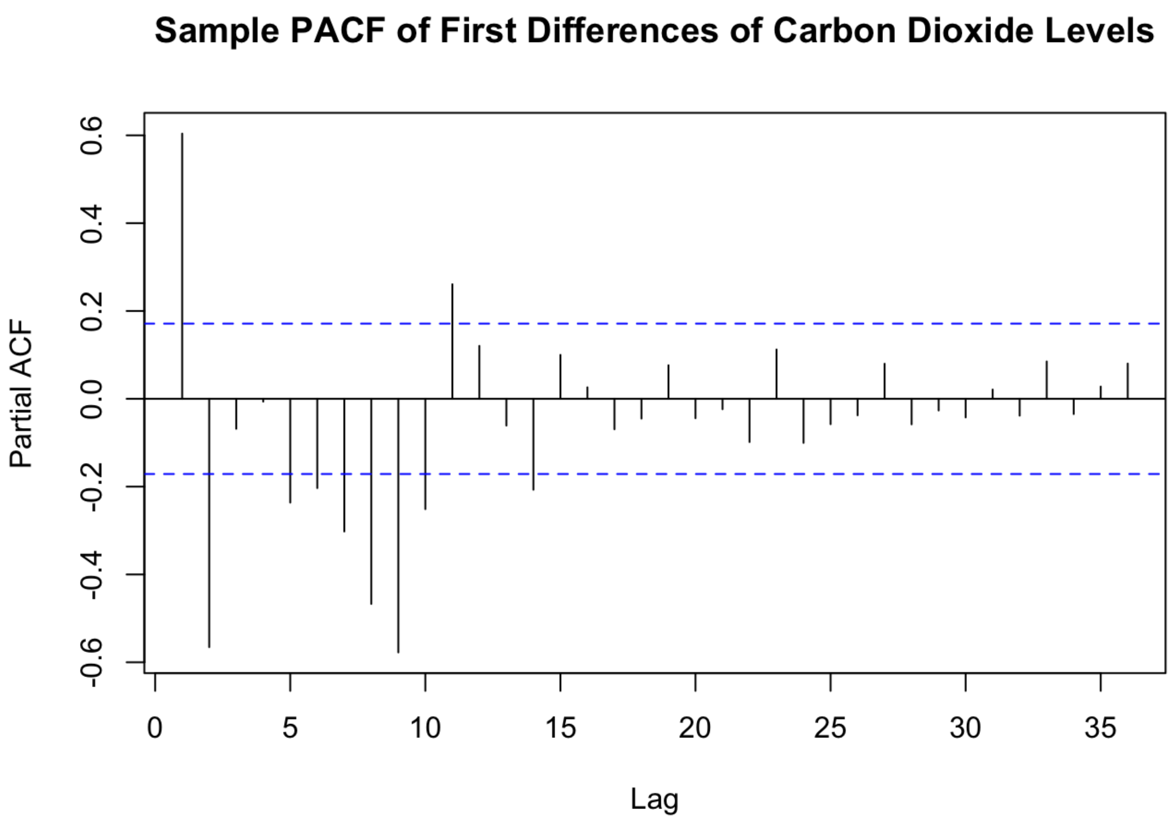
```
plot(diff(co2),ylab='First Difference of CO2',xlab='Time',main = "Time
   Series Plot of the First Differences of Carbon Dioxide Levels")
```




```
acf(as.vector(diff(co2)),lag.max=36,main="Sample ACF of First Differenc
  es of Carbon Dioxide Levels")
```

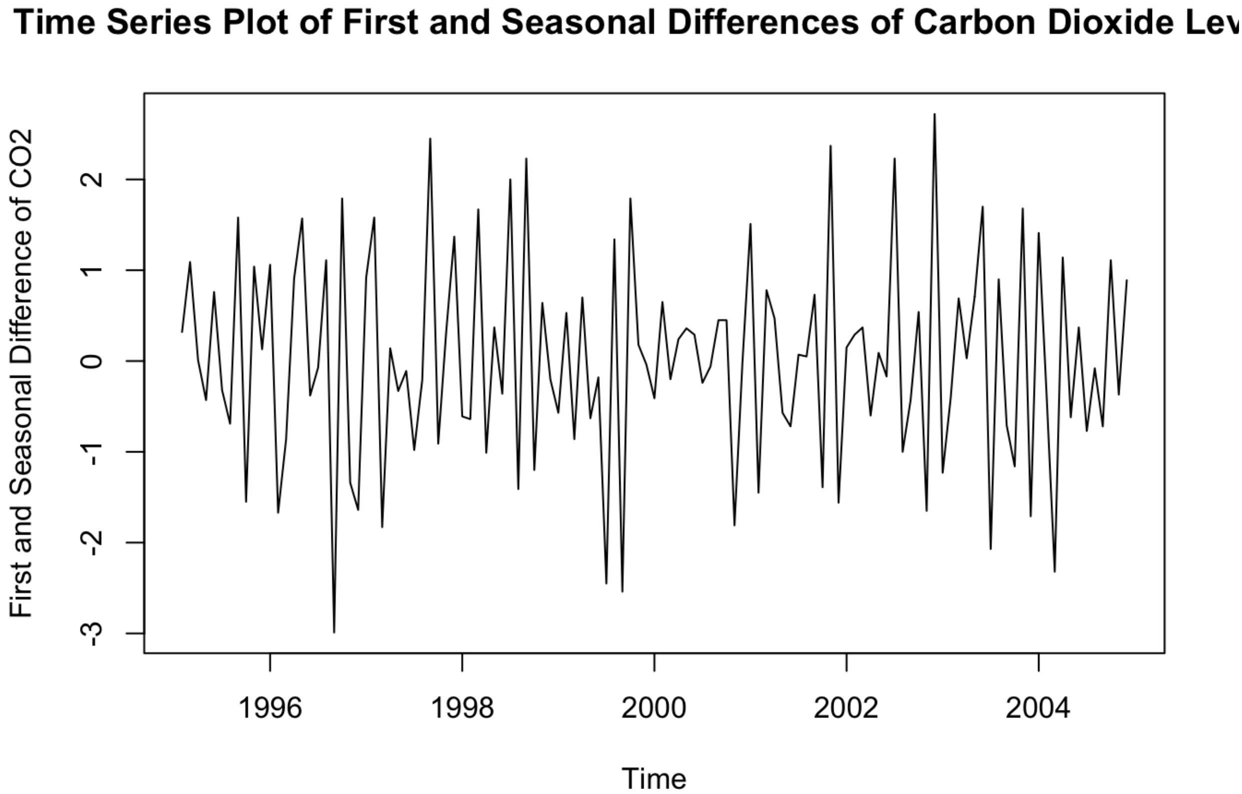



```
pacf(as.vector(diff(co2)),lag.max=36,main="Sample PACF of First Differe
  nces of Carbon Dioxide Levels")
```

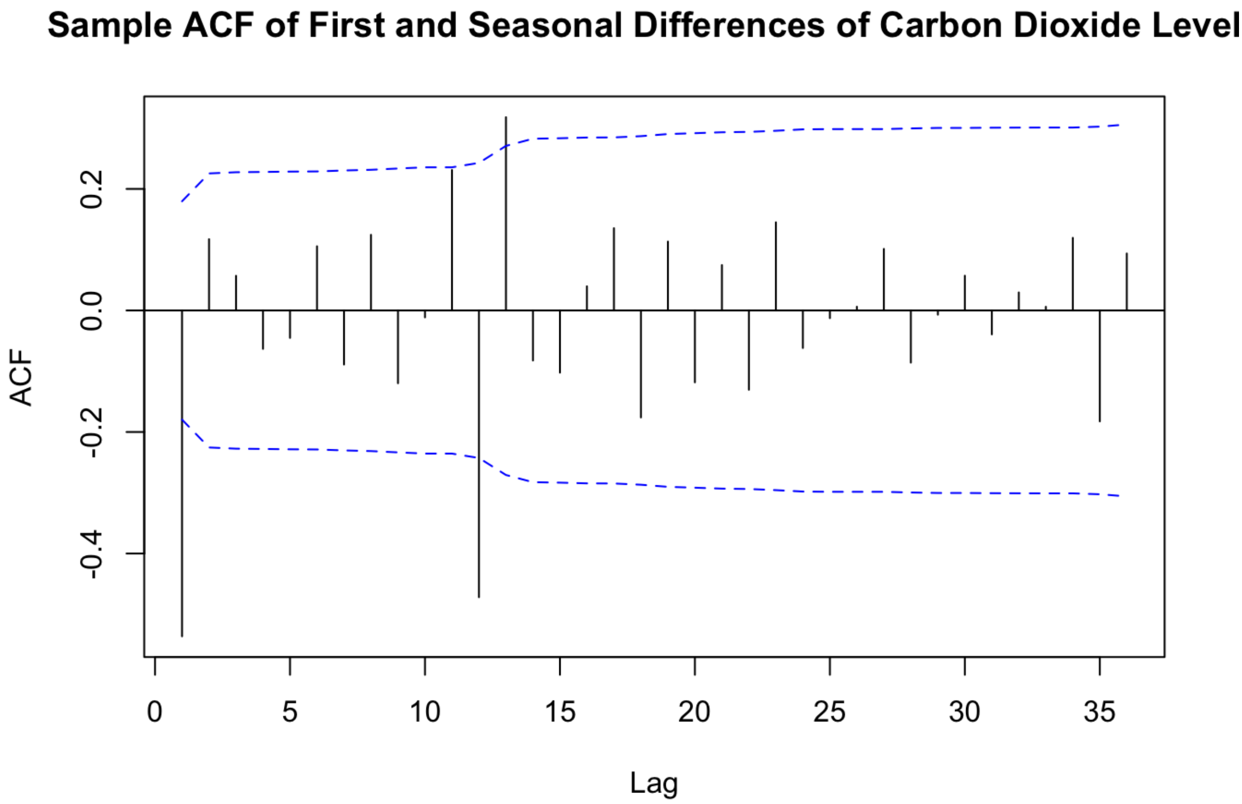


Although the general upward trend is resolved, we still have a seasonal effect pattern in time series plot. So, we consider taking a seasonal difference. Here we have the time series, ACF and PACF plots after taking one ordinary difference and one seasonal difference.

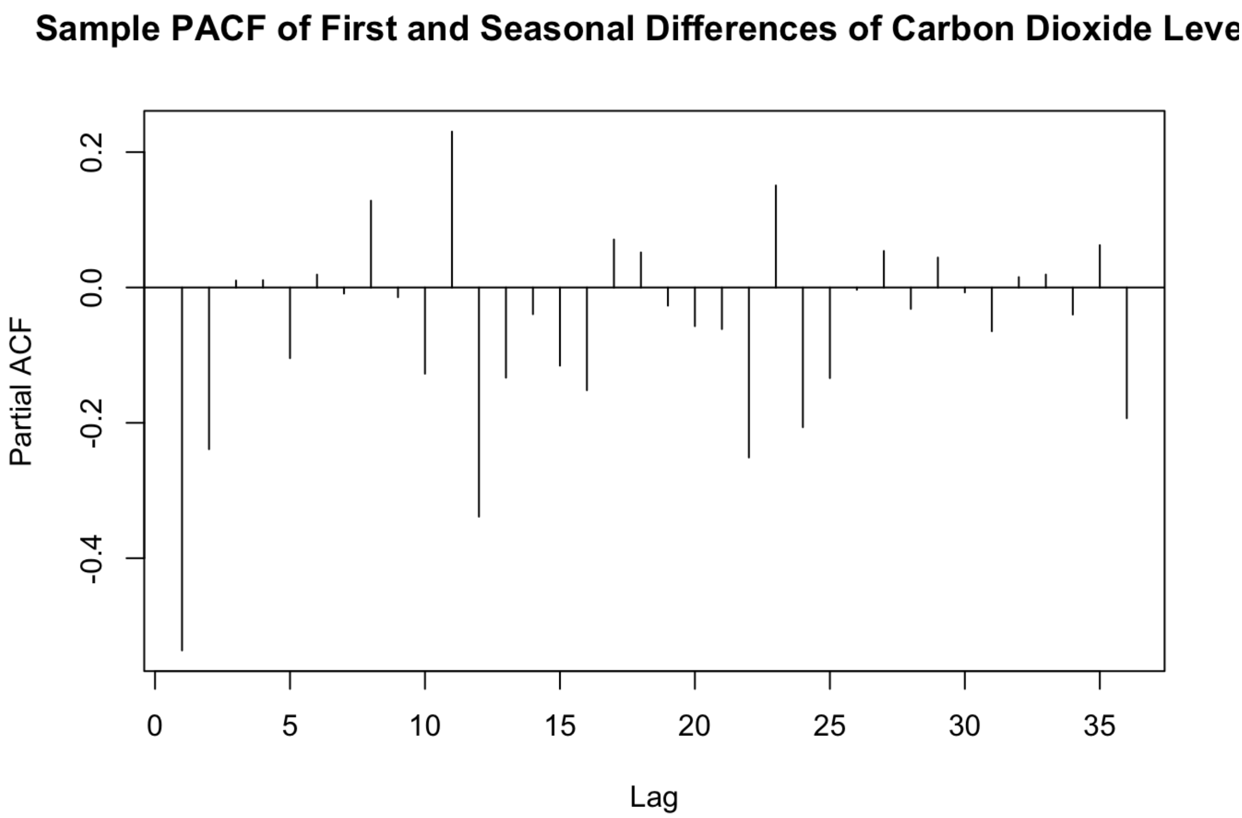
```
plot(diff(diff(co2),lag=12),xlab='Time',ylab='First and Seasonal Differ
  ence of CO2',main= "Time Series Plot of First and Seasonal Difference
  s of Carbon Dioxide Levels")
```




```
acf(as.vector(diff(diff(co2),lag=12)),lag.max=36,ci.type='ma',main="Sam
ple ACF of First and Seasonal Differences of Carbon Dioxide Levels")
```



```
pacf(as.vector(diff(diff(co2),lag=12)),lag.max=36,ci.type='ma',main="Sa
  mple PACF of First and Seasonal Differences of Carbon Dioxide Levels"
  )
```



After taking a seasonal difference, we nearly get rid of the seasonality in the series and very little autocorrelation remains in the series. This plot also suggests that a simple model which incorporates the lag 1 and lag 12 autocorrelations might be adequate. We will consider specifying an ARIMA(0,1,1)×(0,1,1)$_{12}$ model to this series with the following model formulation:

$$\nabla_{12}\nabla Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13}.$$

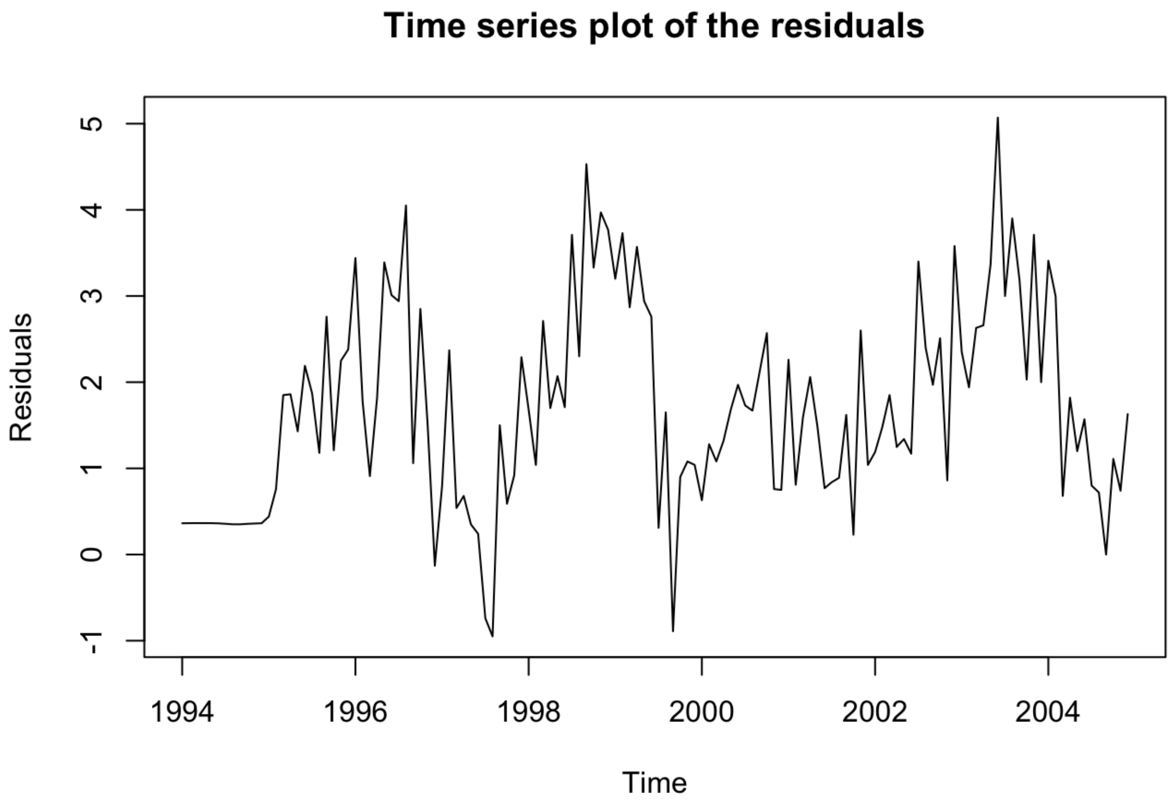
Residuals approach

We will start with displaying time series plot, from which the existence of seasonality and an ordinary trend were obvious.

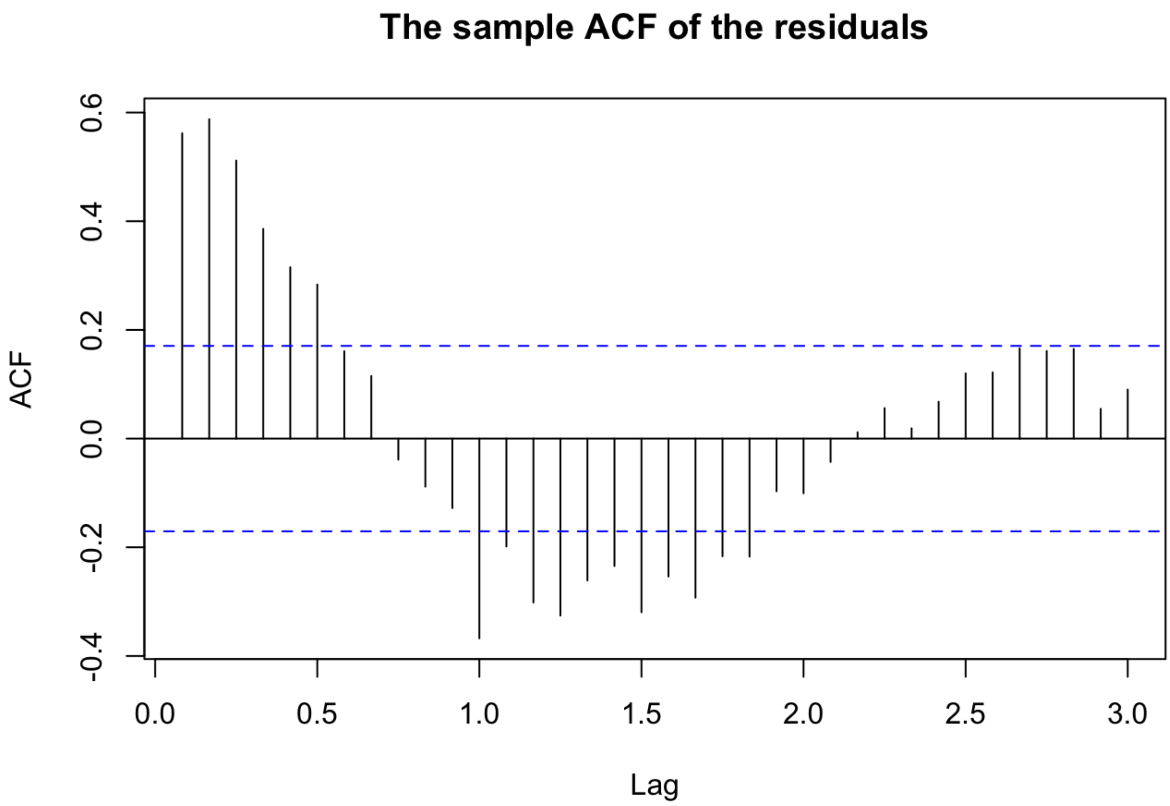
```
data(co2)
plot(co2,ylab='CO2', main = "Monthly Carbon Dioxide Levels at Alert, NW
  T, Canada")
```

Now we will fit SARIMA(0,0,0)x(0,1,0)$_{12}$ model and display time series, ACF and PACF plots of the residuals.

