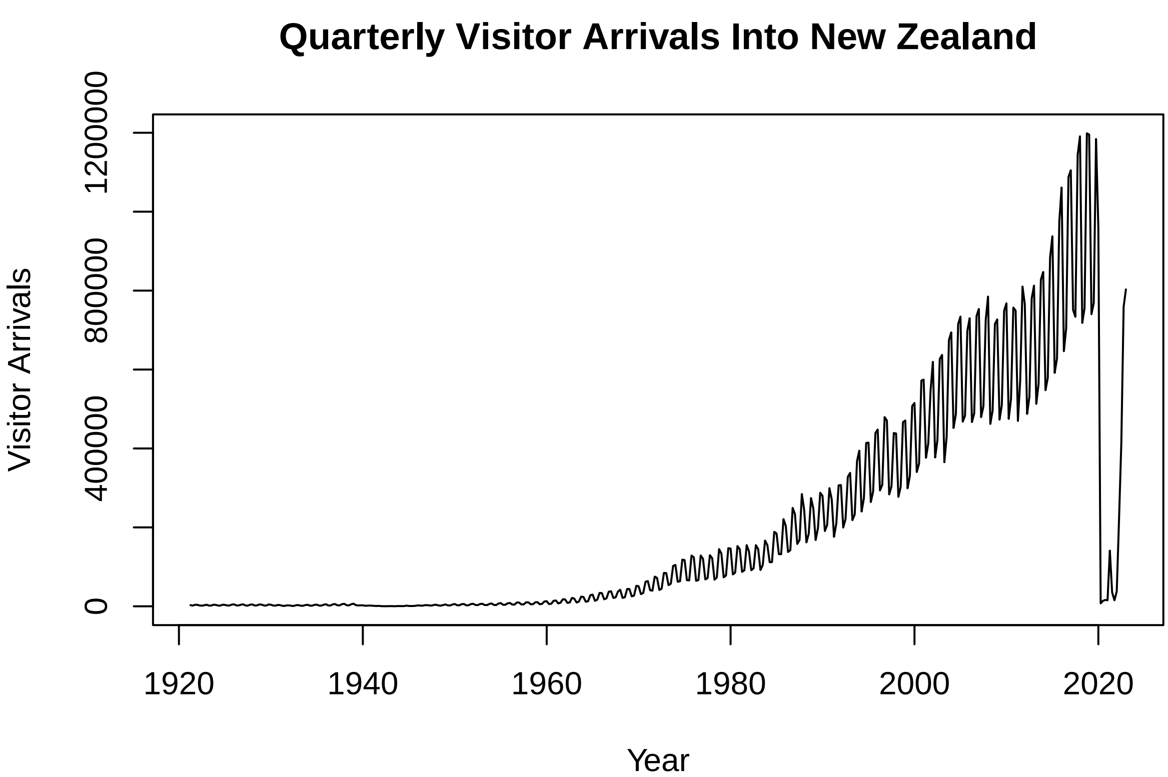
STAT317 Assignment3

2023-09-29

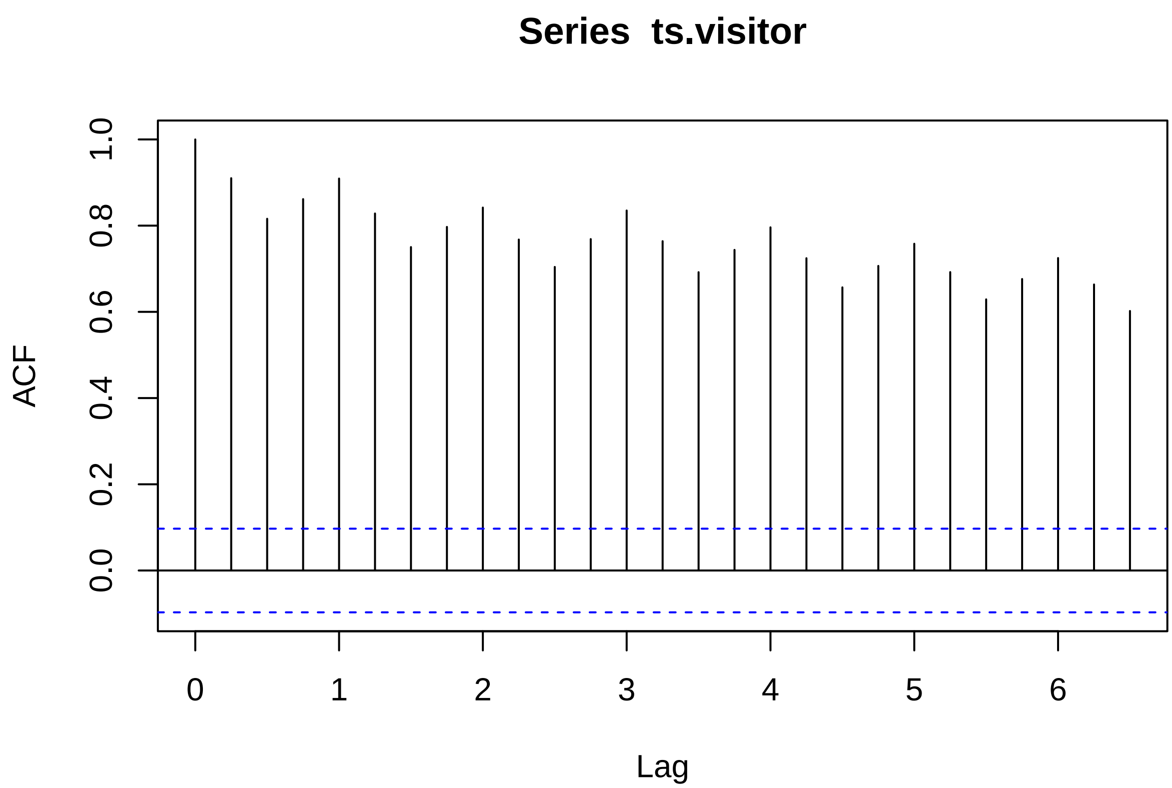
```r
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.3.1

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

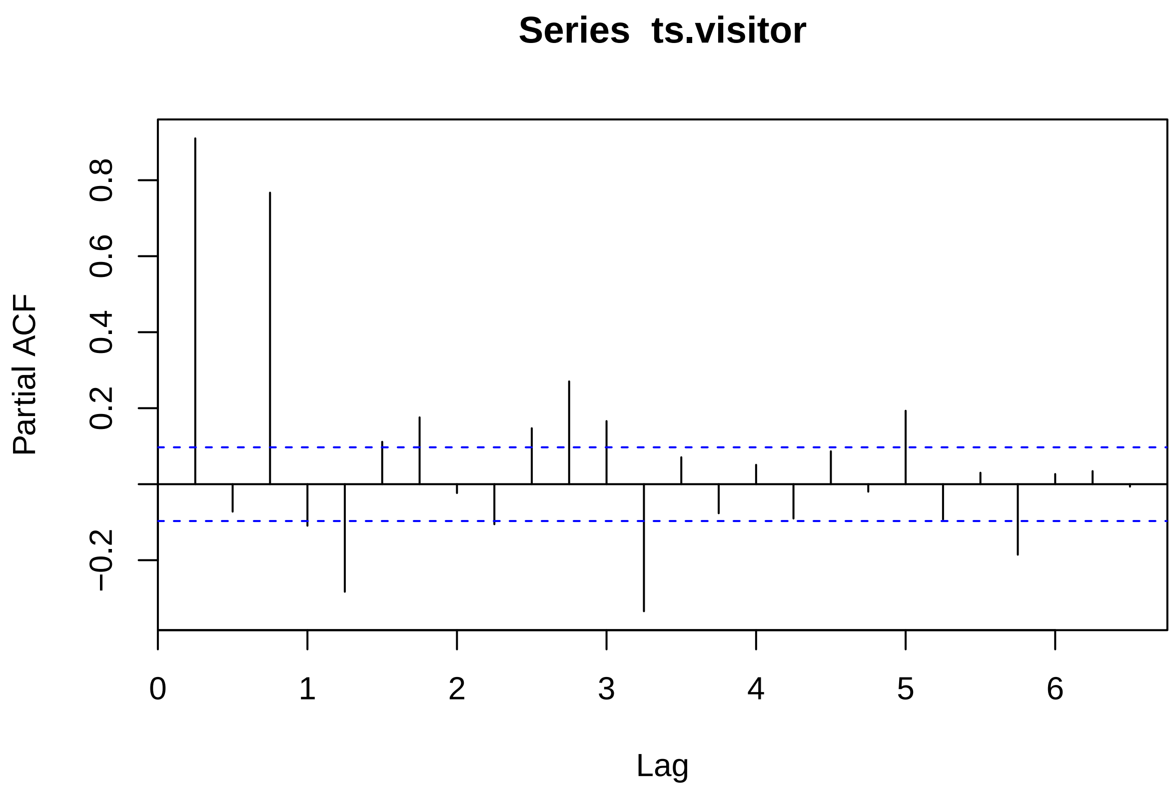
```r
visitor <- read.csv("Visitors.csv")
ts.visitor <- ts(visitor$Actual.Counts, start= c(1921, 2), frequency = 4)
ts.plot(ts.visitor, xlab = "Year", ylab = "Visitor Arrivals", main =
"Quarterly Visitor Arrivals Into New Zealand")
```