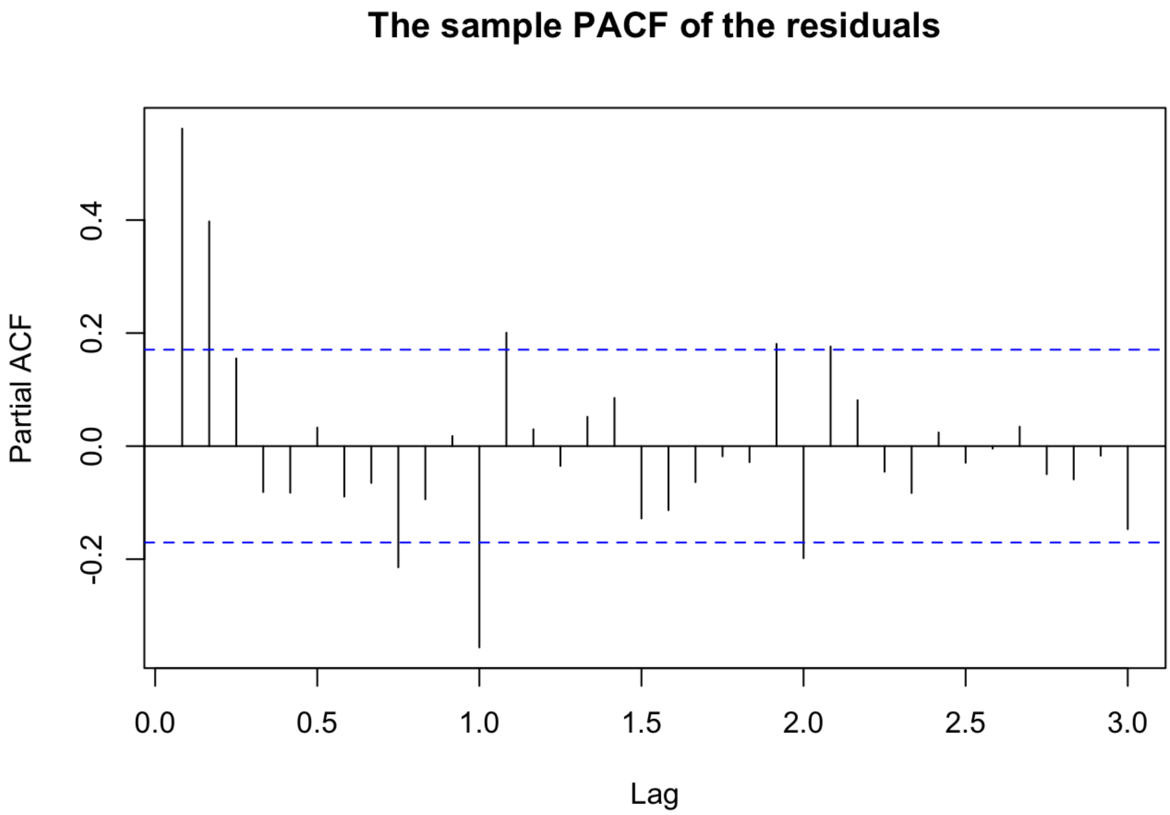
```
m1.co2 = arima(co2,order=c(0,0,0),seasonal=list(order=c(0,1,0), period=
  12))
res.m1 = residuals(m1.co2);
plot(res.m1,xlab='Time',ylab='Residuals',main="Time series plot of the
   residuals")
```




```
acf(res.m1, lag.max = 36, main = "The sample ACF of the residuals")
```



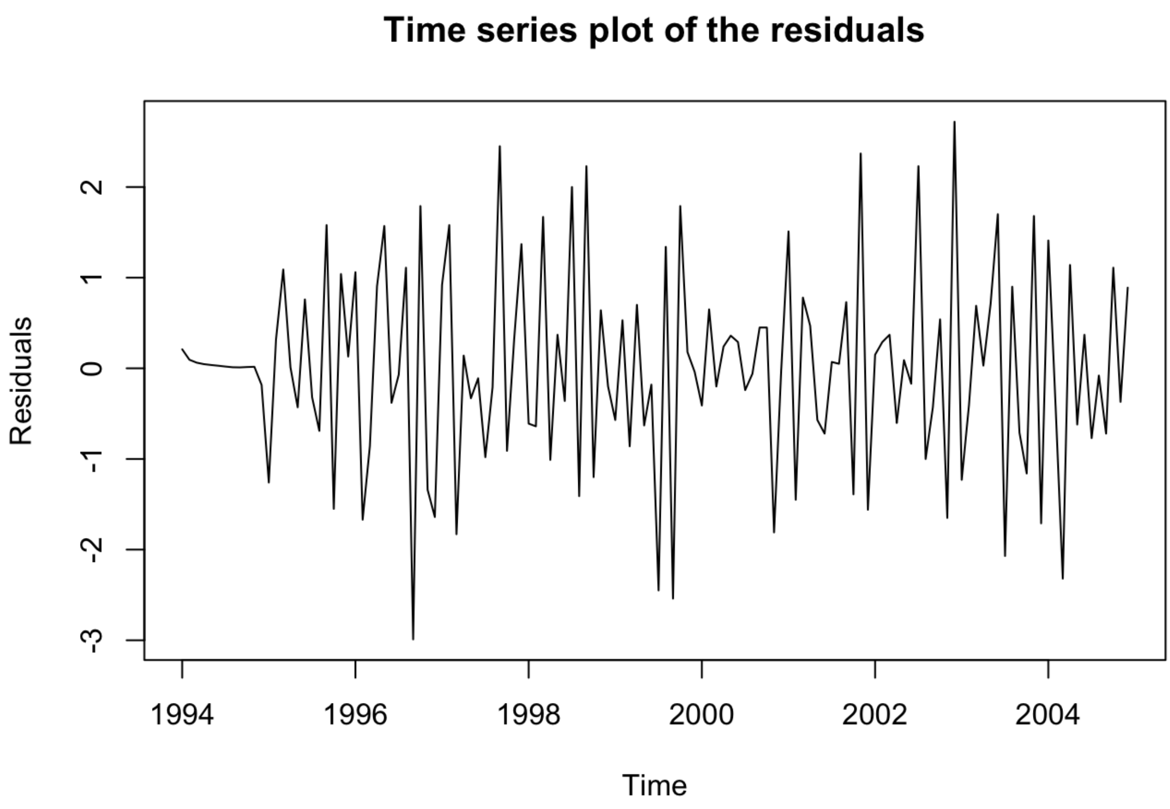

```
pacf(res.m1, lag.max = 36, main = "The sample PACF of the residuals")
```



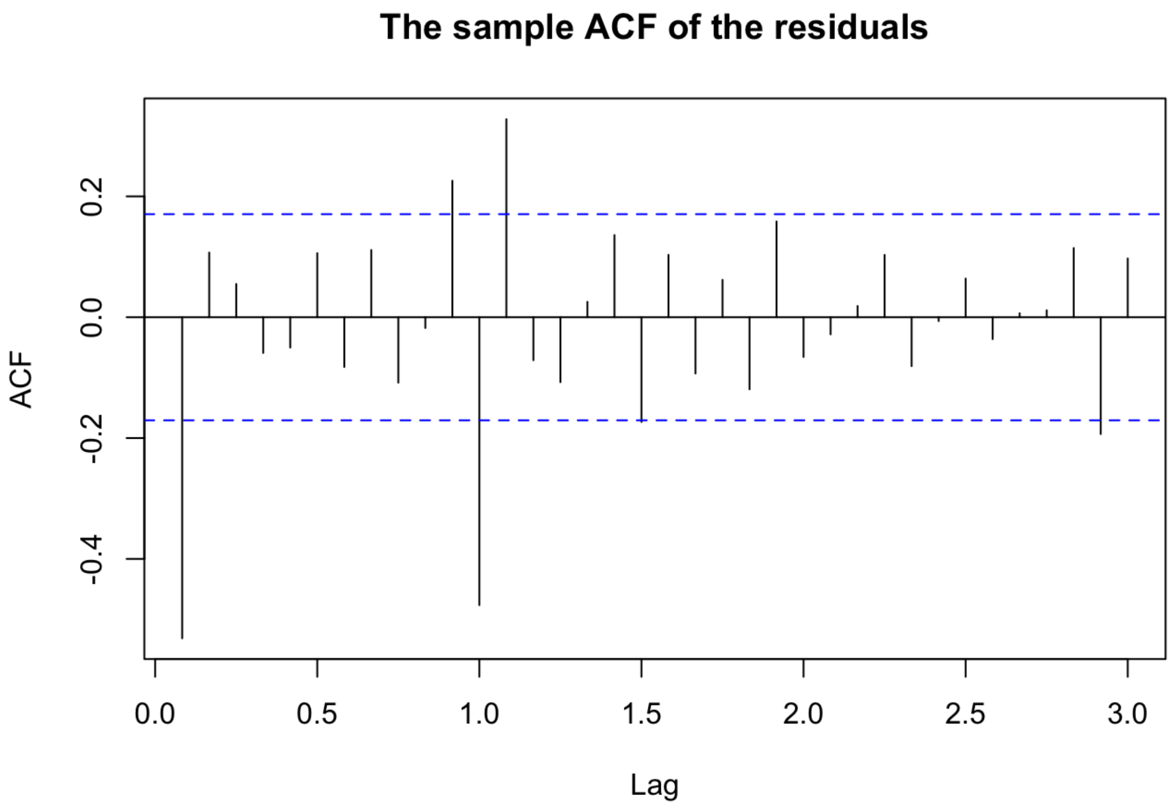
When we focus on the autocorrelations at seasonal lags, there is no pattern implying the existence of a seasonal trend. However, the existence of an ordinary trend is obvious from the time series, ACF and PACF plots. To get rid of the ordinary trend seen in the residuals, we will fit the SARIMA(0,1,0)x$(0,1,0)_{12}$ model.

The following plots show the corresponding time series, ACF, and PACf plots.

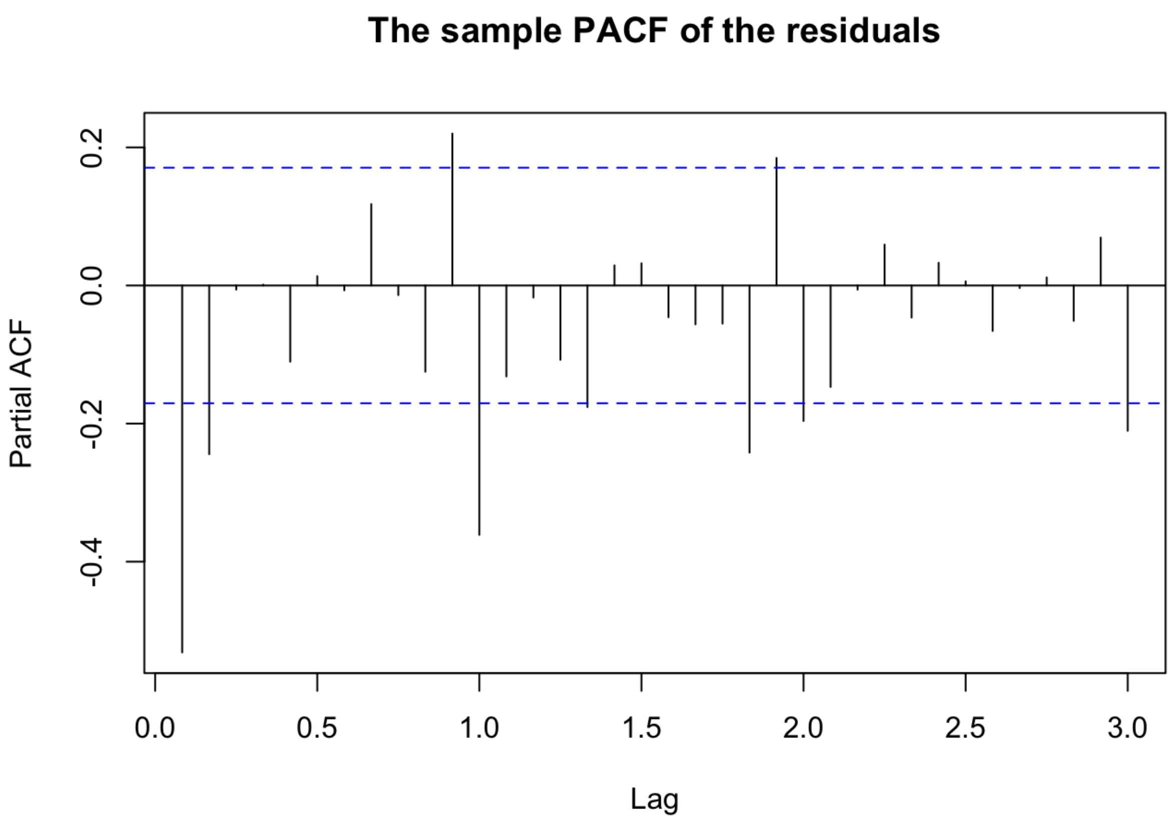
```
m2.co2 = arima(co2,order=c(0,1,0),seasonal=list(order=c(0,1,0), period=
  12))
res.m2 = residuals(m2.co2);
plot(res.m2,xlab='Time',ylab='Residuals',main="Time series plot of the
   residuals")
```




```
acf(res.m2, lag.max = 36, main = "The sample ACF of the residuals")
```




```
pacf(res.m2, lag.max = 36, main = "The sample PACF of the residuals")
```



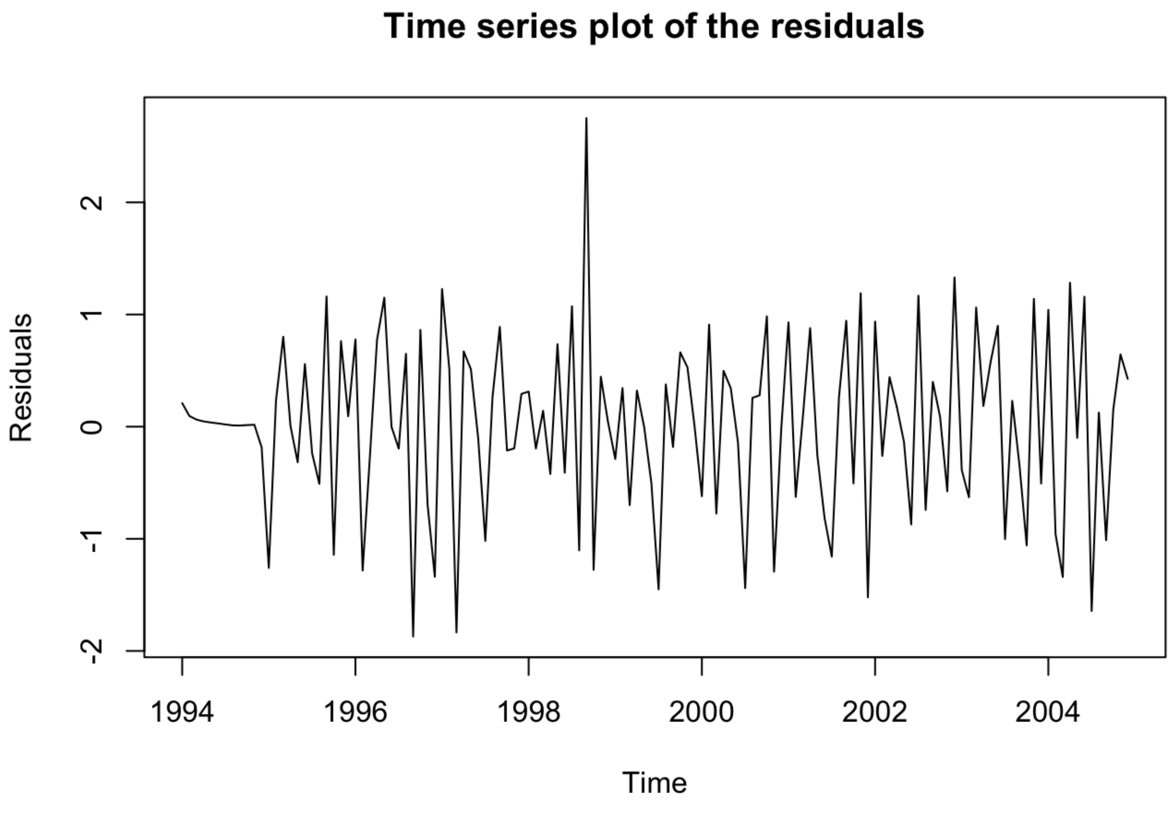
There is no evidence of an ordinary trend left in the latest residuals. The residuals of the model SARIMA(0,1,0)x$(0,1,0)_{12}$ now includes SARMA and ARMA components.

For the orders of SARMA component, we will consider the lags $s, 2s, 3, \ldots$, i.e. 12, 24, 36, … These are shown as 1, 2, 3, … in both the ACF and PACF plots.

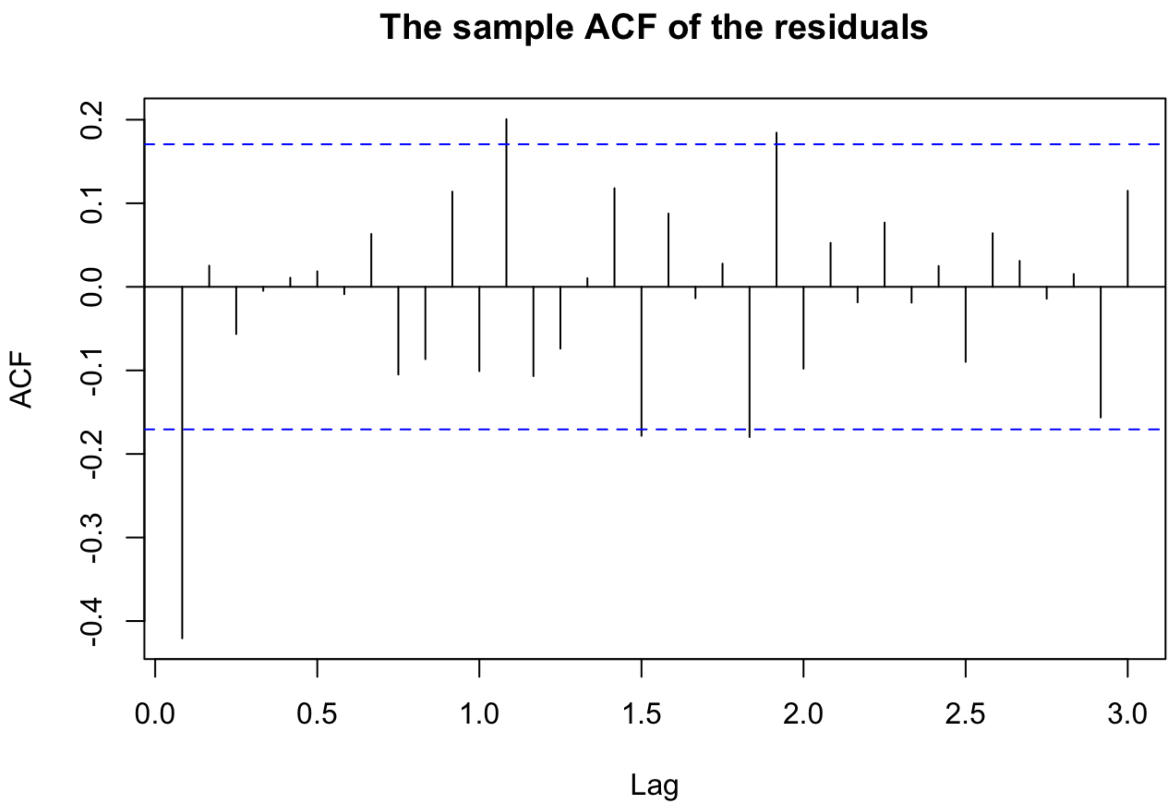
When we consider the lags 1, 2, 3, … of ACF plot, the correlation at lag 1 is significant, and in PACF there is a decreasing pattern at lags 1, 2, 3, … This implies the existence of an SMA(1) component.

Now, we will fit SARIMA(0,1,0)x(0,1,1)$_{12}$ model and try to see if we get rid of the effect of the seasonal component on the residuals.

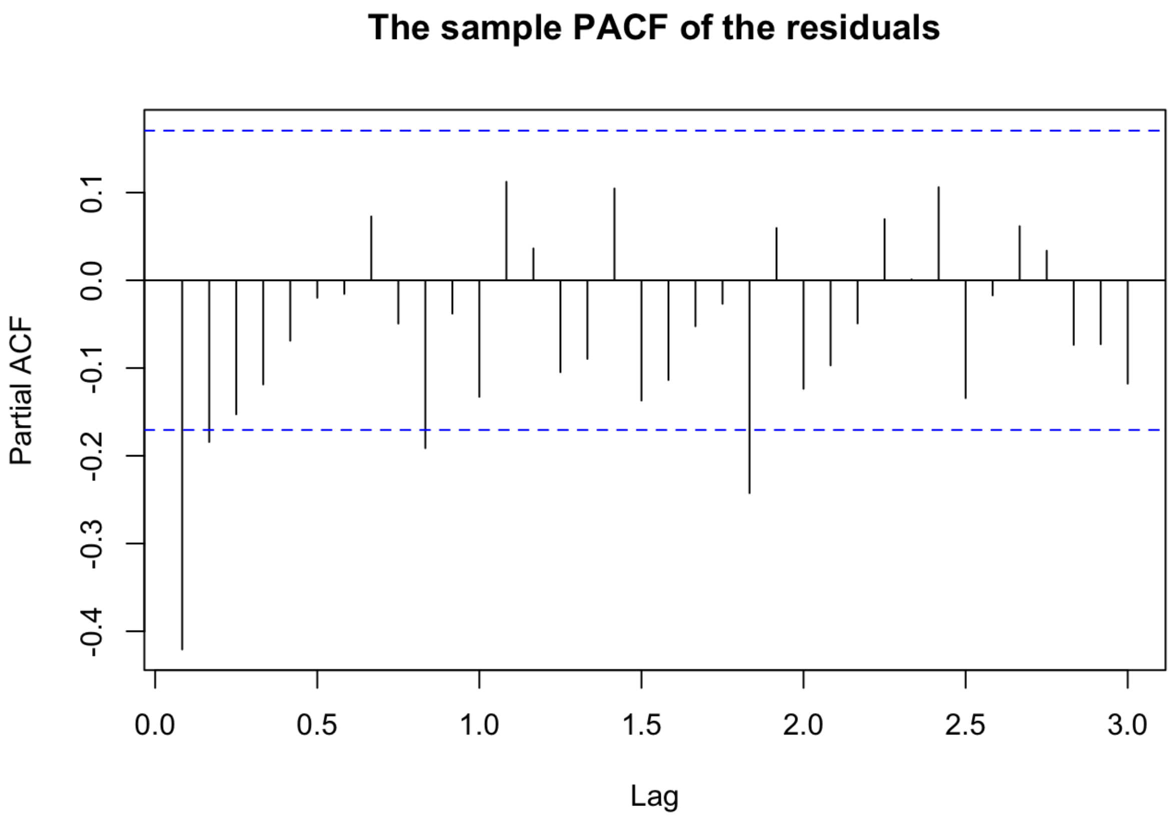
```
m3.co2 = arima(co2,order=c(0,1,0),seasonal=list(order=c(0,1,1), period=
  12))
res.m3 = residuals(m3.co2);
plot(res.m3,xlab='Time',ylab='Residuals',main="Time series plot of the
   residuals")
```




```
acf(res.m3, lag.max = 36, main = "The sample ACF of the residuals")
```




```
pacf(res.m3, lag.max = 36, main = "The sample PACF of the residuals")
```

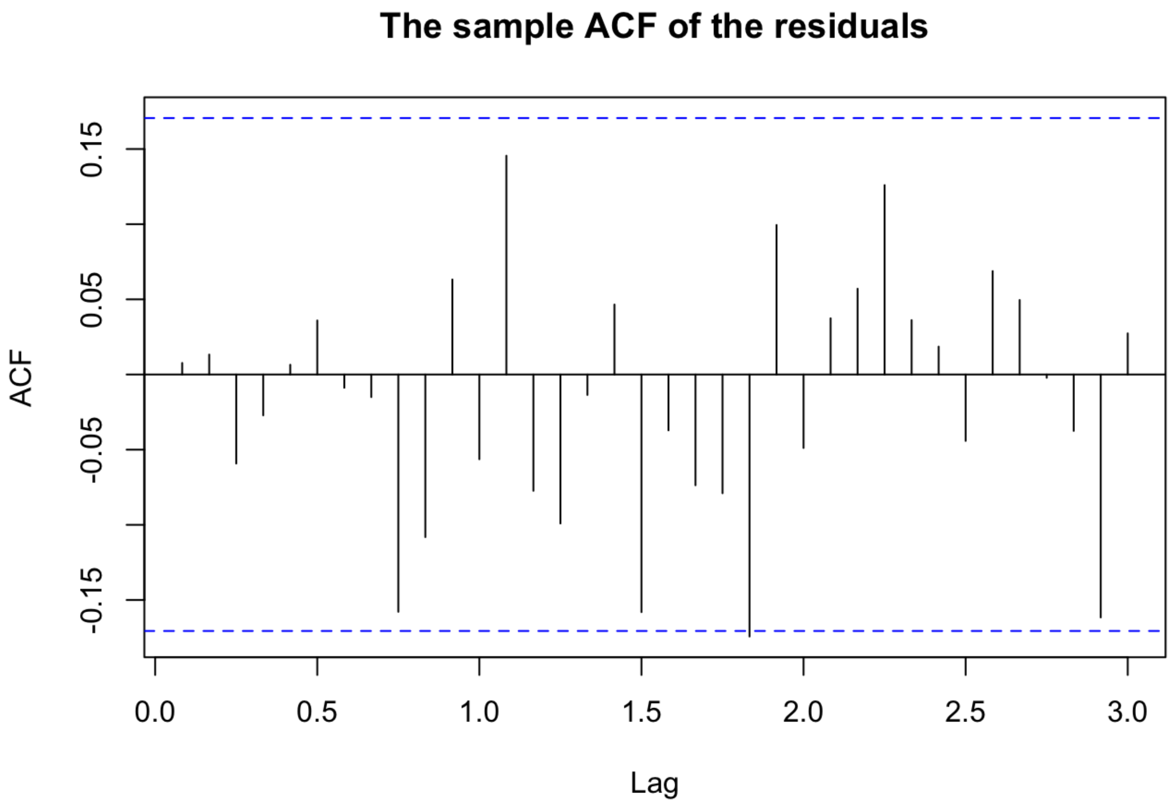


At the seasonal lags, especially at the first seasonal lag in ACF, the autocorrelations became insignificant or slightly significant after adding the seasonal component. We will use the resulting ACF and PACF plots to specify the orders of ARMA component as there is no highly significant correlation at the lags $s, 2s, 3s, \ldots$.

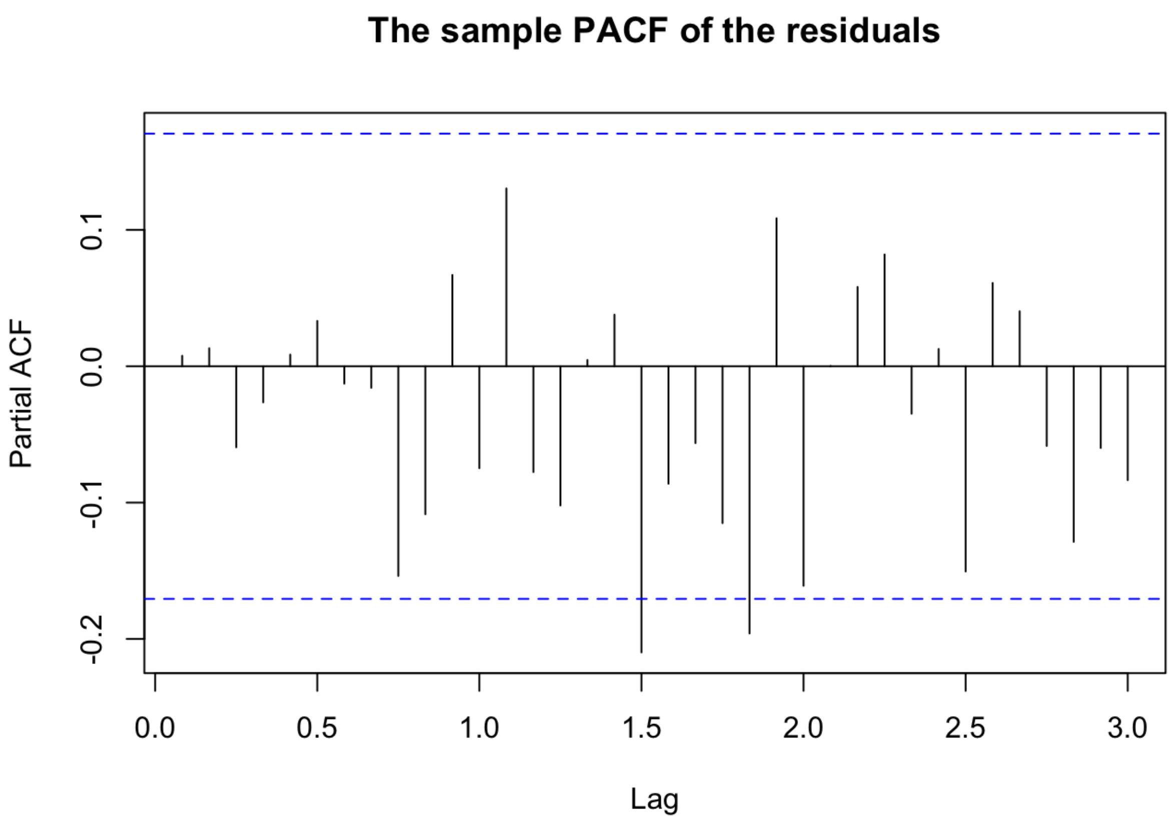
The last ACF-PACF pair has significant pikes at the first lags implying an ARMA(1,1) model or due to the (insignificant) decrease of correlations in PACF, we would specify ARMA(0,1) for the residuals.

Now we will fit SARIMA(0,1,1)x(0,1,1)$_{12}$ and SARIMA(1,1,1)x(0,1,1)$_{12}$ models.

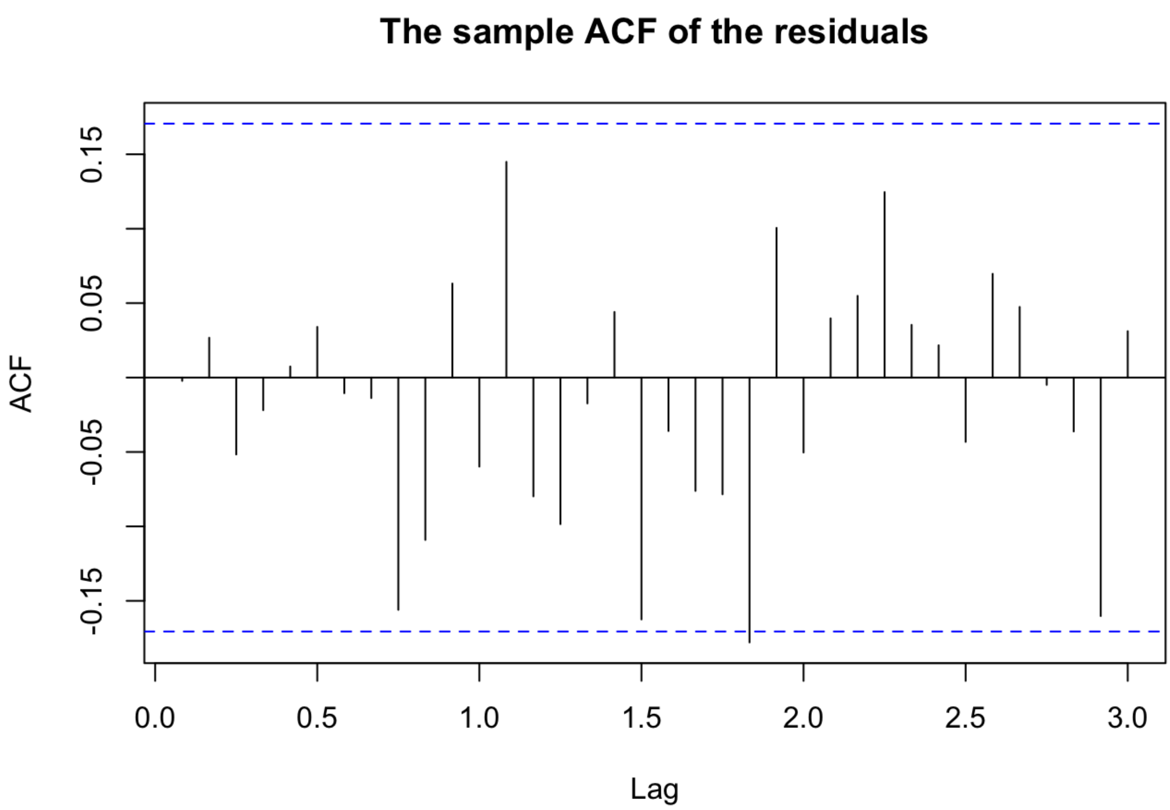
```
# SARIMA(0,1,1)x(0,1,1)_12 MODEL
m4.co2 = arima(co2,order=c(0,1,1),seasonal=list(order=c(0,1,1), period=
  12))
res.m4 = residuals(m4.co2);
acf(res.m4, lag.max = 36, main = "The sample ACF of the residuals")
```



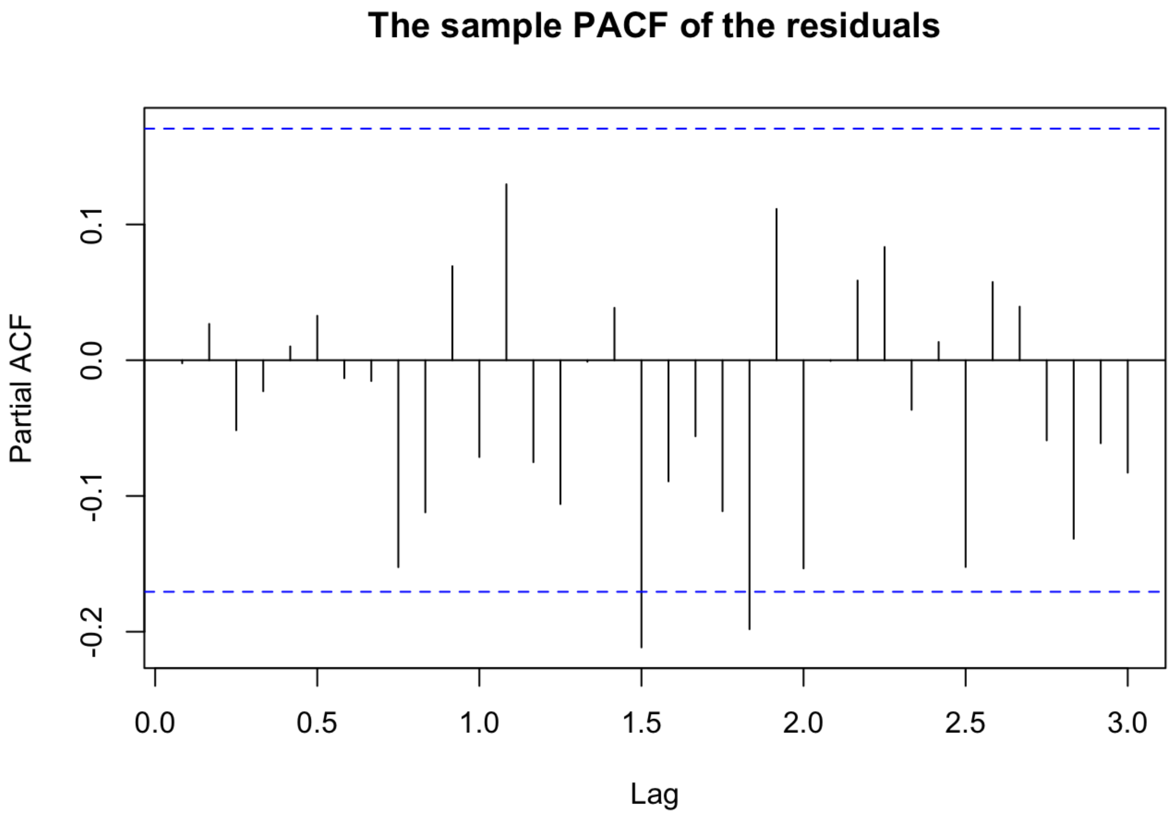

```
pacf(res.m4, lag.max = 36, main = "The sample PACF of the residuals")
```




```
# SARIMA(1,1,1)x(0,1,1)_12 MODEL
m5.co2 = arima(co2,order=c(1,1,1),seasonal=list(order=c(0,1,1), period=
  12))
res.m5 = residuals(m5.co2);
acf(res.m5, lag.max = 36, main = "The sample ACF of the residuals")
```



```
pacf(res.m5, lag.max = 36, main = "The sample PACF of the residuals")
```



We got nearly white noise residual series for both of the SARIMA(0,1,1)x(0,1,1)$_{12}$ and SARIMA(1,1,1)x(0,1,1)$_{12}$ models. The residuals of SARIMA(0,1,1)x(0,1,1)$_{12}$ model are closer to the white noise. So, we can conclude with the orders $p = 0$, $d = 1$, $q = 1$, $P = 0$, $D = 1$, $Q = 1$, and $s = 12$ of SARIMA(p,d,q)x(P,D,Q)$_s$ model.