```
acf(ts.visitor)
```



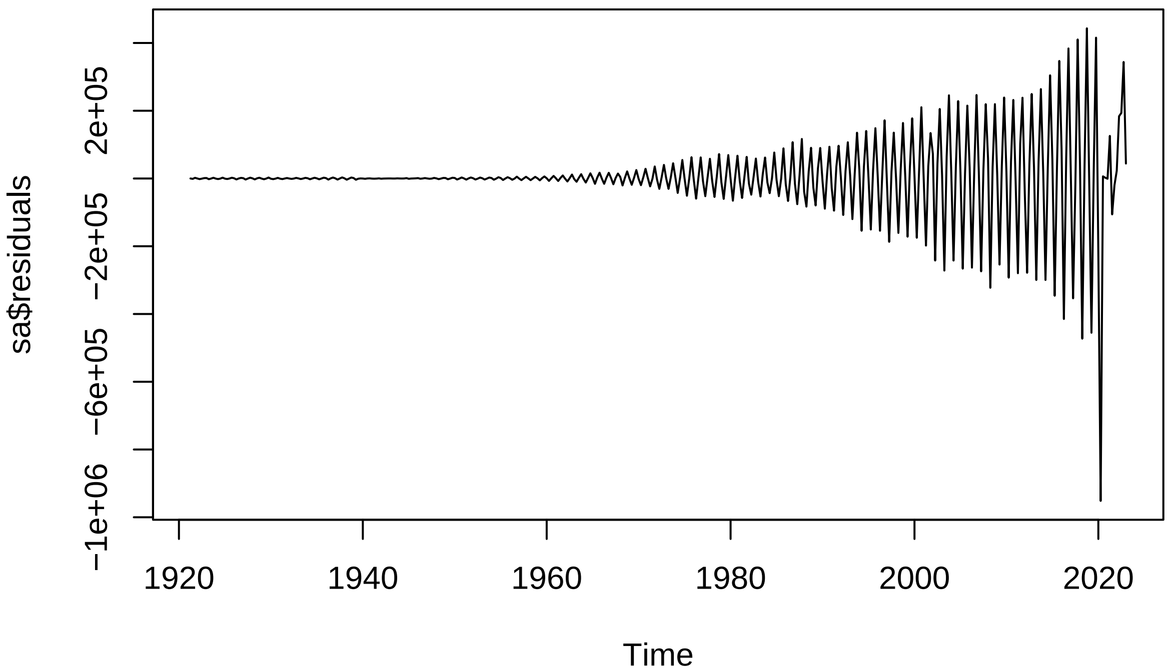

```
pacf(ts.visitor)
```



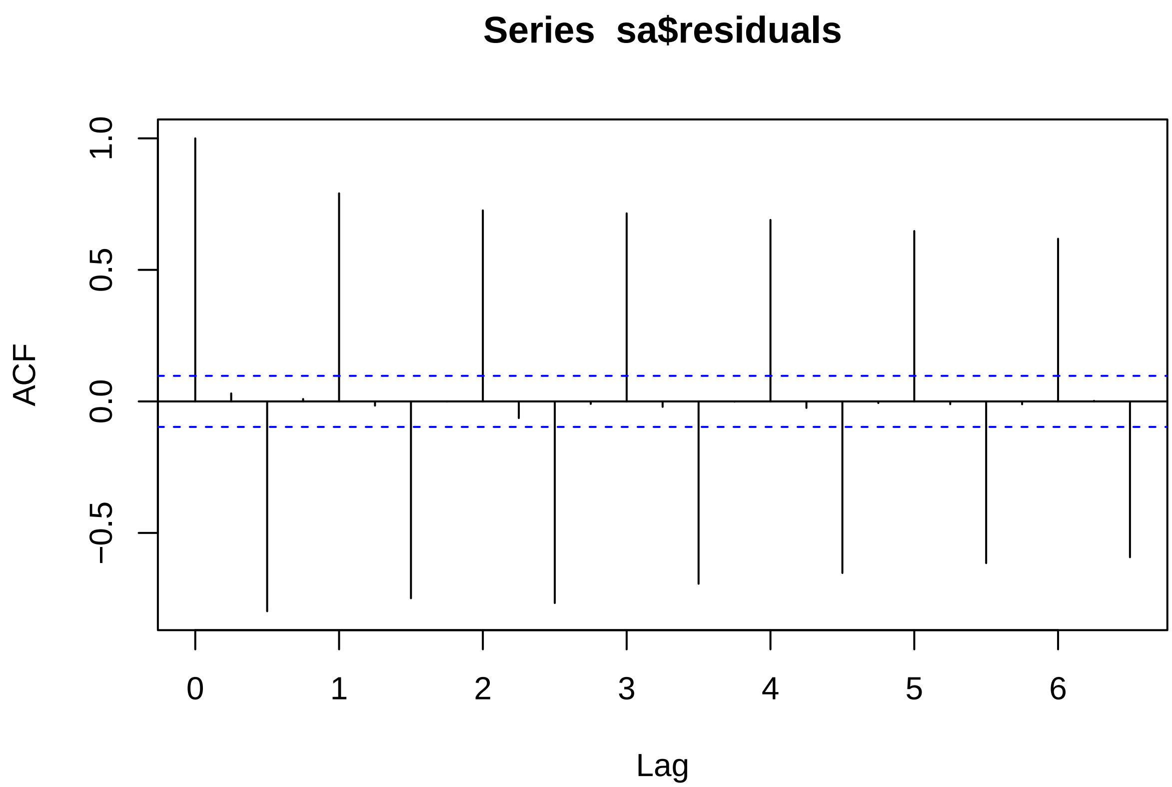
The plot has a clear trend and therefore the series is not stationary. We should then take the first difference to try eliminate the trend.

The diferenced plot does not have a positive trend like the original series. This is because the trend is removed when we take the first difference of a series. The new series does not have a constant variance, so the series is still not stationary. we can take the log and first difference of the series to try have a stationary series.

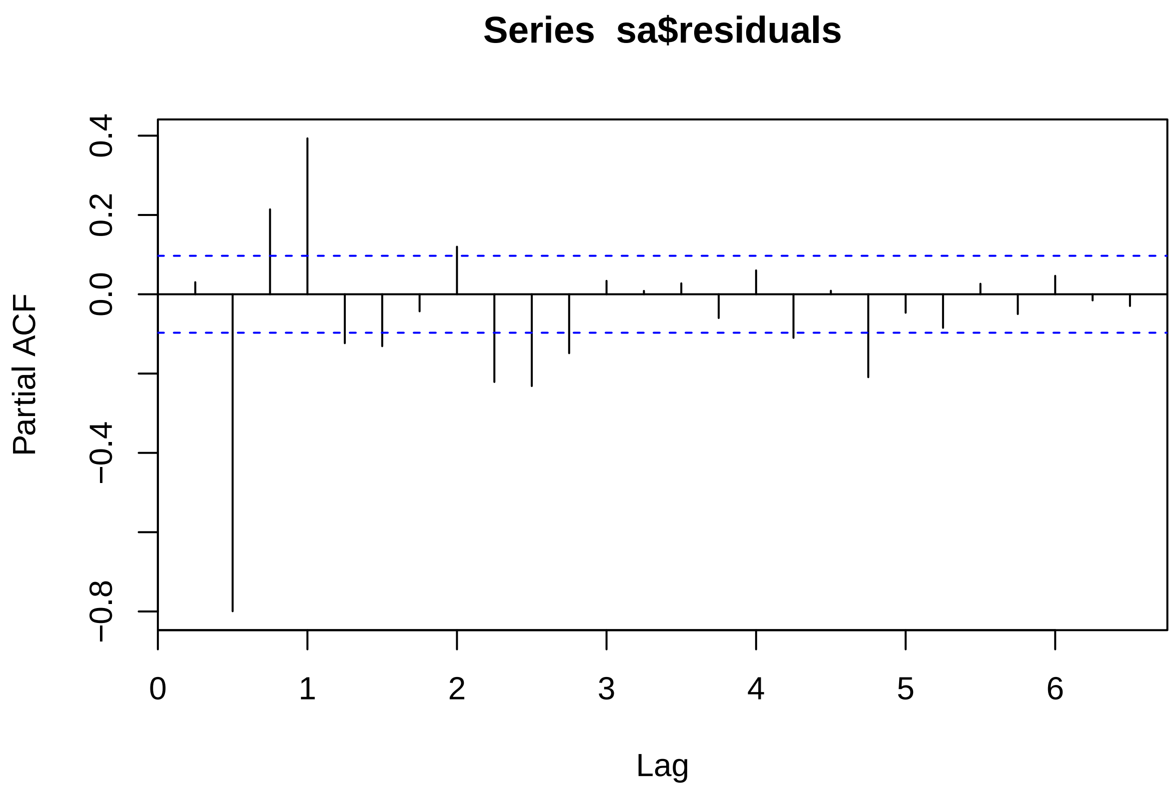
```
sa <- arima(ts.visitor, order=c(0,1,0),seasonal = list(order=c(0,0,0), period=4))
ts.plot(sa$residuals)
```