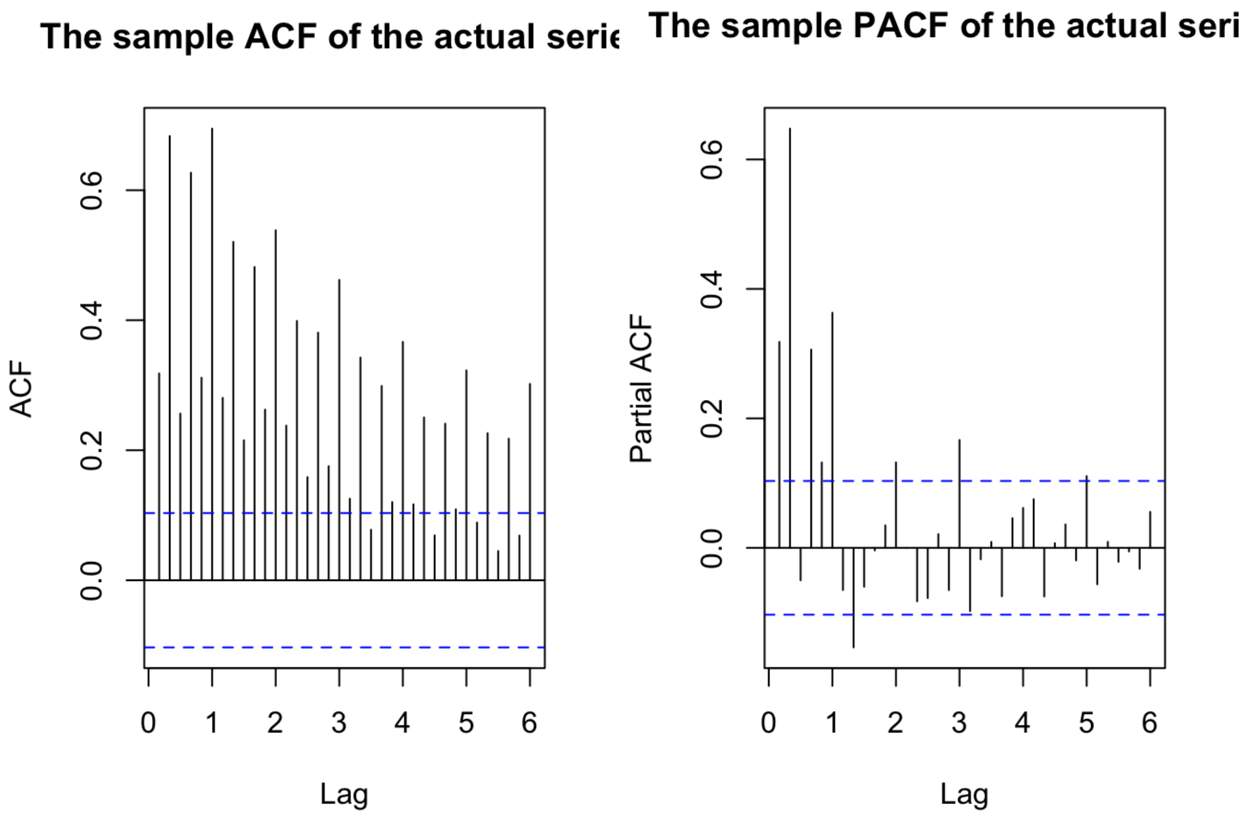
As another example, let's consider a simulated series from the SARIMA(1,1,0)x(0,1,1)$_6$ model generated by the following code chunk:

```
sarima.data = sarima.Sim(n=60, period=6,
                model = list(order = c(1,1,0),ar=-0.95),
                list(order= c(0,1,1), ma = -0.75))
```

So, the model includes, ordinary and seasonal trends and AR and SMA components.

ACF and PACF of the actual series are displayed below.

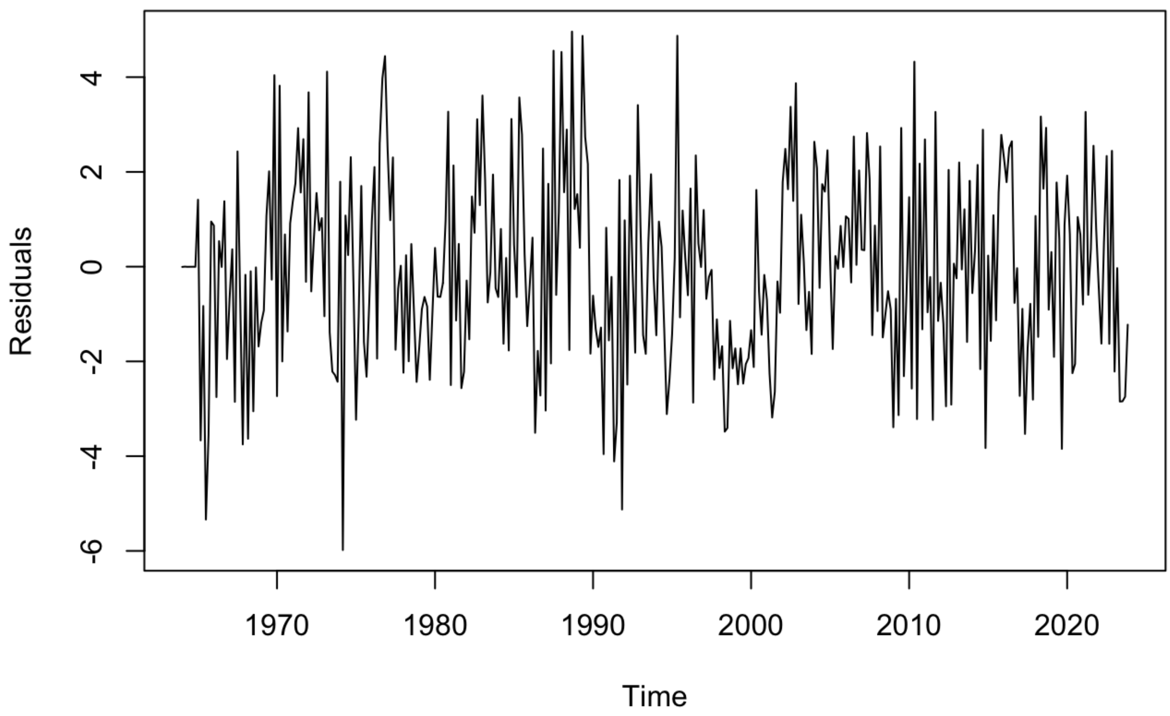
```
par(mfrow=c(1,2))
acf(sarima.data, lag.max = 36, main = "The sample ACF of the actual ser
  ies")
pacf(sarima.data, lag.max = 36, main = "The sample PACF of the actual s
  eries")
```



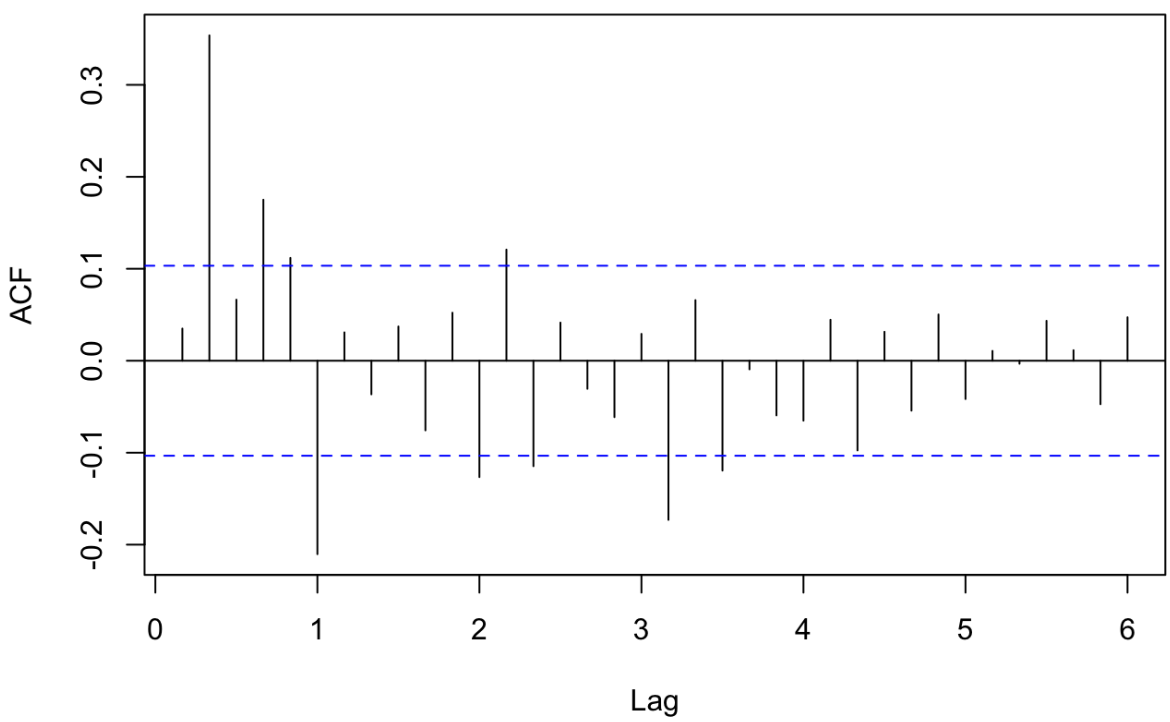
The jumps in ACF at lags 6, 12, 18, and so on indicate that the period of the series is 6

Now, we will specify the orders of the model following the residuals approach. First, we fit a plain model with only the first seasonal difference with order D = 1and see if we can get rid of the seasonal trend effect by inspecting the autocorrelation structure of the residuals.

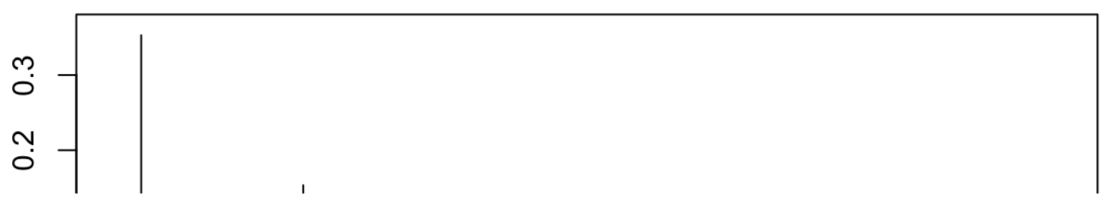
Time series plot of the residuals

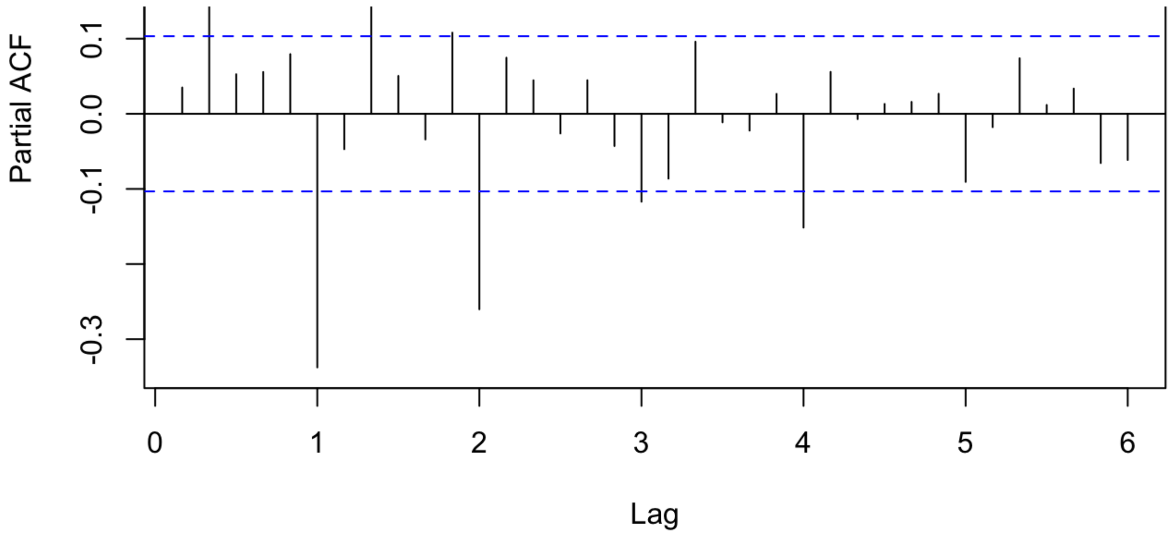


The sample ACF of the residuals



The sample PACF of the residuals

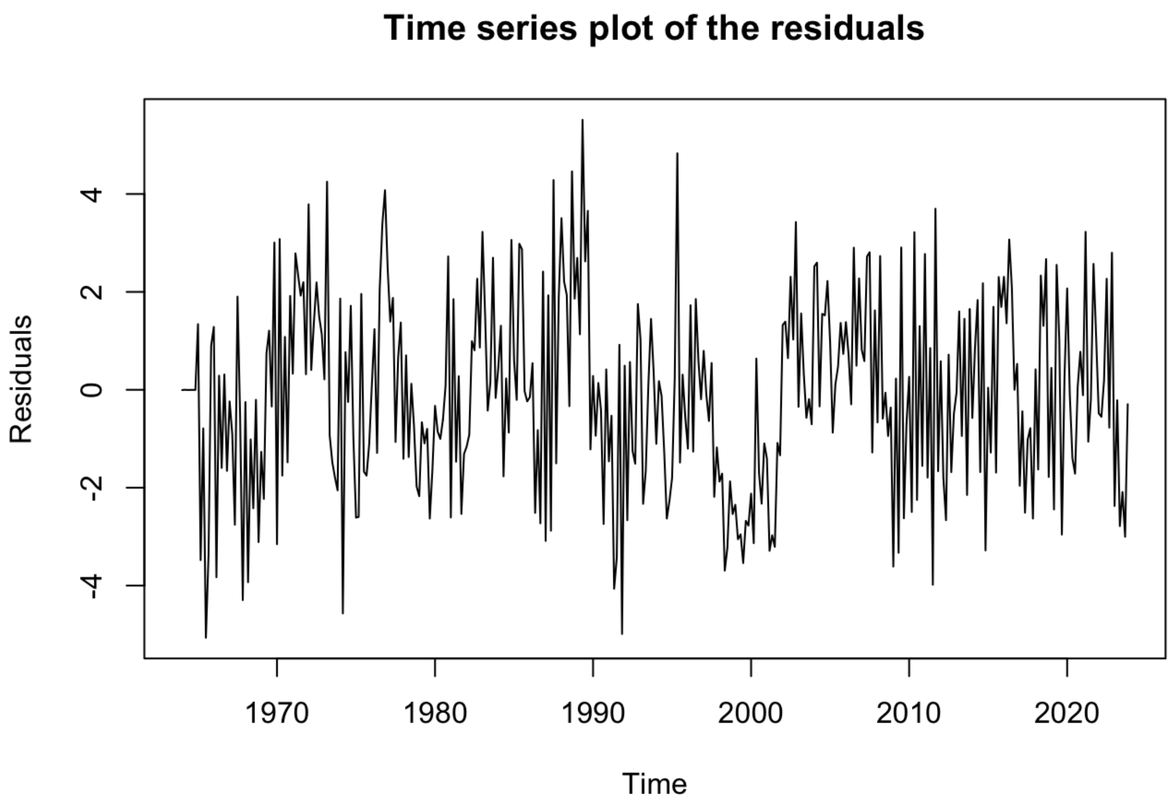




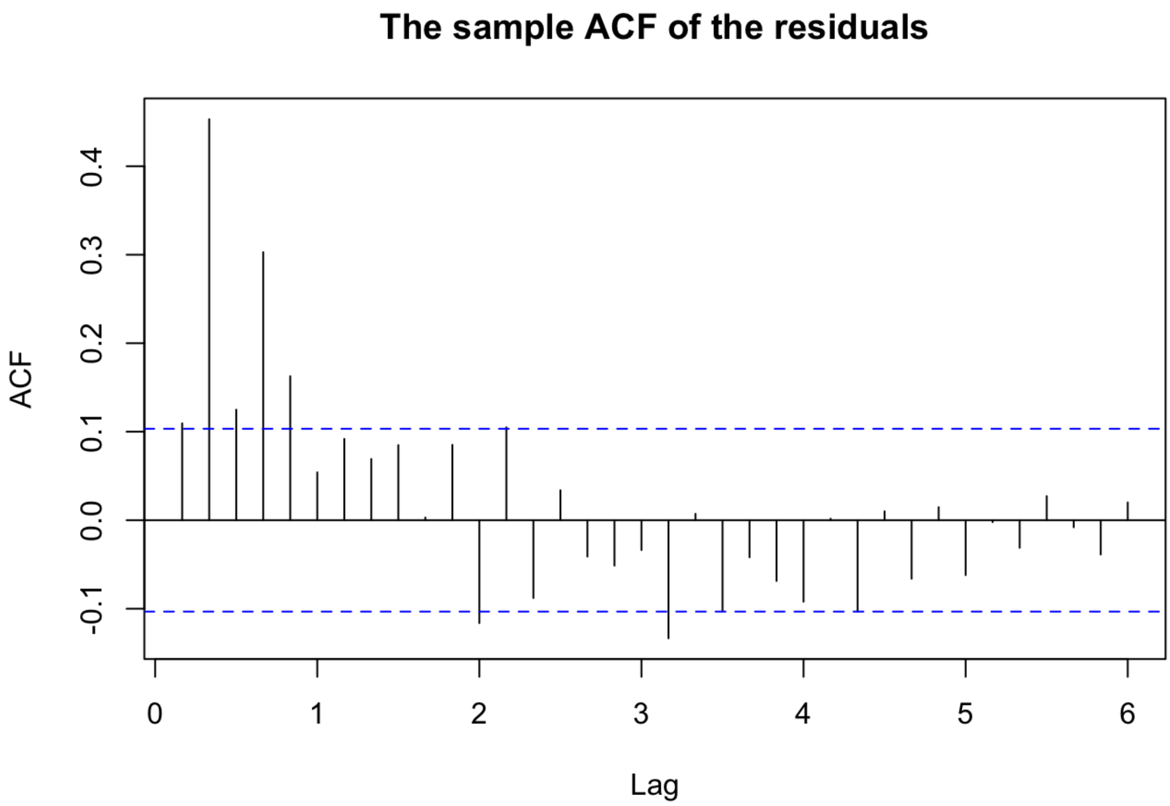
In the ACF and PACF, when we only consider the lags s, 2s, 3s, and so on (which correspond to lags 1, 2, 3, … in the ACF and PACF plots) to figure out the SAR and SMA orders, we see that there is a decreasing wave pattern in PACF and highly significant autocorrelation at the first lag in ACF. So, we consider the SMA(1) model for the seasonal part.