```
acf(sa$residuals)
```

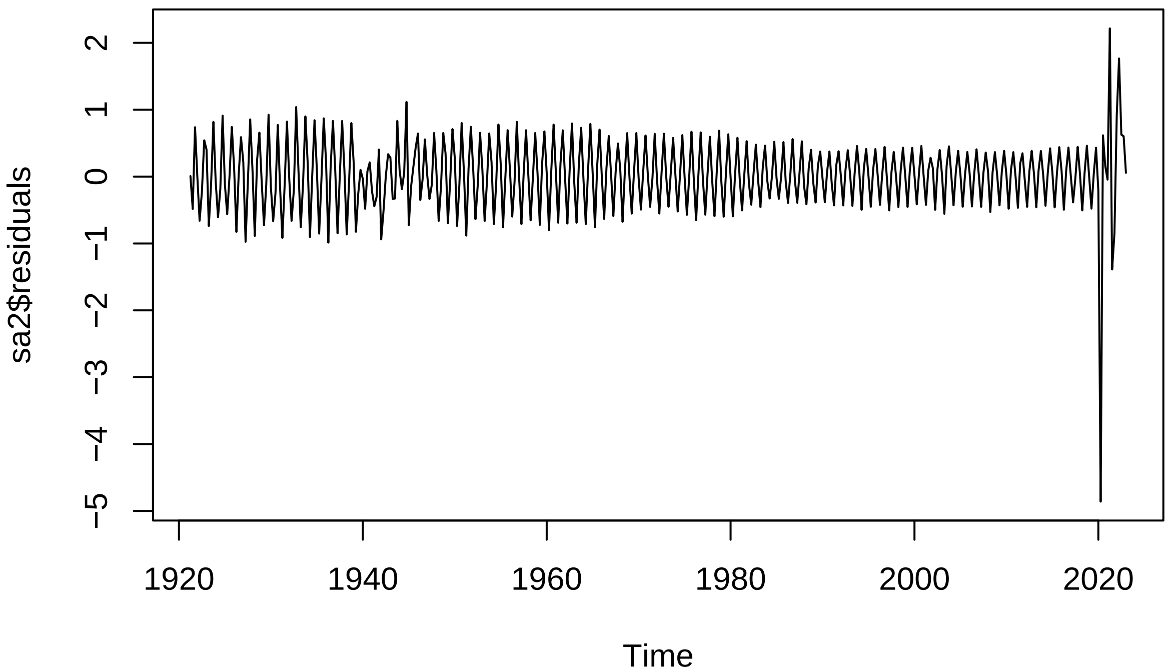



```
pacf(sa$residuals)
```

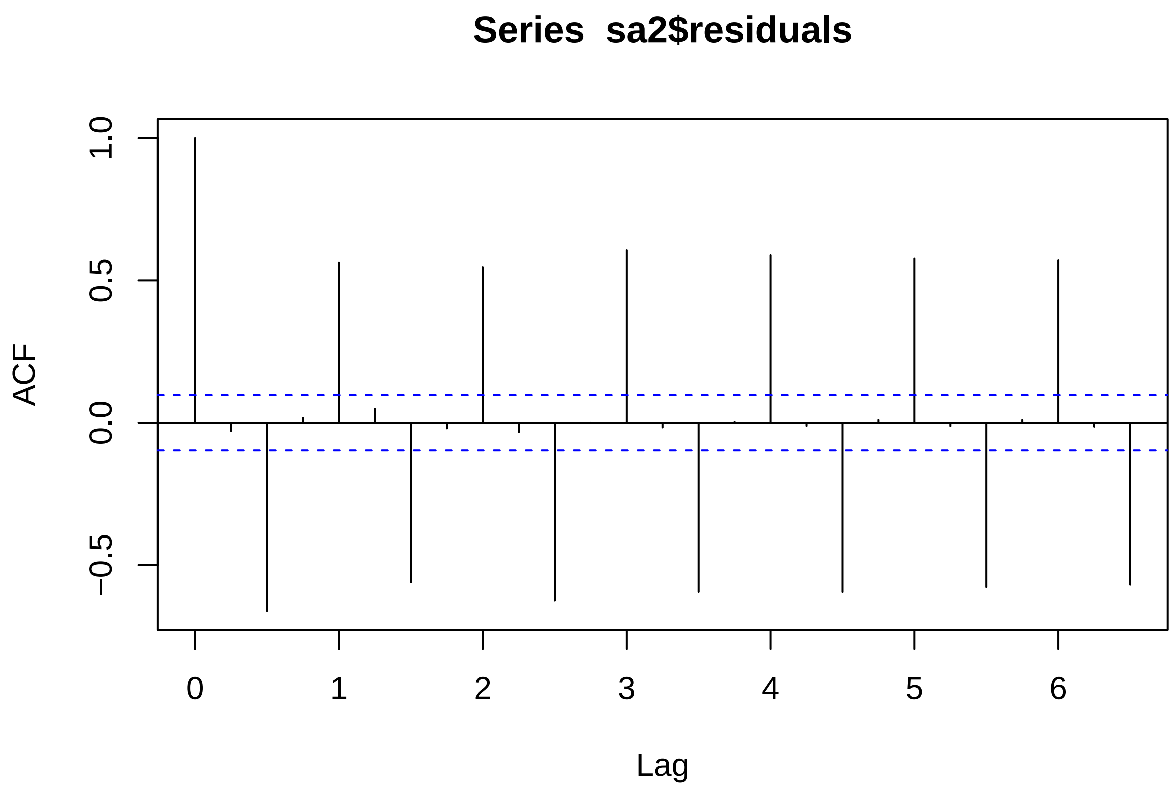


The series is still not stationary as the variance is not constant. We can take the log of the series to try reach a constant variance