Second, we will add the SMA(1) component and see if we get rid of the effect of the seasonal component in the residuals.

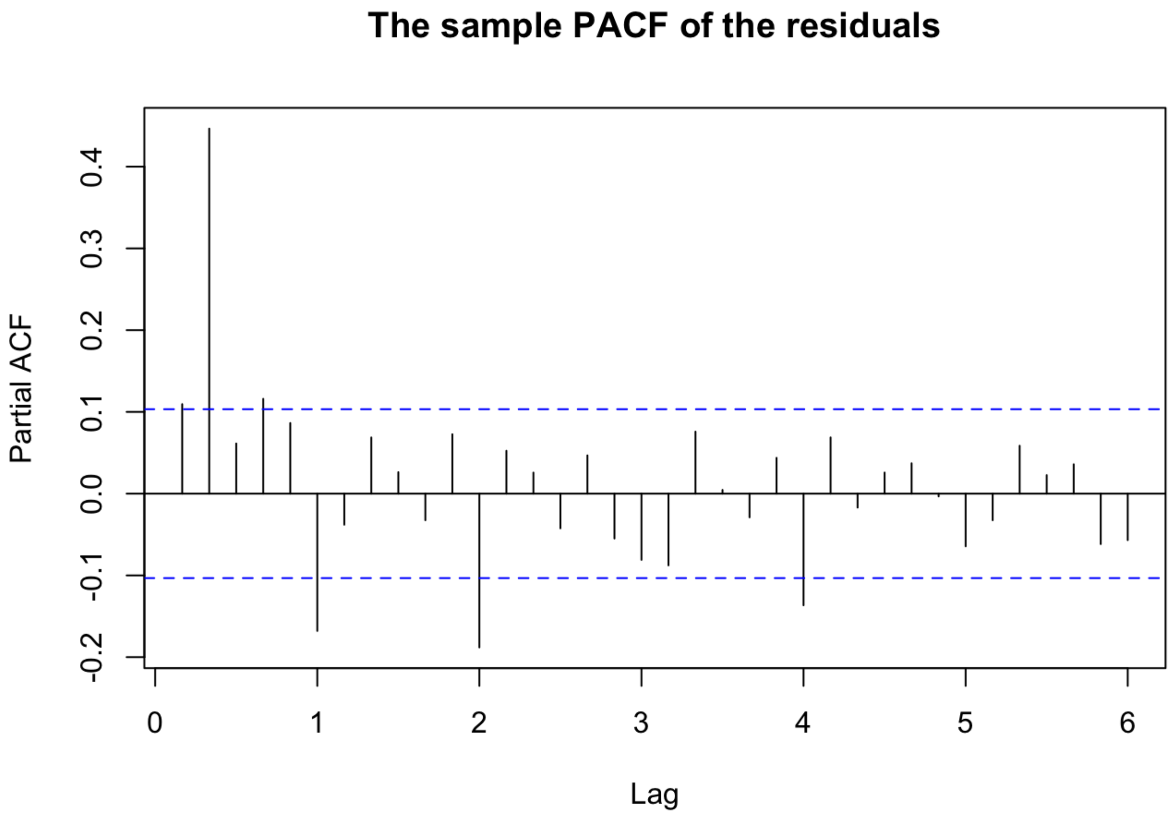
```
m2.sim = arima(sarima.data,order=c(0,0,0),seasonal=list(order=c(0,1,1),
   period=6))
res.m2 = residuals(m2.sim);
plot(res.m2,xlab='Time',ylab='Residuals',main="Time series plot of the
   residuals")
```



```
acf(res.m2, lag.max = 36, main = "The sample ACF of the residuals")
```

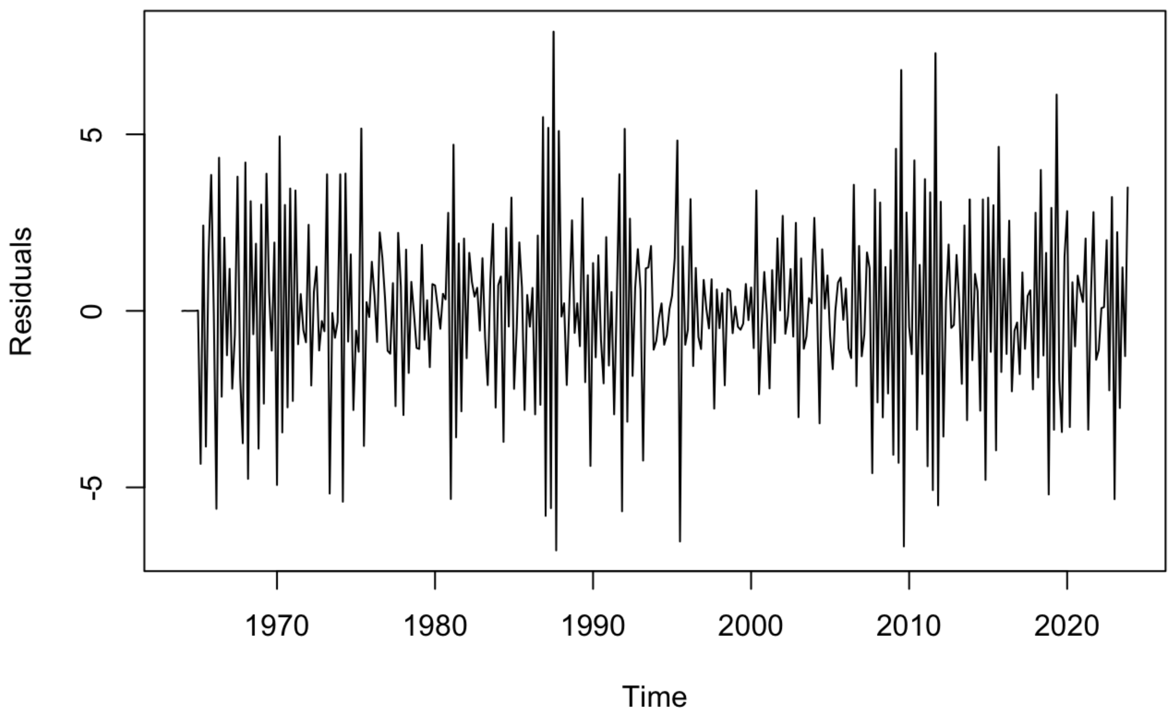



```
pacf(res.m2, lag.max = 36, main = "The sample PACF of the residuals")
```

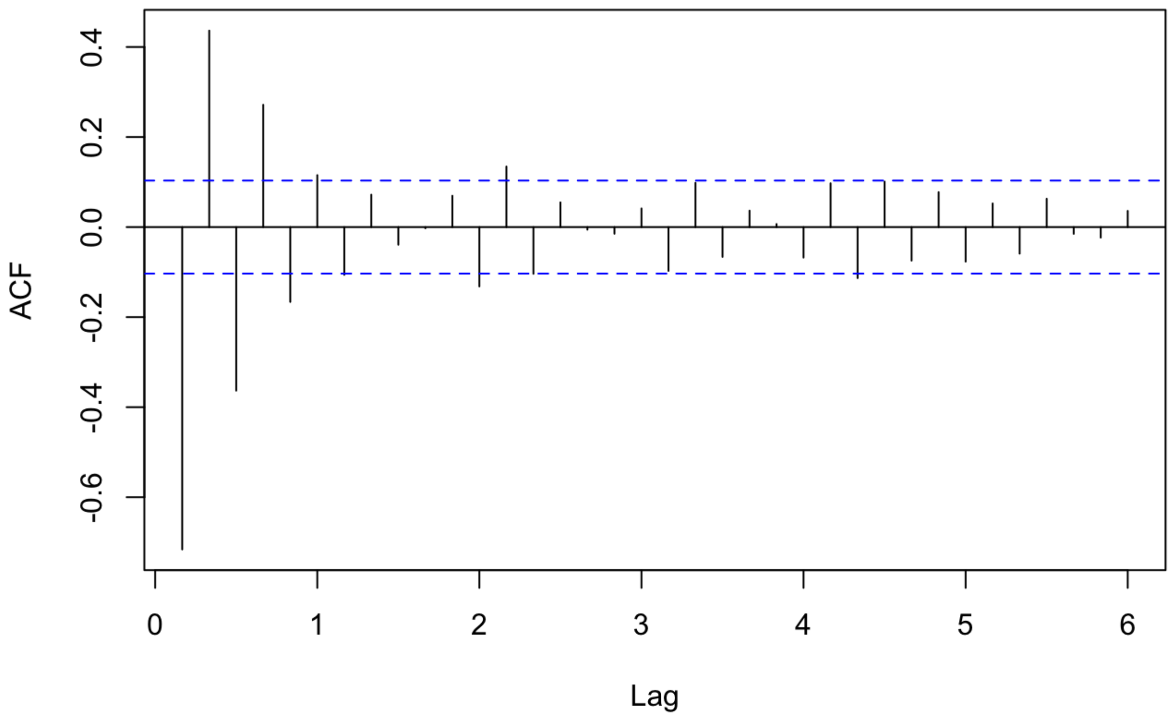


Although there is a slightly significant correlation at seasonal lag 2 in ACF and significant correlations at lags 1 and 2 in PACF, we can go on with ordinary differencing to get rid of the ordinary trend. We add the ordinary first difference to the model and inspect the residuals to see if there is still some evidence for ordinary trend and if not, we use the ACF and PACF to specify the orders of ARIMA part for the next step.

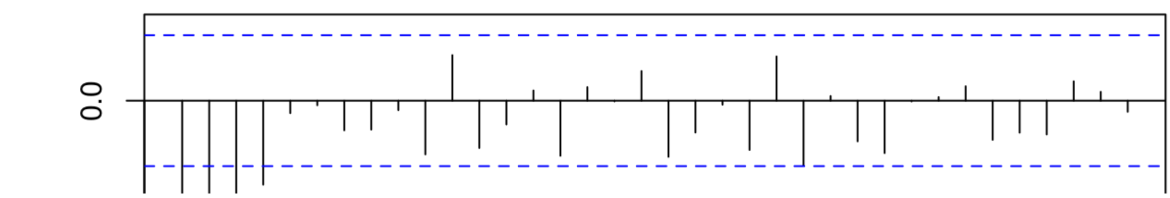
Time series plot of the residuals

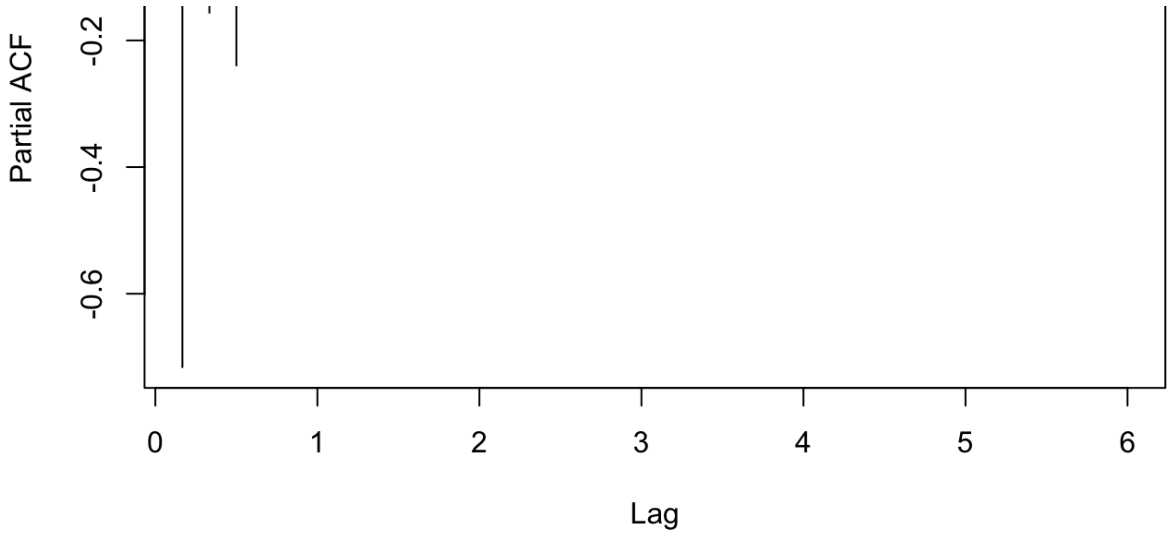


The sample ACF of the residuals



The sample PACF of the residuals

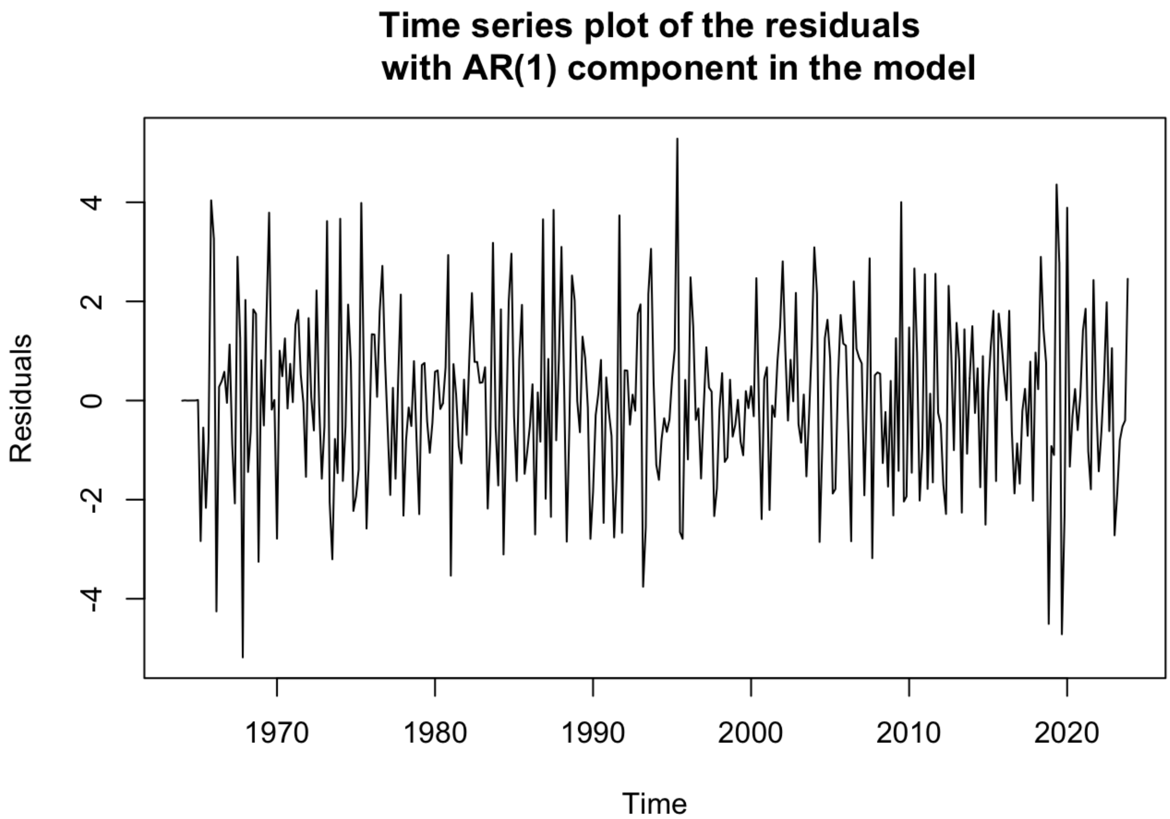




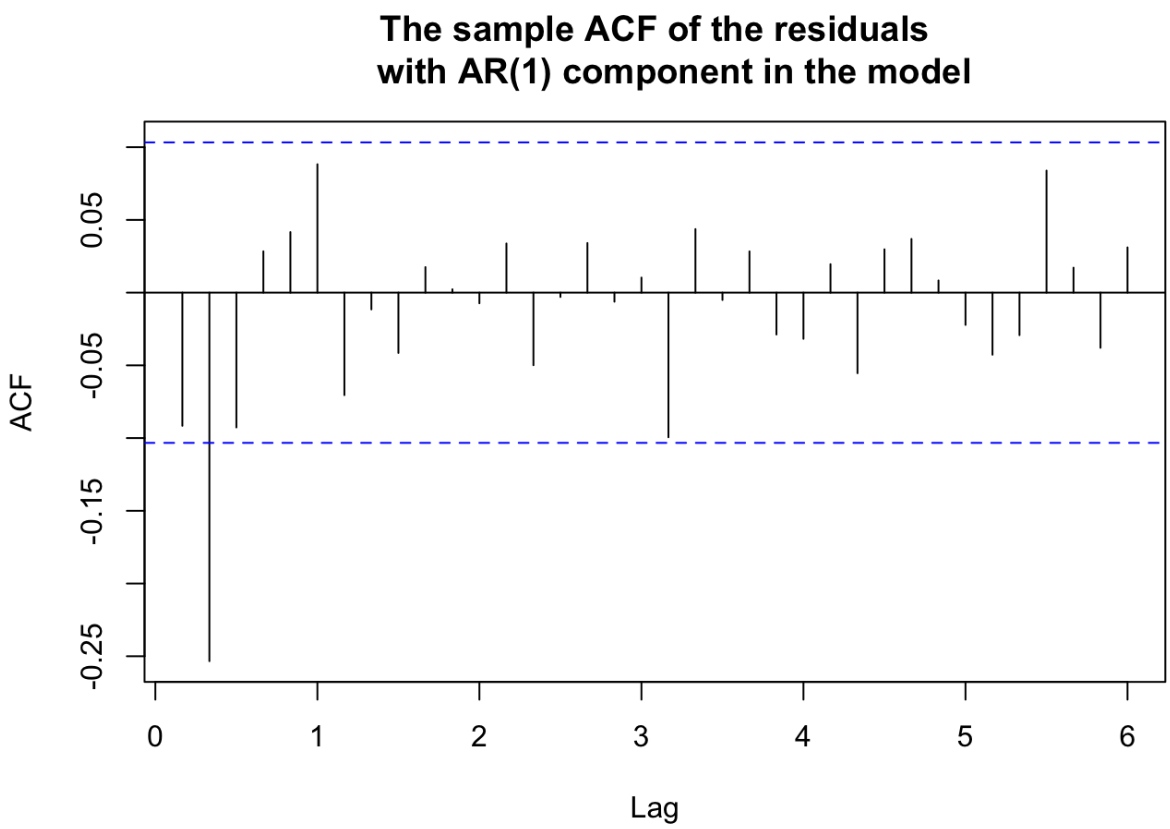
After adding the first ordinary difference, we have no longer evidence for neither ordinary trend nor seasonal component. So, we can go on with the specification of ARIMA orders. Because among the first 6 autocorrelations, there is a decreasing wave pattern in ACF and four highly significant correlations at lag 1, 2, 3 and 4. So, we can set p = 1, 2, 3, 4 and q = 0.

Now, we will try adding AR(1), AR(2), AR(3), and AR(4) ordinary parts to the model one by one and display ACF and PACF of residuals.

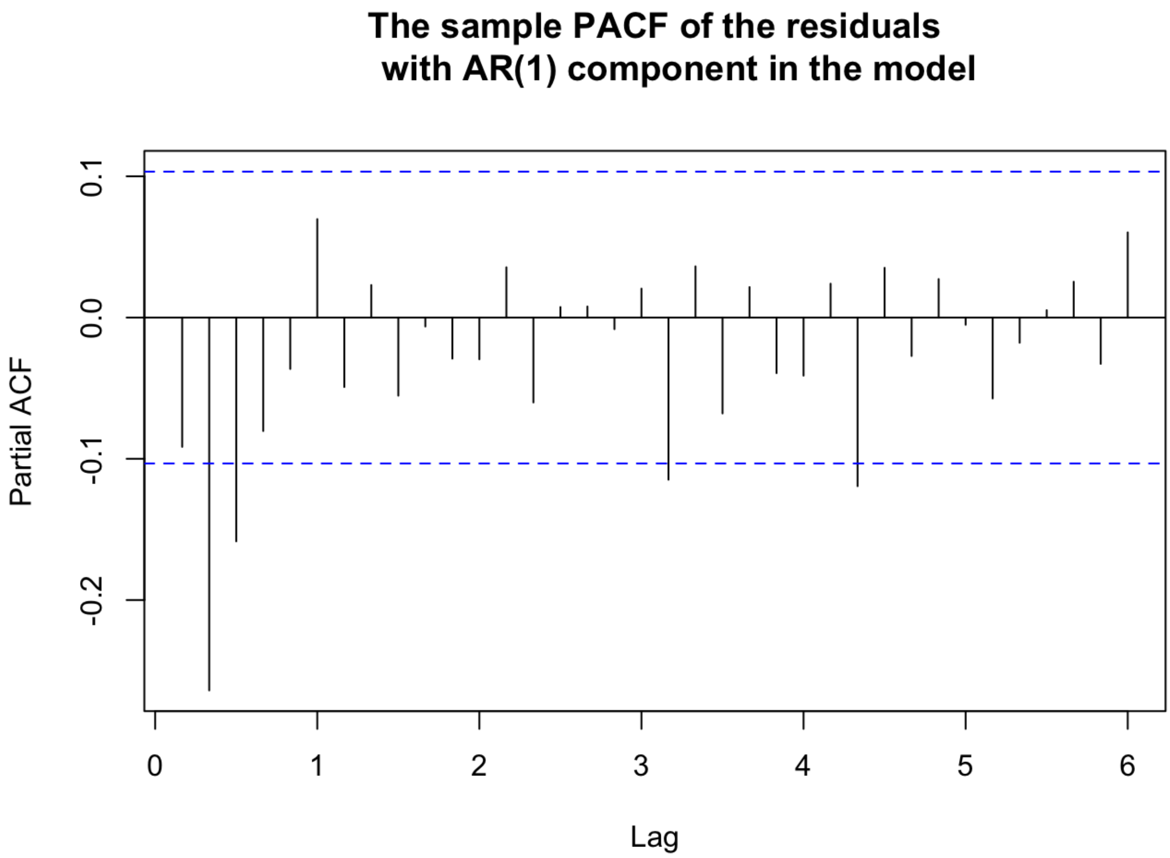
```
# SARIMA(1,1,0)x(0,1,1)_6 MODEL
m4.sim = arima(sarima.data,order=c(1,1,0),seasonal=list(order=c(0,1,1),
   period=6))
res.m4 = residuals(m4.sim);
plot(res.m4,xlab='Time',ylab='Residuals',main="Time series plot of the
   residuals
     with AR(1) component in the model")
```



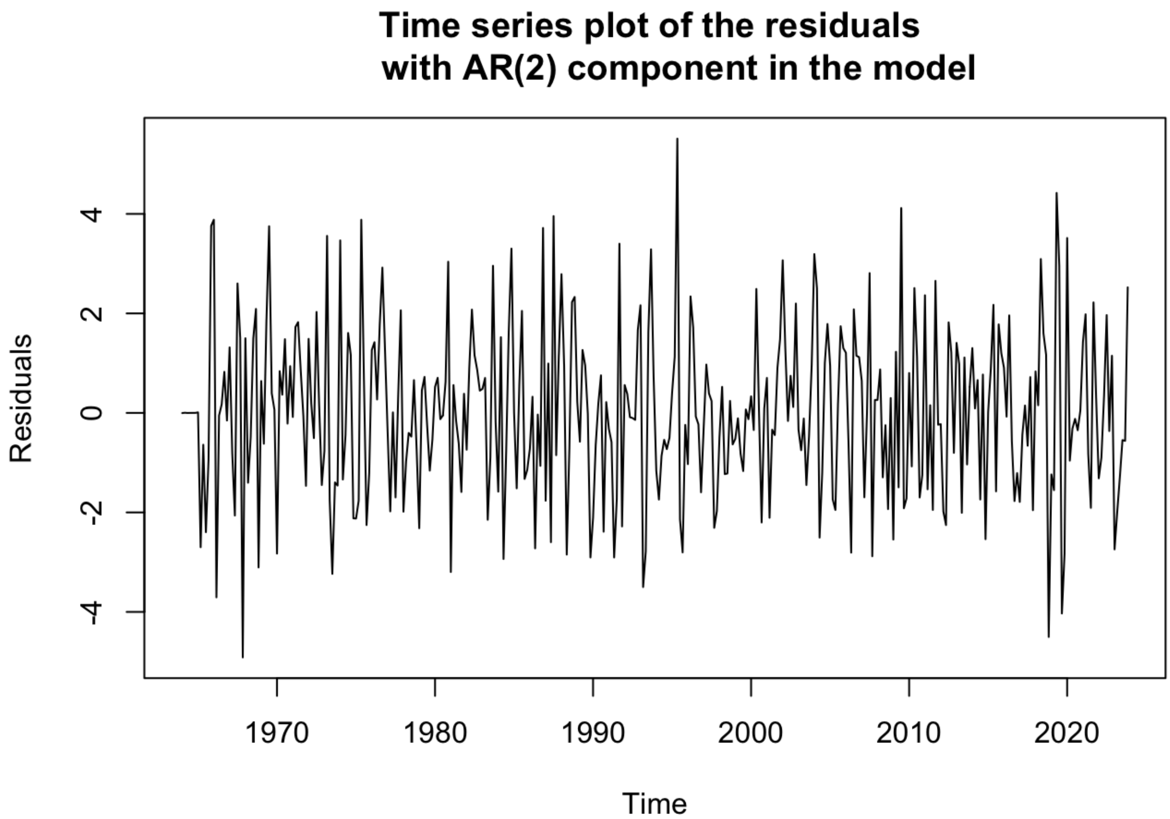

```
acf(res.m4, lag.max = 36, main = "The sample ACF of the residuals
    with AR(1) component in the model")
```



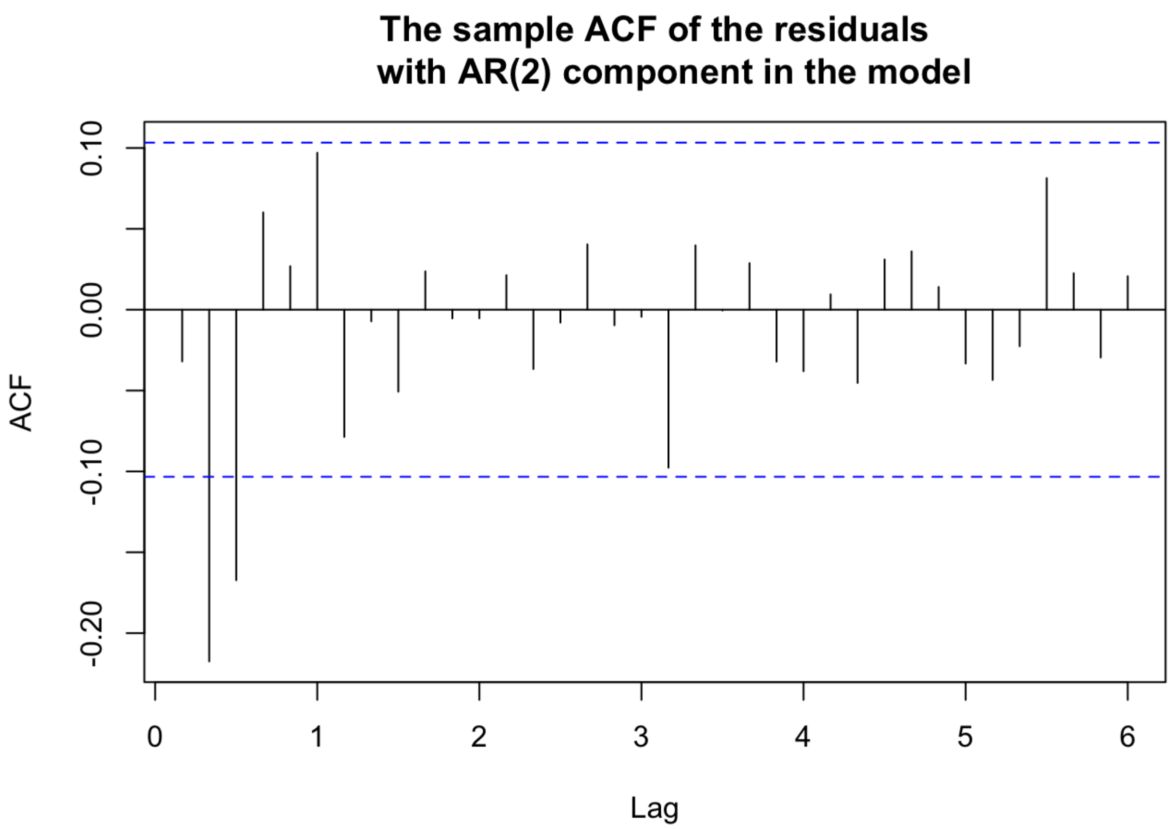
```
pacf(res.m4, lag.max = 36, main = "The sample PACF of the residuals
     with AR(1) component in the model")
```



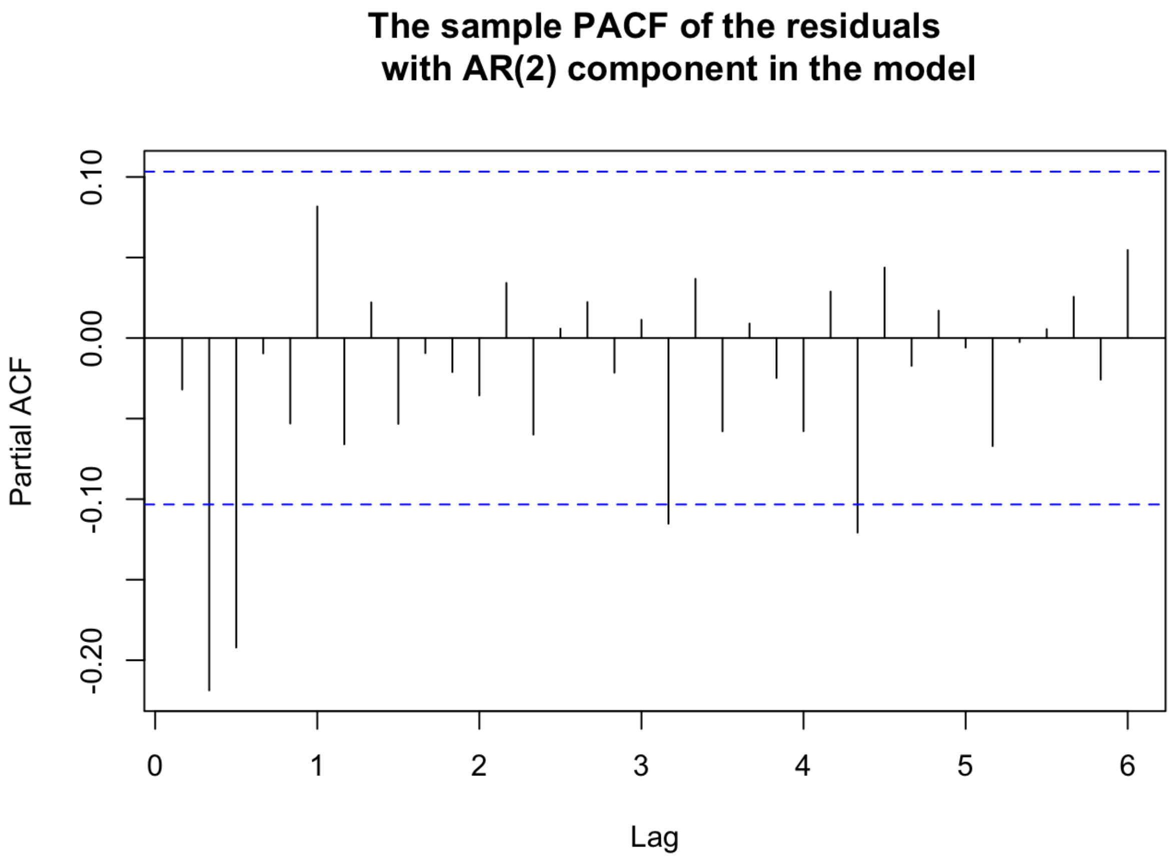

```
# SARIMA(2,1,0)x(0,1,1)_6 MODEL
m5.sim = arima(sarima.data,order=c(2,1,0),seasonal=list(order=c(0,1,1),
   period=6))
res.m5 = residuals(m5.sim);
plot(res.m5,xlab='Time',ylab='Residuals',main="Time series plot of the
   residuals
     with AR(2) component in the model")
```



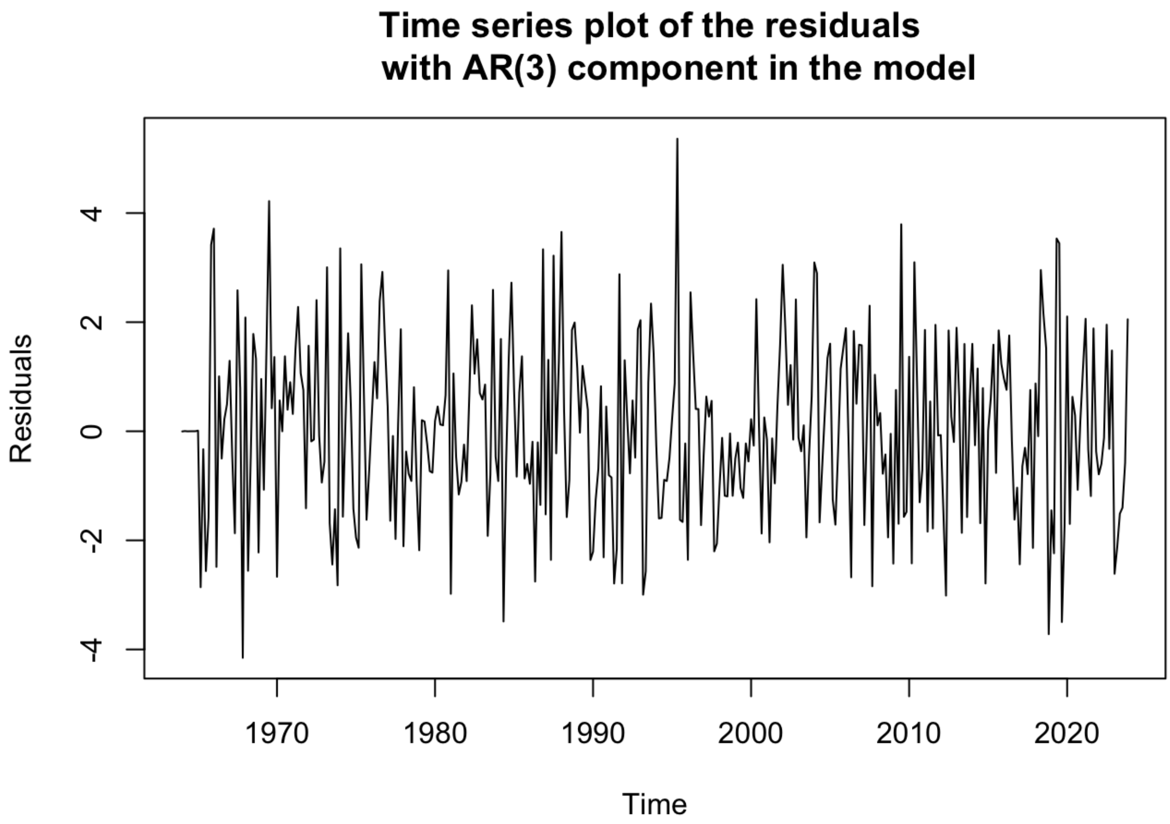

```
acf(res.m5, lag.max = 36, main = "The sample ACF of the residuals
    with AR(2) component in the model")
```



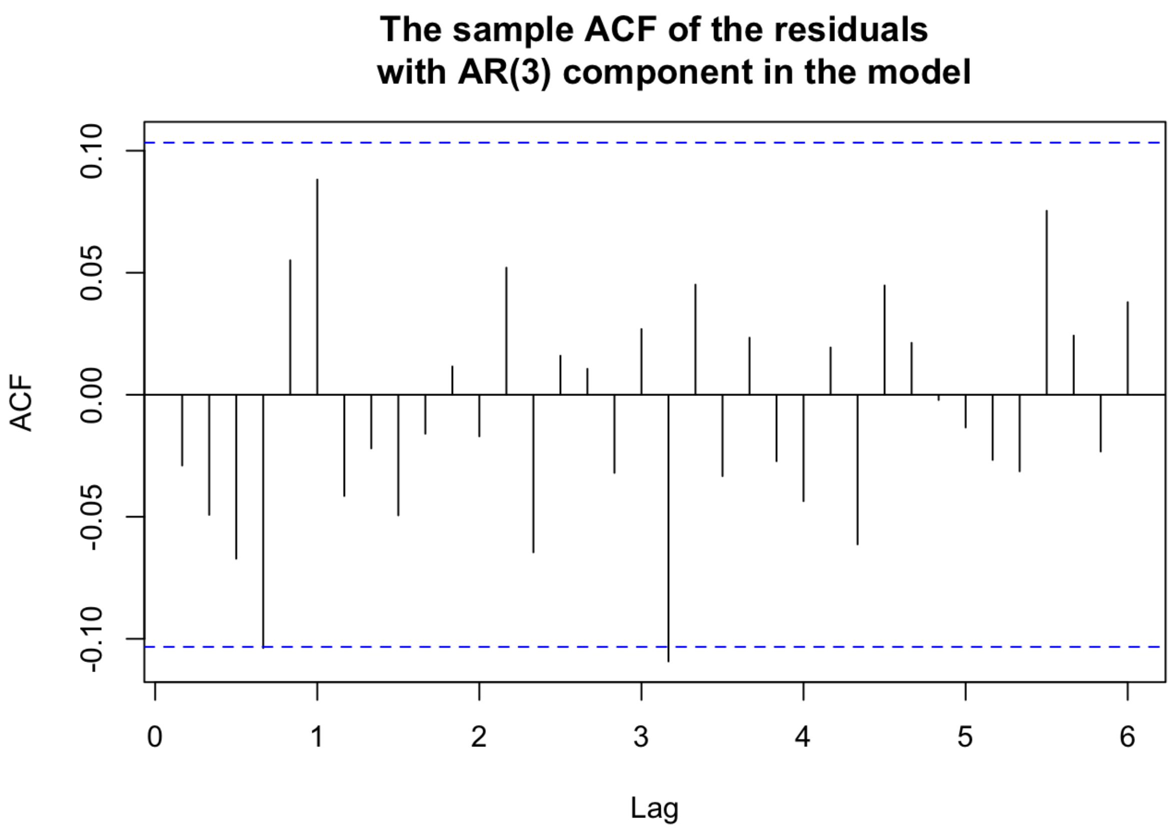
```
pacf(res.m5, lag.max = 36, main = "The sample PACF of the residuals
     with AR(2) component in the model")
```