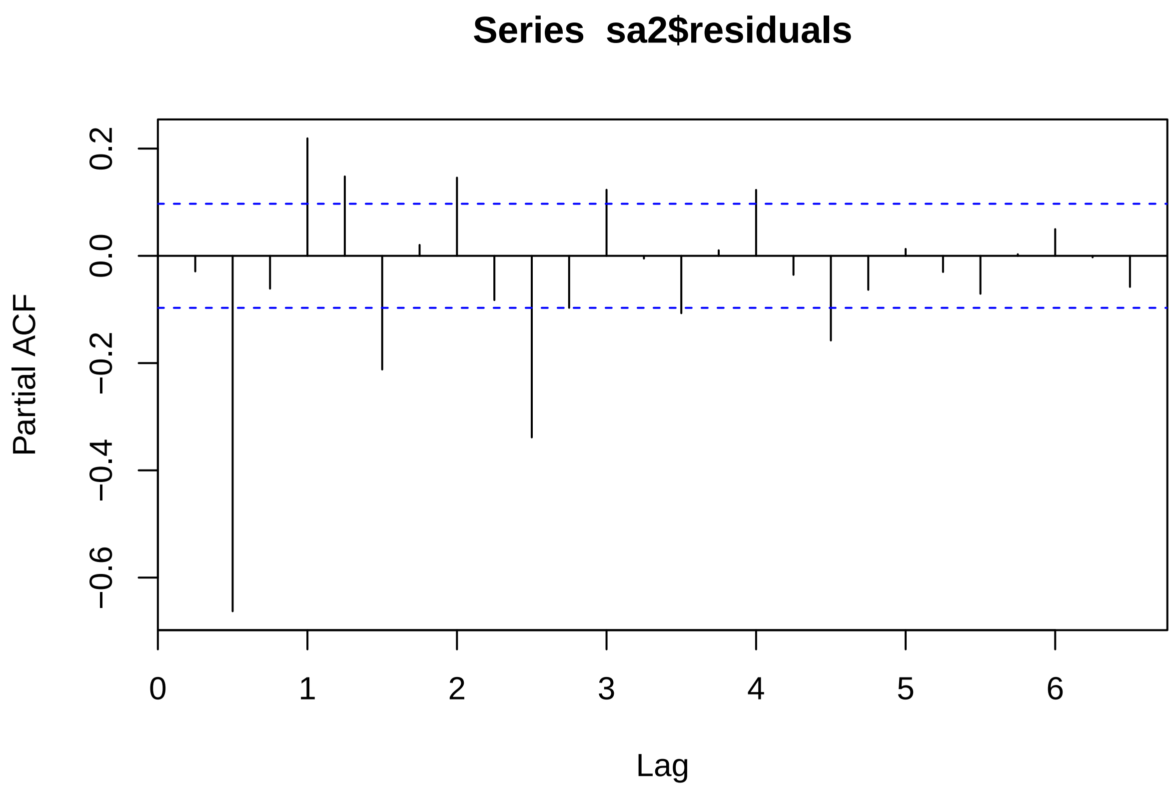
```
logvis = log(ts.visitor)
sa2 <- arima(logvis, order=c(0,1,0), seasonal = list(order=c(0,0,0), period=4))
ts.plot(sa2$residuals)
```



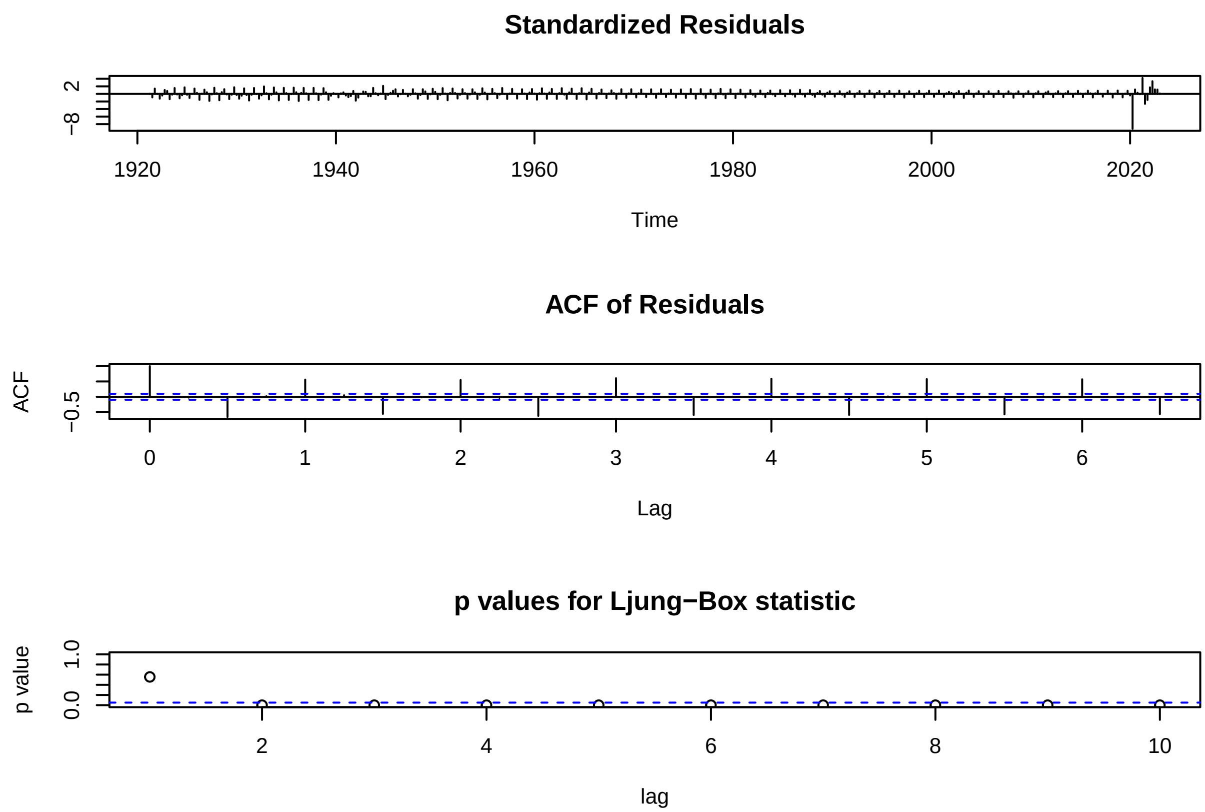

```
acf(sa2$residuals)
```




```
pacf(sa2$residuals)
```



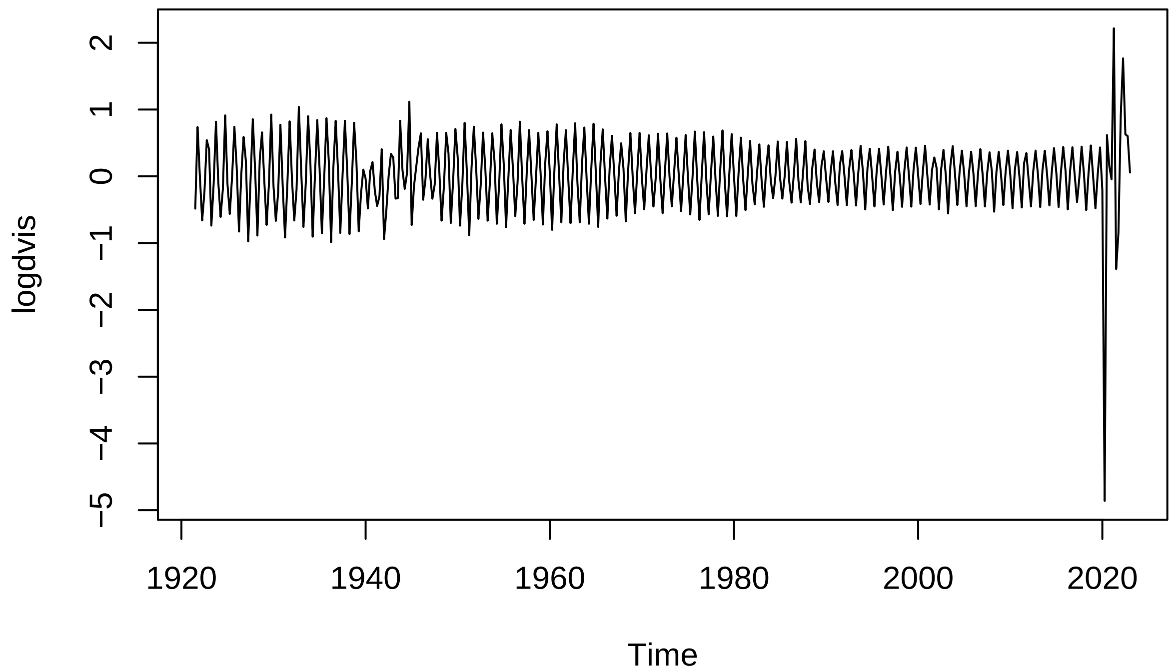

```
tsdiag(sa2)
```



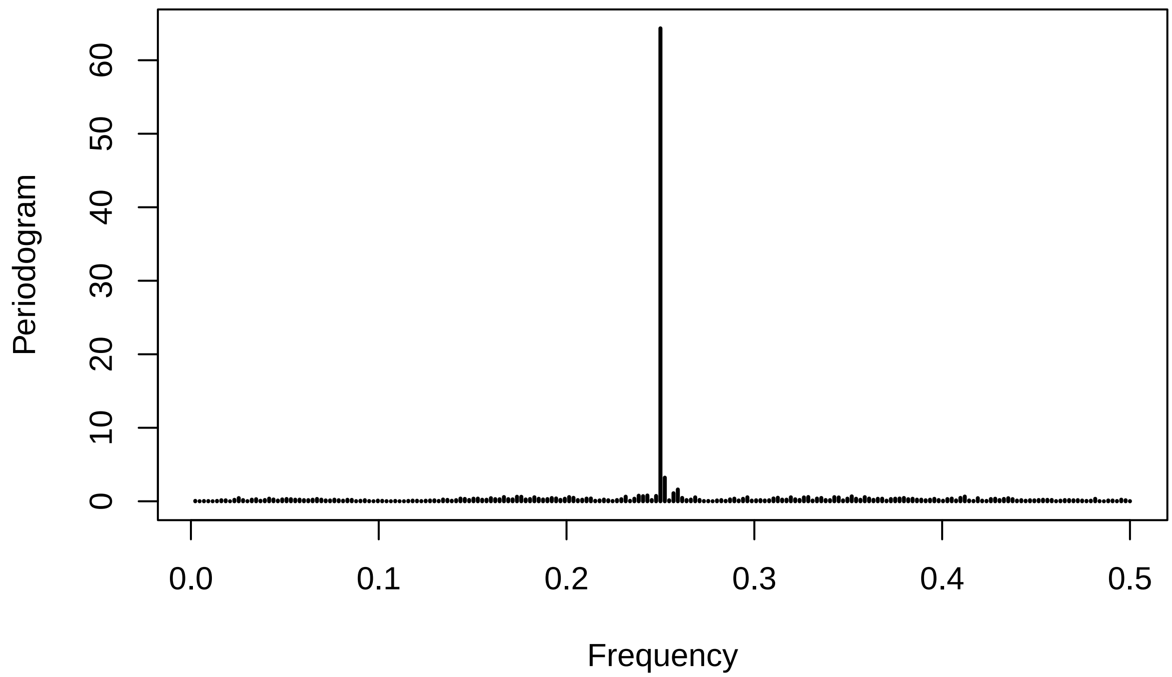

```
library(TSA)
```

```
## Warning: package 'TSA' was built under R version 4.3.1

## Registered S3 methods overwritten by 'TSA':
##   method       from
##   fitted.Arima forecast
##   plot.Arima   forecast

##
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':
##
##     acf, arima

## The following object is masked from 'package:utils':
##
##     tar
```

```
logdvis <- diff(logvis)
plot(logdvis)
```



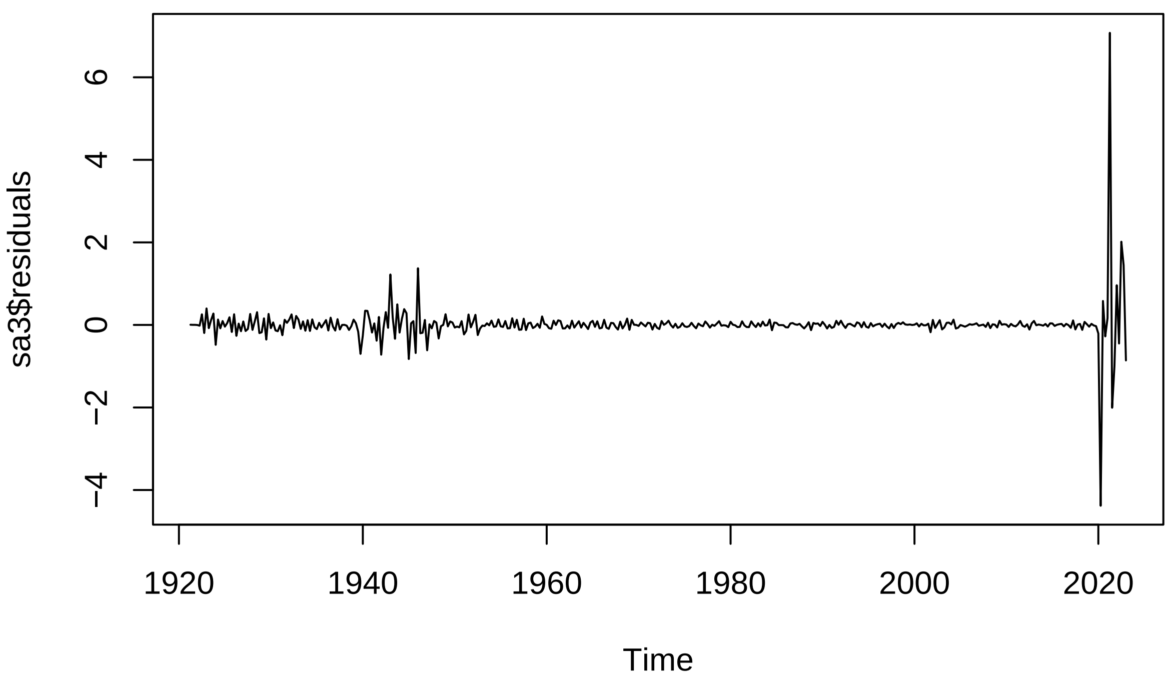