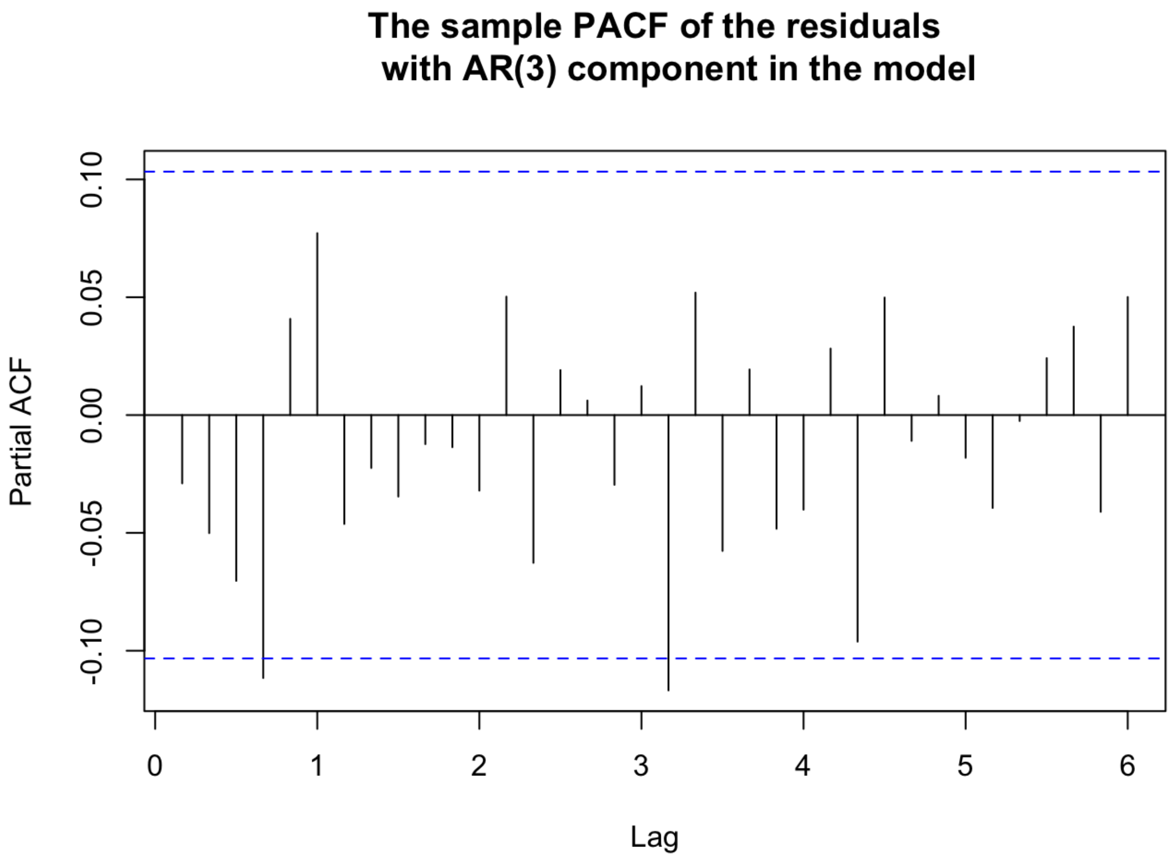

```
# SARIMA(3,1,0)x(0,1,1)_6 MODEL
m6.sim = arima(sarima.data,order=c(3,1,0),seasonal=list(order=c(0,1,1),
   period=6))
res.m6 = residuals(m6.sim);
plot(res.m6,xlab='Time',ylab='Residuals',main="Time series plot of the
   residuals
     with AR(3) component in the model")
```



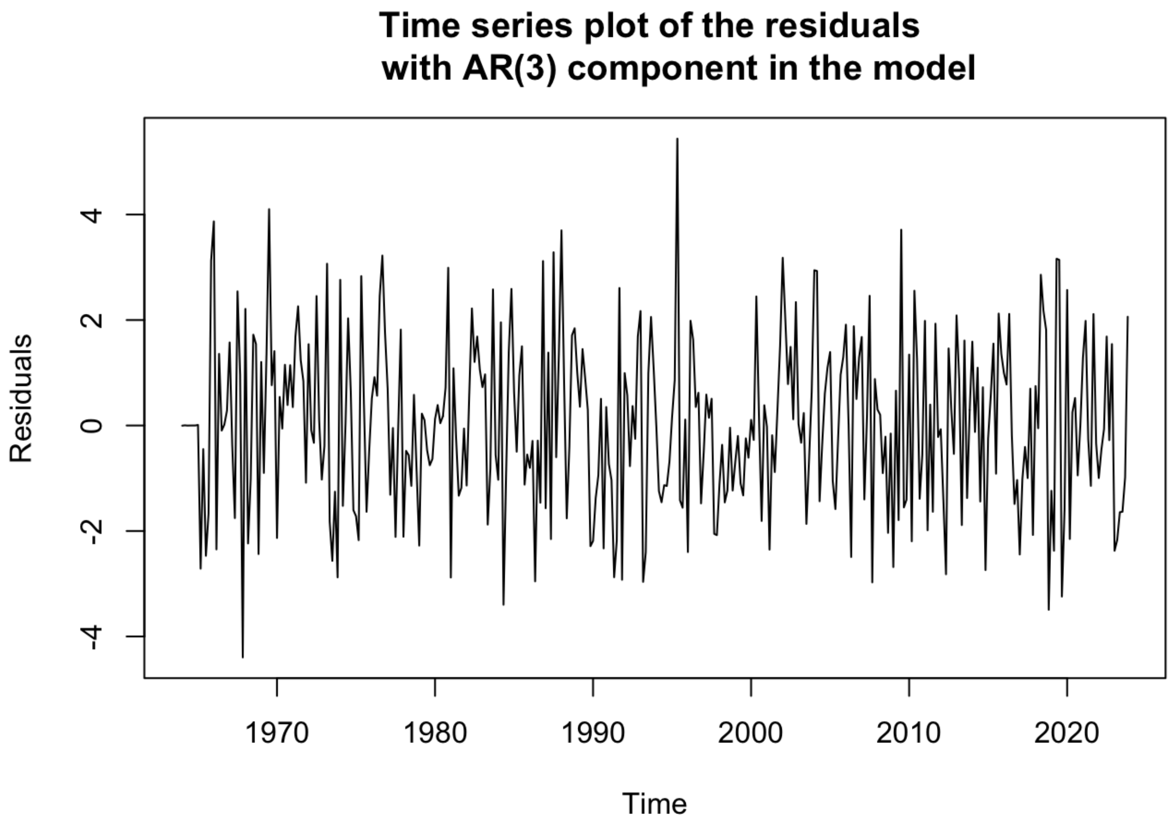

```
acf(res.m6, lag.max = 36, main = "The sample ACF of the residuals
    with AR(3) component in the model")
```



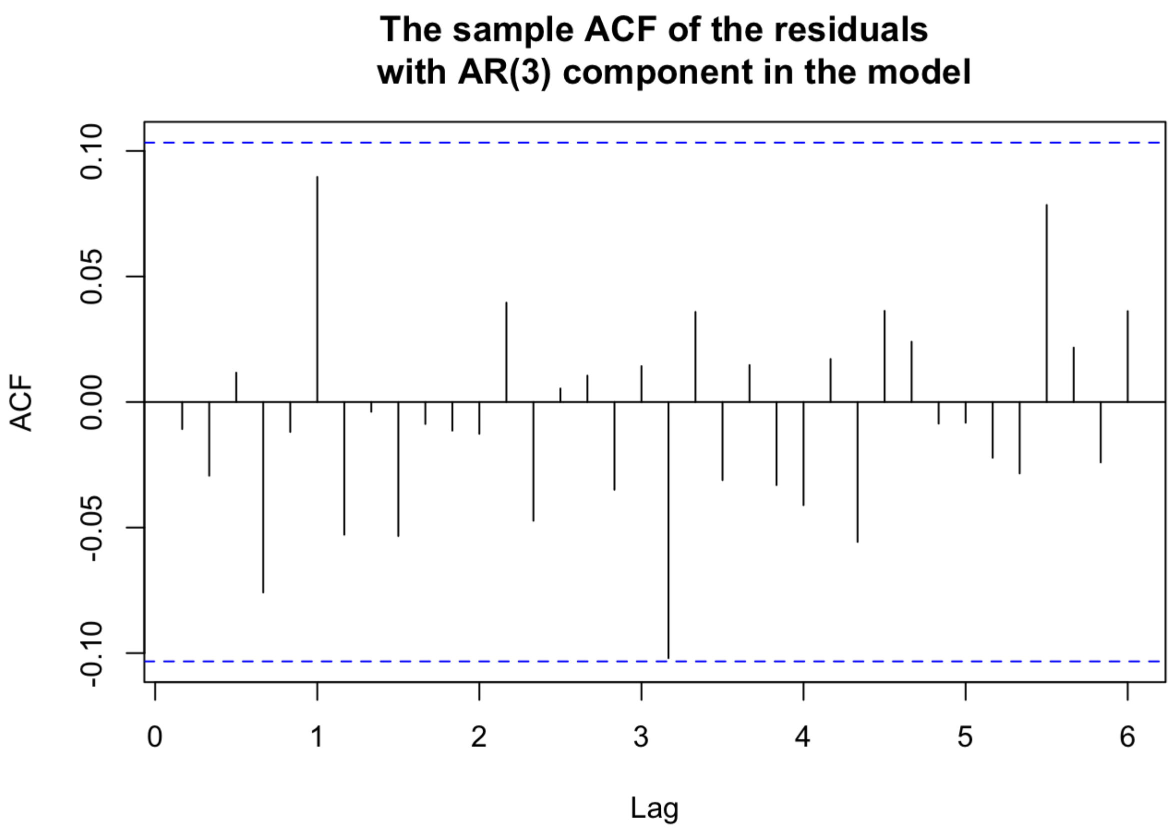
```
pacf(res.m6, lag.max = 36, main = "The sample PACF of the residuals
     with AR(3) component in the model")
```



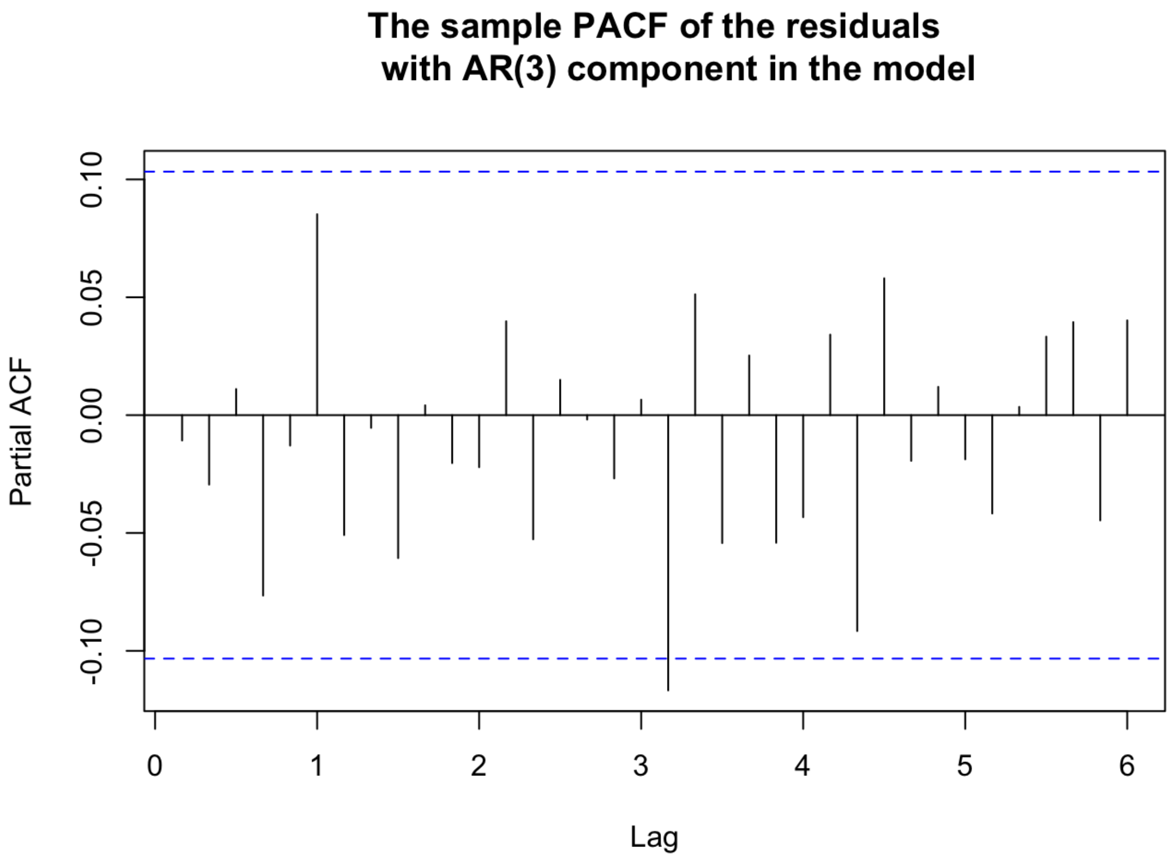

```
# SARIMA(4,1,0)x(0,1,1)_6 MODEL
m7.sim = arima(sarima.data,order=c(4,1,0),seasonal=list(order=c(0,1,1),
   period=6))
res.m7 = residuals(m7.sim);
plot(res.m7,xlab='Time',ylab='Residuals',main="Time series plot of the
   residuals
     with AR(3) component in the model")
```




```
acf(res.m7, lag.max = 36, main = "The sample ACF of the residuals
    with AR(3) component in the model")
```



```
pacf(res.m7, lag.max = 36, main = "The sample PACF of the residuals
     with AR(3) component in the model")
```



In the residuals of the model with either AR(1) or AR(2) components, we have highly significant first lag autocorrelations. Those for the model with AR(3) component, there are some slightly significant autocorrelations. It is the same as for the model with AR(4) component. Thus, we can conclude that for the orders p = 3 and p = 4, we get white noise residuals. Therefore, we conclude that the suitable models for this series are SARIMA(3,1,0)x$(0,1,1)_6$ and SARIMA(4,1,0)x$(0,1,1)_6$

Model Fitting

Model fitting for parameter estimation is straightforward by the use of R. We get ML estimates using the following code chunk.

```
m1.co2 = arima(co2,order=c(0,1,1),seasonal=list(order=c(0,1,1),
                                      period=12),method = "ML")
coeftest(m1.co2)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ma1  -0.579181   0.079075 -7.3244 2.399e-13 ***
## sma1 -0.820609   0.113731 -7.2154 5.378e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the standard error estimates are relatively smaller than the corresponding parameter estimates, we think that the coefficient estimates are all highly significant. This is also confirmed by the significance tests.

We can also use conditional least squares approach:

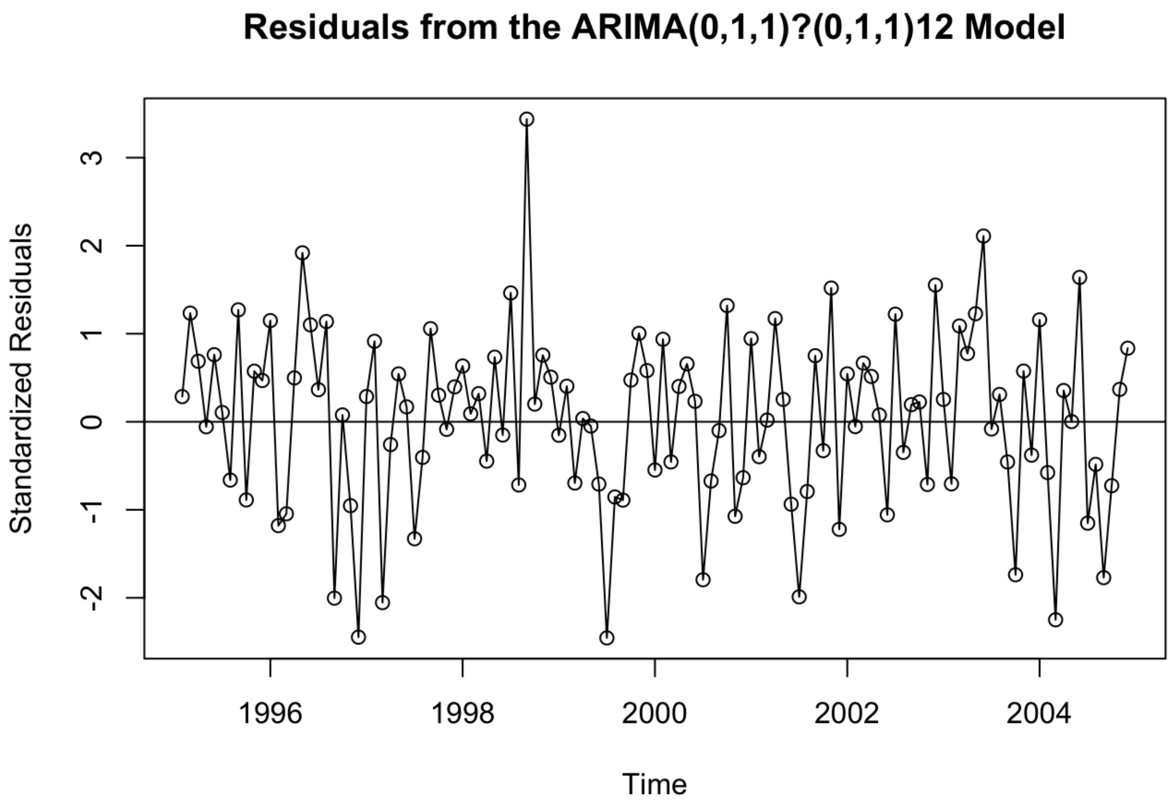
```
m1.co2.CSS = arima(co2,order=c(0,1,1),seasonal=list(order=c(0,1,1),
                                          period=12),method ="CSS")
coeftest(m1.co2.CSS)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ma1  -0.551228   0.074591  -7.390 1.468e-13 ***
## sma1 -0.719833   0.063052 -11.417 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnostic Checking

First, we look at the time series plot of residuals for the diagnostic checking. The following plot exhibits the time series plot for the standardised residuals.