```
periodogram(logdvis)
```



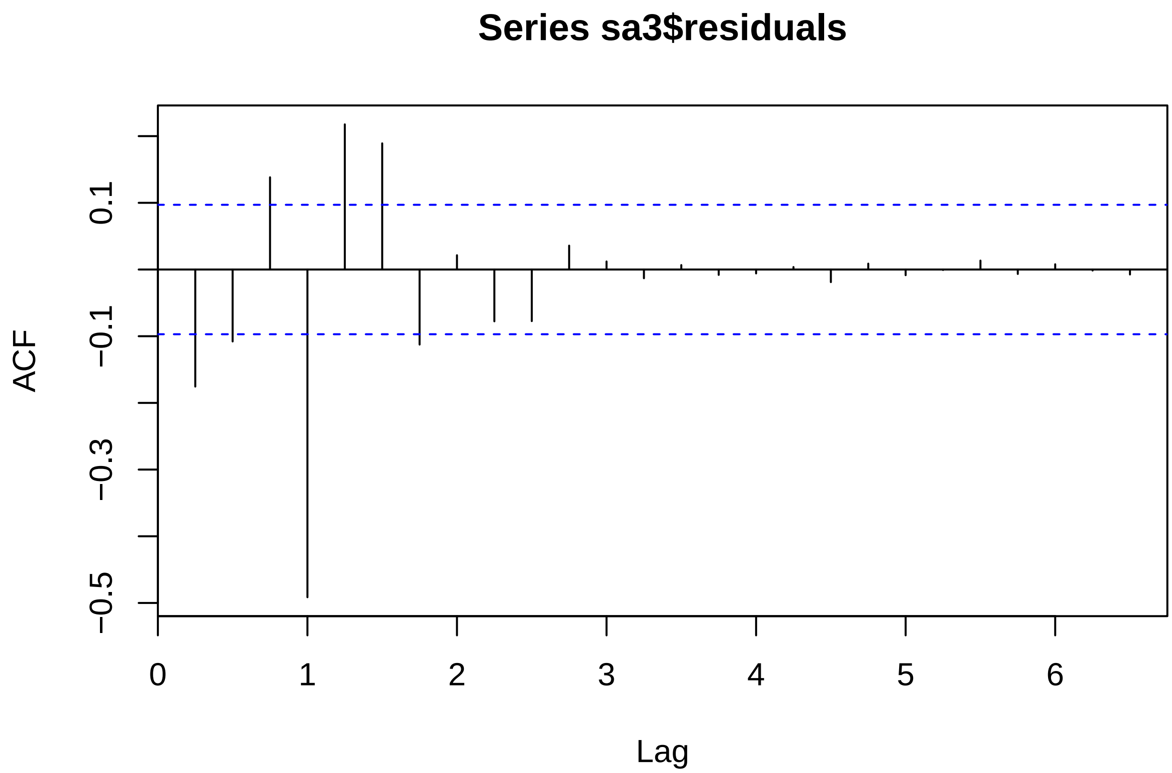
The series now looks stationary as the mean and variance appear to be constant. The ACF shows a clear seasonal trend that we need to get rid of.

The periodigram backs up this observation as it clearly shows a peak every four quarters

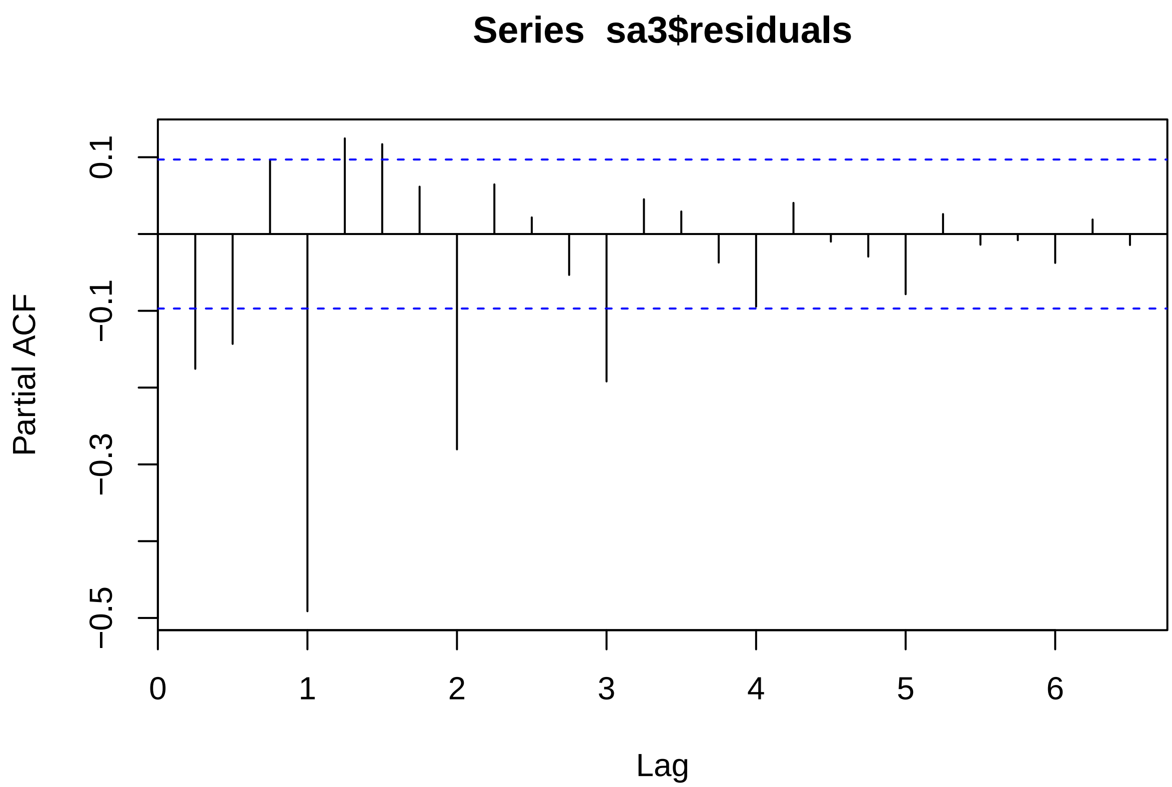
```
sa3 <- arima(logvis, order=c(0,1,0), seasonal = list(order=c(0,1,0), period=4))
ts.plot(sa3$residuals)
```



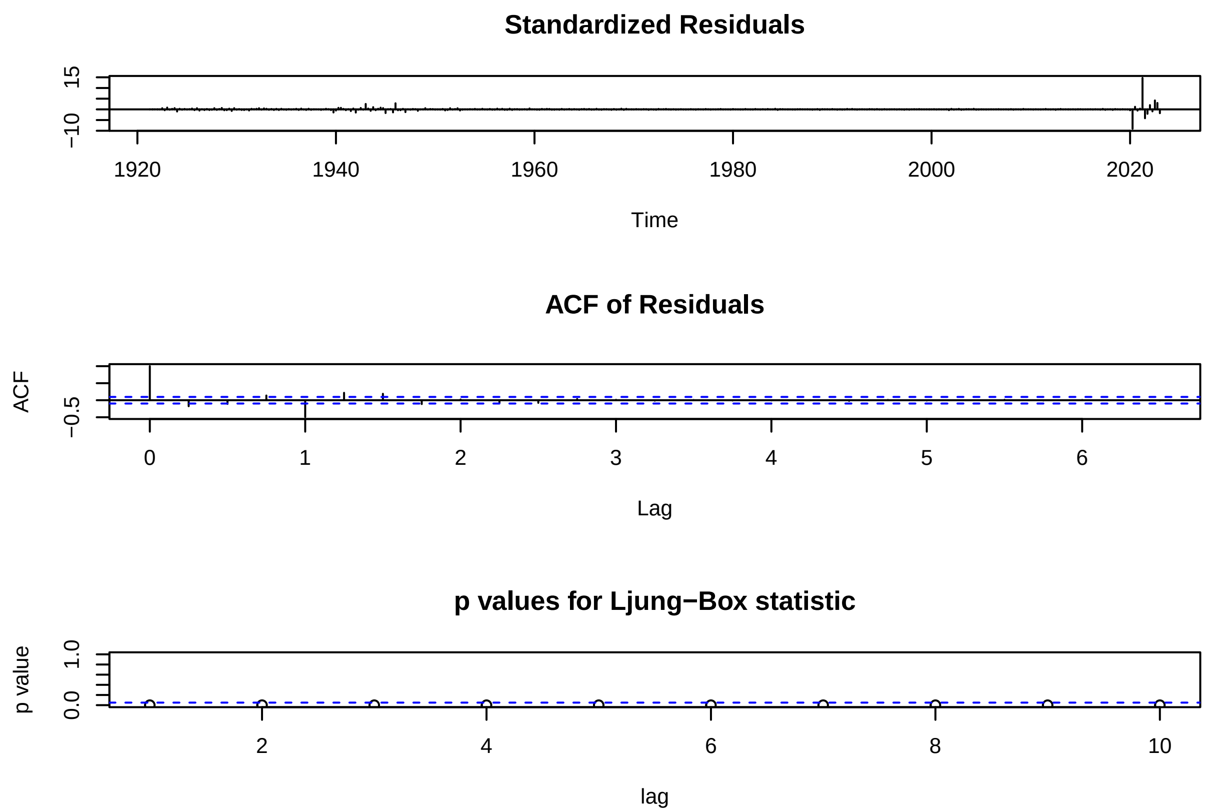

```
acf(sa3$residuals)
```




```
pacf(sa3$residuals)
```

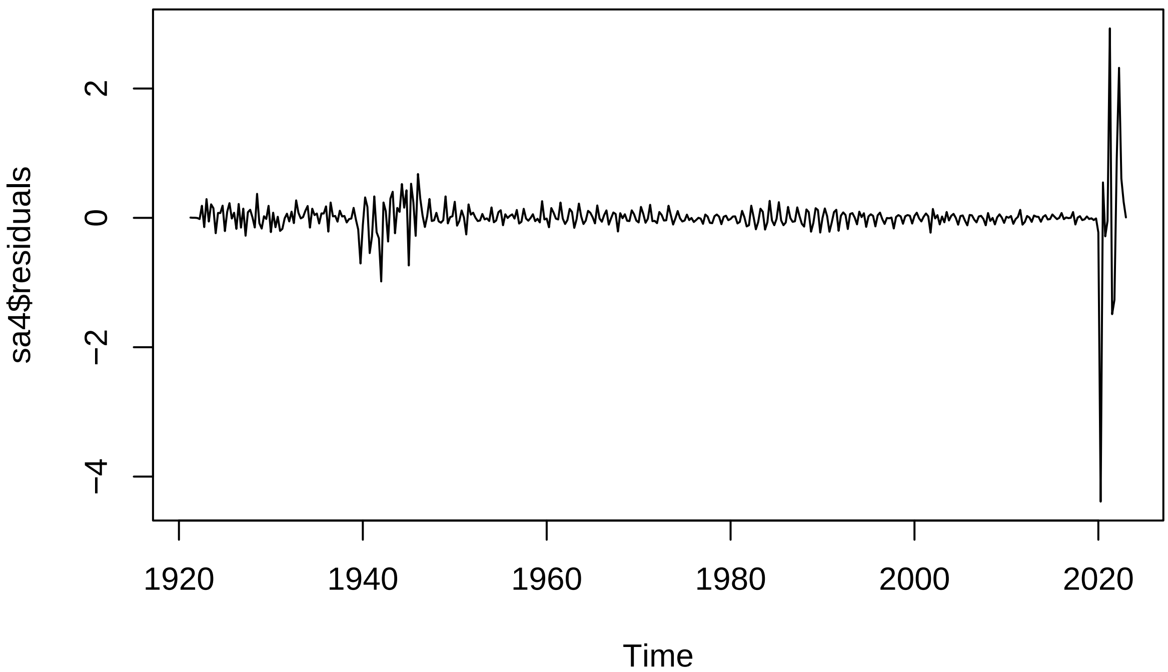



```
tsdiag(sa3)
```

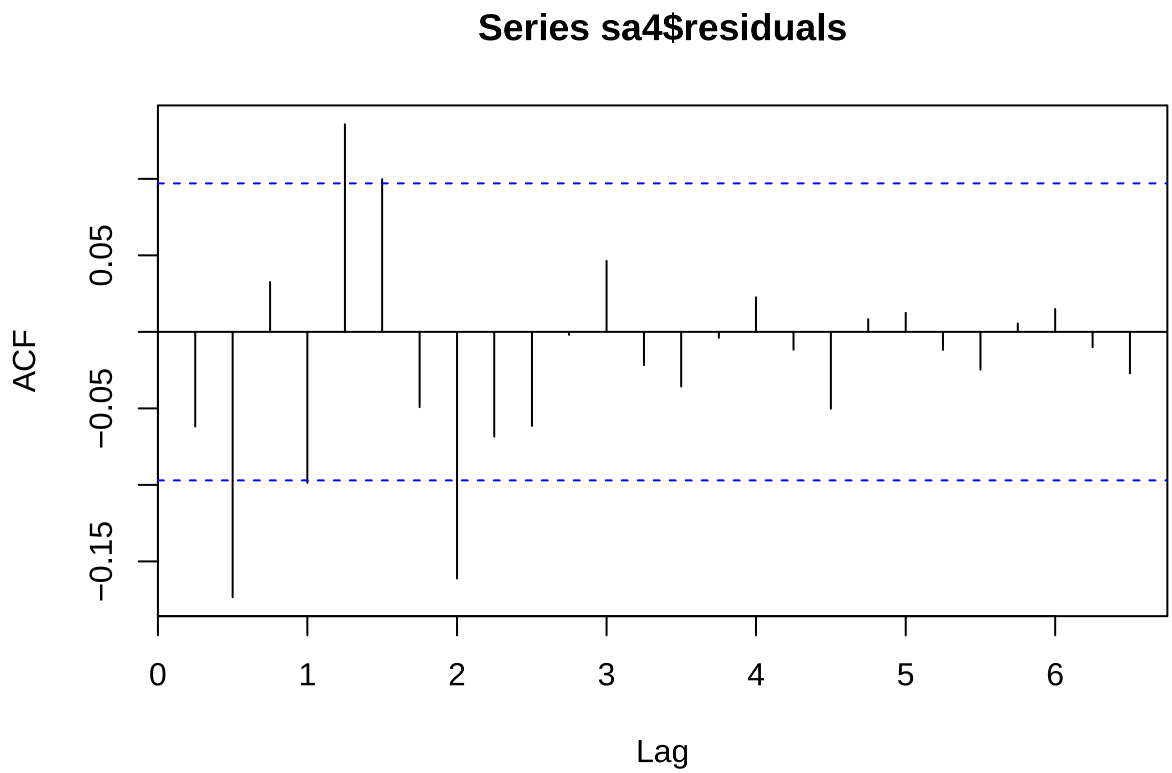


The PACF seems to be affected by a seasonal AR component