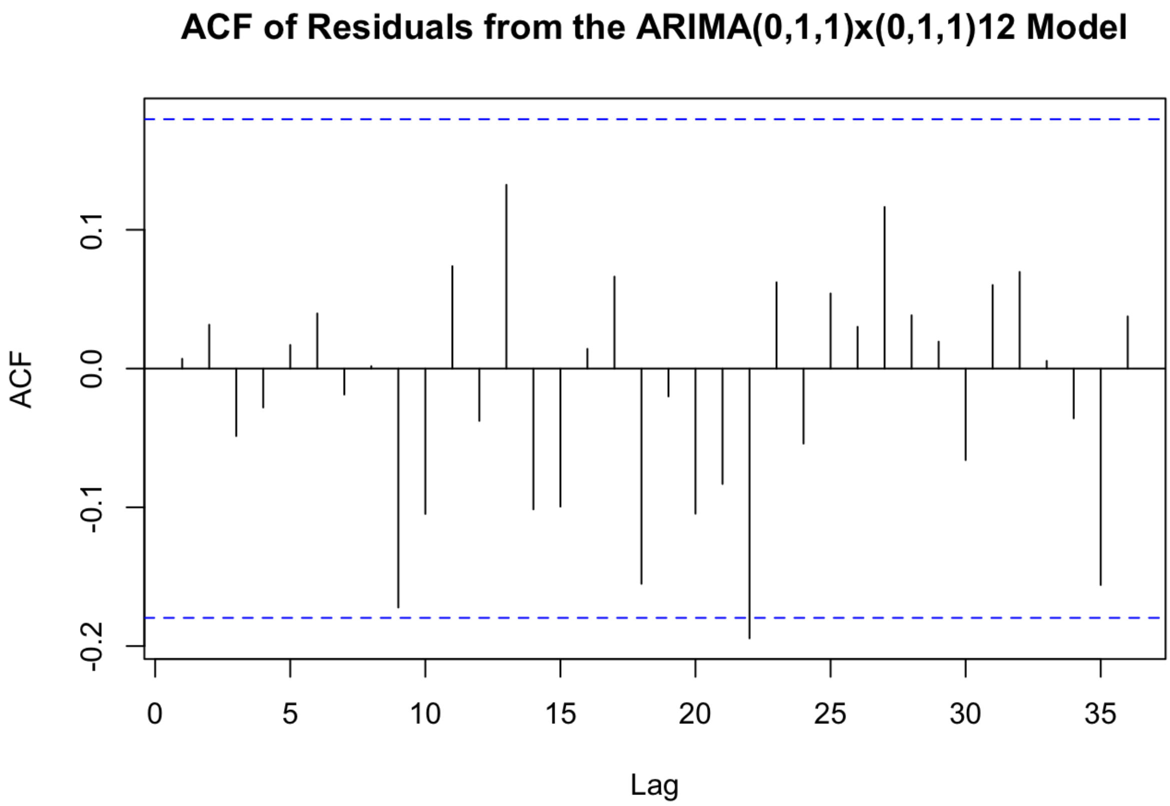
```
plot(window(rstandard(m1.co2),start=c(1995,2)), ylab='Standardized Resi
  duals',type='o', main="Residuals from the ARIMA(0,1,1)?(0,1,1)12 Mode
  l")
abline(h=0)
```



Although we have some strange behavior in the middle of the series, this plot does not suggest any major irregularities with the model. But we may need to investigate the model further for outliers, as the standardized residual at September 1998 looks suspicious.

To go further, we plot the ACF of stadardised residuals.

```
acf(as.vector(window(rstandard(m1.co2),start=c(1995,2))),lag.max=36,mai
  n="ACF of Residuals from the ARIMA(0,1,1)x(0,1,1)12 Model")
```



Except for the slightly significant autocorrelation at lag 22, there is not any sign of violation of the independence of residuals.

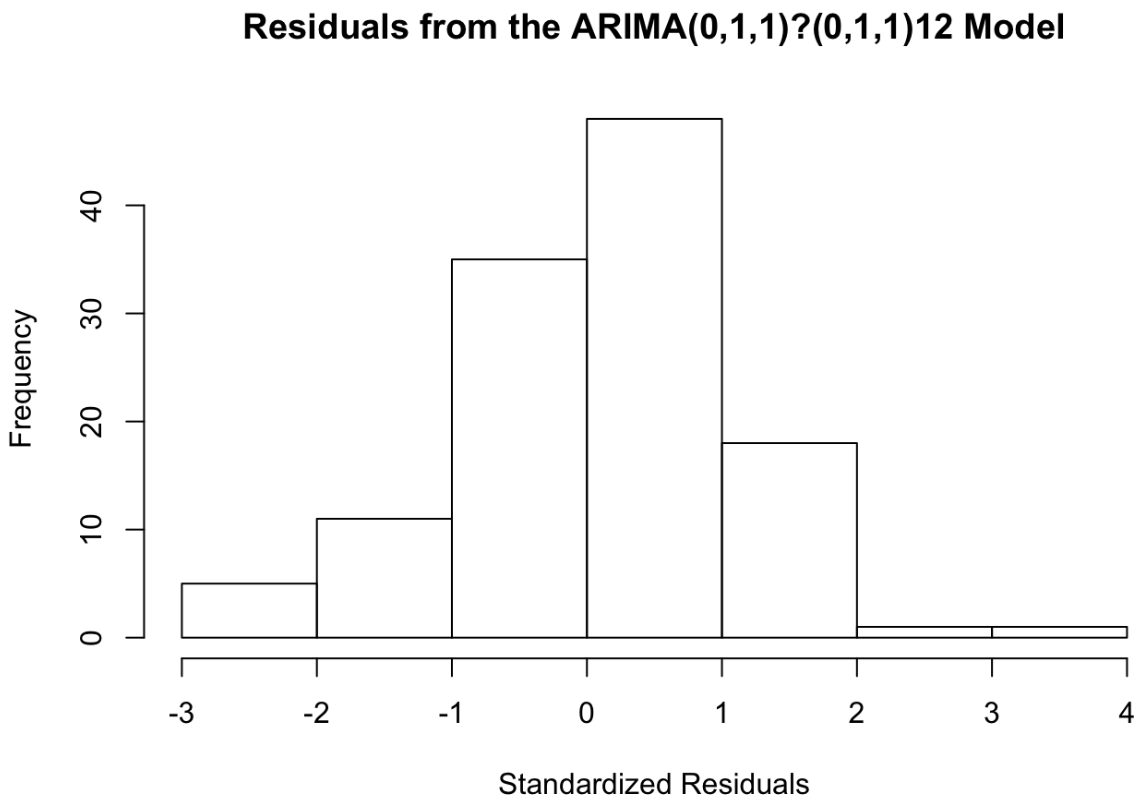
To get an overall look, we apply the Ljung-Box test. We observe that there is no problem in terms of independence of errors in the overall sense.

```
Box.test(window(rstandard(m1.co2),start=c(1995,2)), lag = 22, type = "L
  jung-Box", fitdf = 0)
```

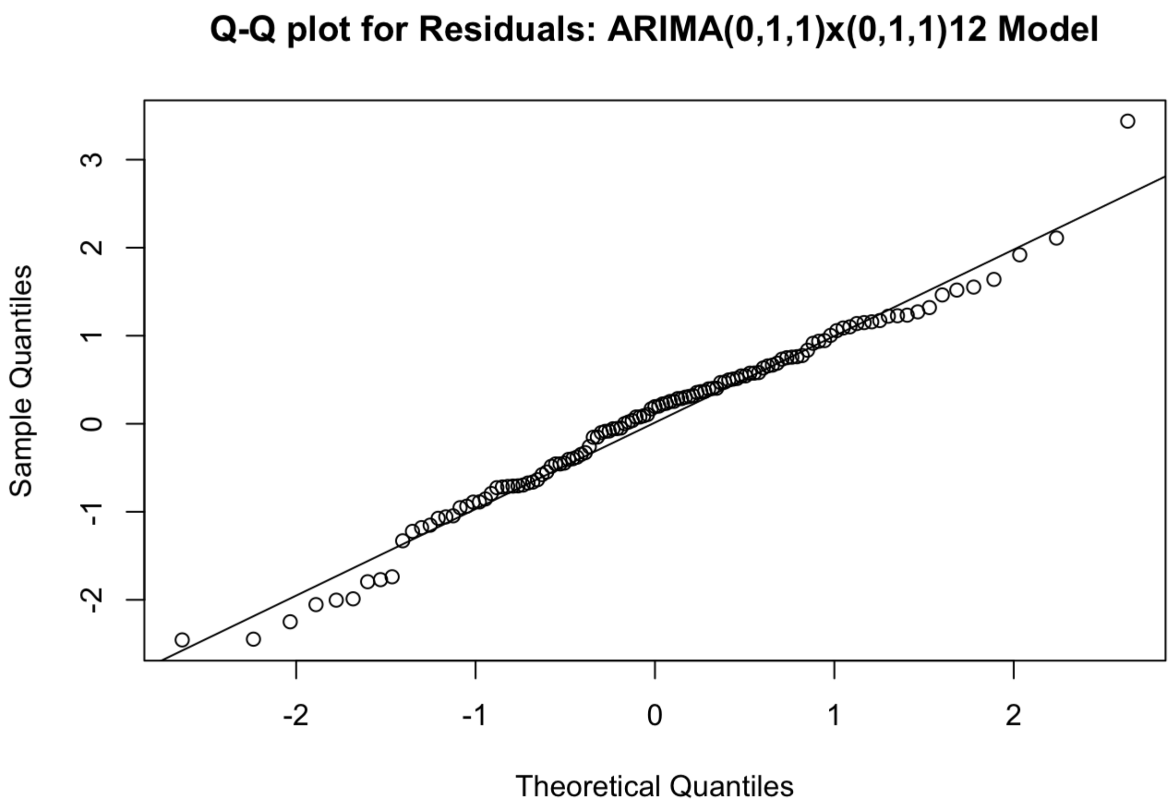
```
##
##  Box-Ljung test
##
## data:  window(rstandard(m1.co2), start = c(1995, 2))
## X-squared = 24.564, df = 22, p-value = 0.3184
```

To investigate the normality assumption, histogram and Q-Q plot of the residuals are examined and there is no suspicious but the outlier in the upper tail.

```
hist(window(rstandard(m1.co2),start=c(1995,2)),xlab='Standardized Resid
  uals',main="Residuals from the ARIMA(0,1,1)?(0,1,1)12 Model")
```




```
qqnorm(window(rstandard(m1.co2),start=c(1995,2)),main="Q-Q plot for Res
  iduals: ARIMA(0,1,1)x(0,1,1)12 Model")
qqline(window(rstandard(m1.co2),start=c(1995,2)))
```



Then, we apply the Shapiro-Wilk test to check normality.

```
shapiro.test(window(rstandard(m1.co2),start=c(1995,2)))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  window(rstandard(m1.co2), start = c(1995, 2))
## W = 0.98205, p-value = 0.1134
```

The normality of residuals is not rejected at 5% significance level with the Shapiro-Wilk test.

Lastly, we consider overfitting by fitting an ARIMA(0,1,2)×(0,1,1)_12 model with the following code chunk.

```
m2.co2=arima(co2,order=c(0,1,2),seasonal=list(order=c(0,1,1),period=12
  ))
coeftest(m2.co2)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ma1  -0.571436   0.089695 -6.3708 1.880e-10 ***
## ma2  -0.016474   0.094806 -0.1738     0.862
## sma1 -0.827381   0.122404 -6.7594 1.385e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When we compare these results with those obtained by the ARIMA(0,1,1)×(0,1,1)$_{12}$ model, we observe that the estimates of θ_1 and Θ have changed very little-especially when the size of the standard errors is taken into consideration. In addition, the estimate of the new parameter, θ_2, is not statistically significant. Also, the estimate and the log-likelihood have not changed much while we have a worse (increased) AIC value

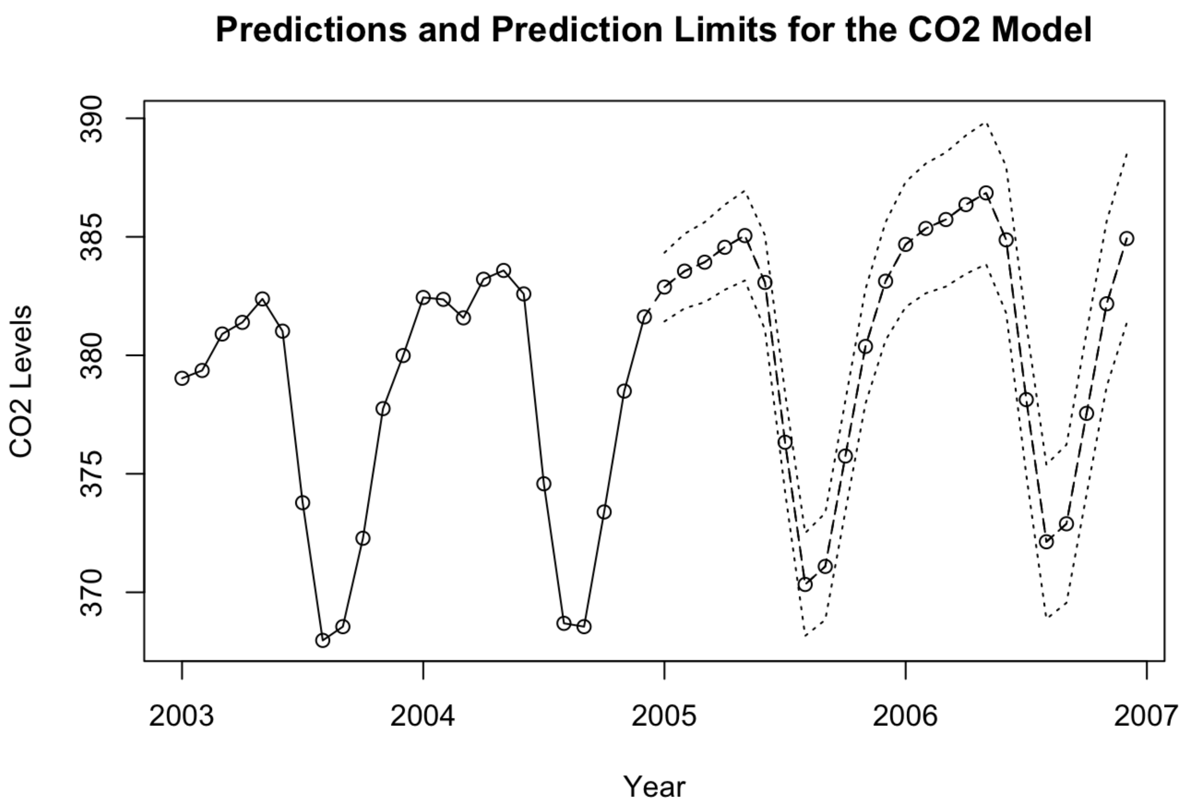
```
AIC(m1.co2,m2.co2)
```

```
##        df      AIC
## m1.co2  3 285.0769
## m2.co2  4 287.0466
```

Forecasting with Seasonal Models

We will consider prediction limits for the predictions of future realizations of time series. Let's consider the carbon dioxide series. The next plot shows the forecasts and 95% forecast limits for a lead time of two years for the ARIMA(0,1,1)×(0,1,1)$_{12}$ model that we fit.

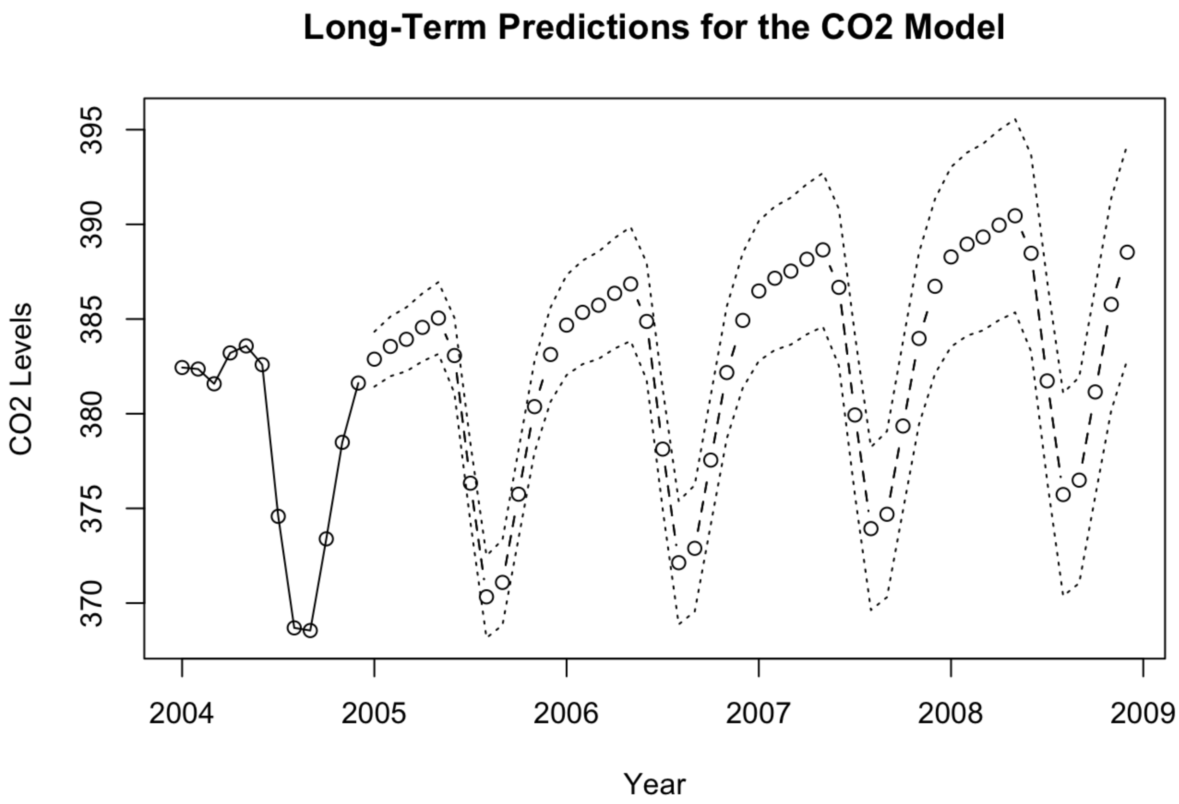
```
plot(m1.co2,n1=c(2003,1),n.ahead=24,xlab='Year',type='o', ylab='CO2 Lev
  els', main = "Predictions and Prediction Limits for the CO2 Model")
```



The last two years of observed data are also shown. The forecasts mimic the stochastic periodicity in the data quite well, and the forecast limits give a good feeling for the precision of the forecasts.

The following plot shows the long term predictions of the carbon dioxide series.

```
plot(m1.co2,n1=c(2004,1),n.ahead=48,xlab='Year',type='b',ylab='CO2 Leve
  ls', main="Long-Term Predictions for the CO2 Model")
```



The forecast limits are getting wider, as there is more uncertainty in the forecasts.

Summary

In this module, we investigated multiplicative seasonal ARIMA models that provide an economical way to model time series whose seasonal tendencies are not as regular as we would have with a deterministic seasonal trend model.

Then, we proceed with the classical steps of time series modelling:

- Model specification,
- Model fitting,
- Diagnostic checking, and
- Use of the model to produce predictions for the next realizations of the series.