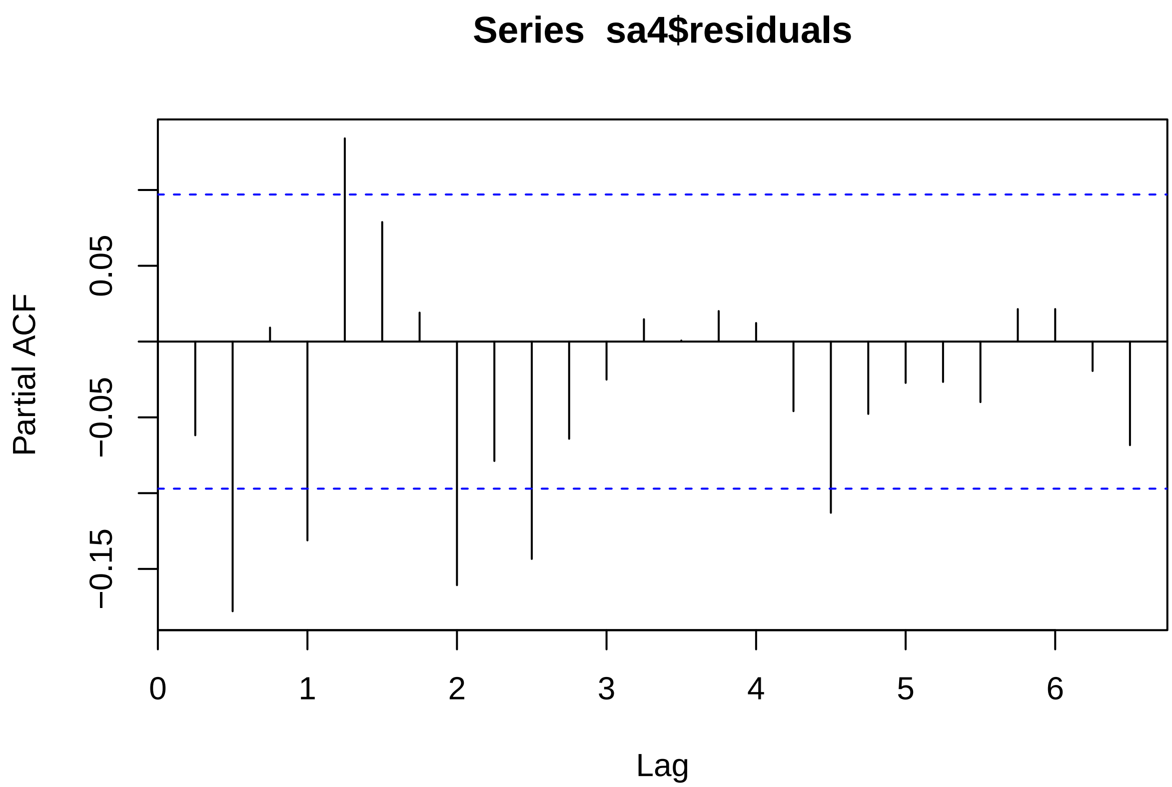
```
sa4 <- arima(logvis, order=c(0,1,0), seasonal = list(order=c(0,1,1), period=4))
ts.plot(sa4$residuals)
```



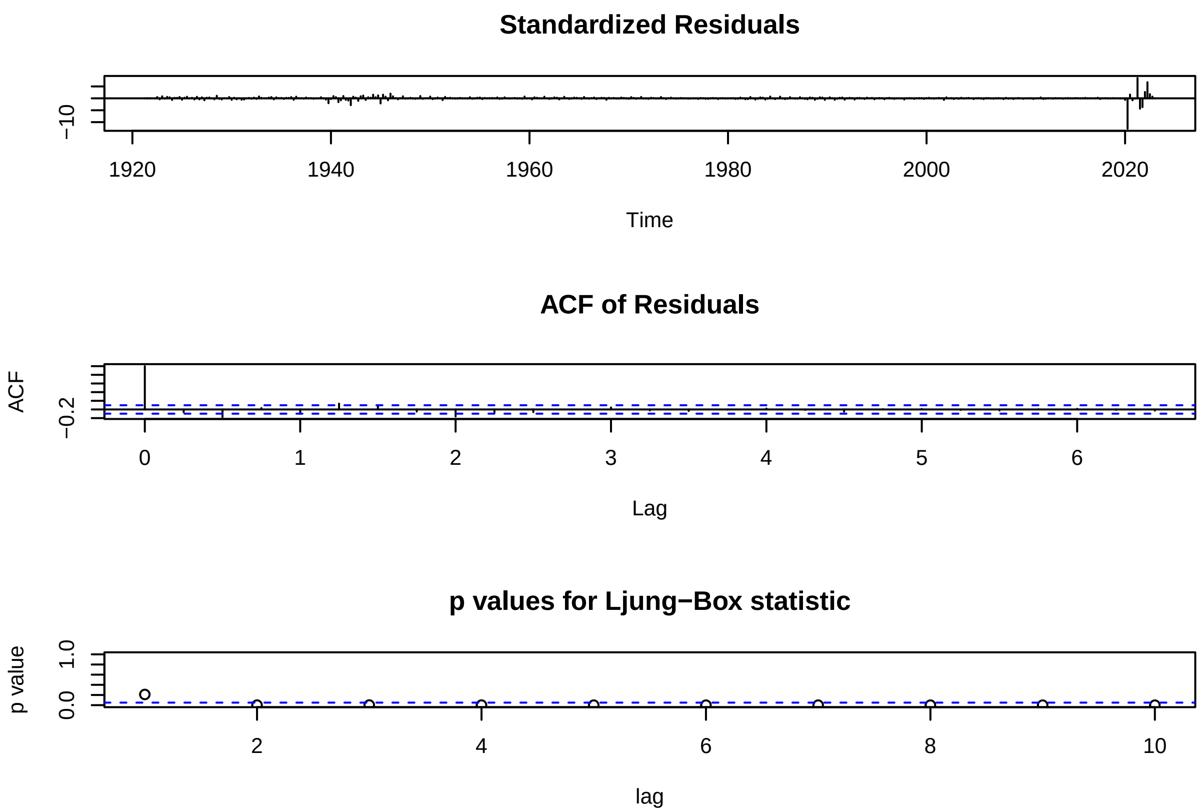

```
acf(sa4$residuals)
```




```
pacf(sa4$residuals)
```

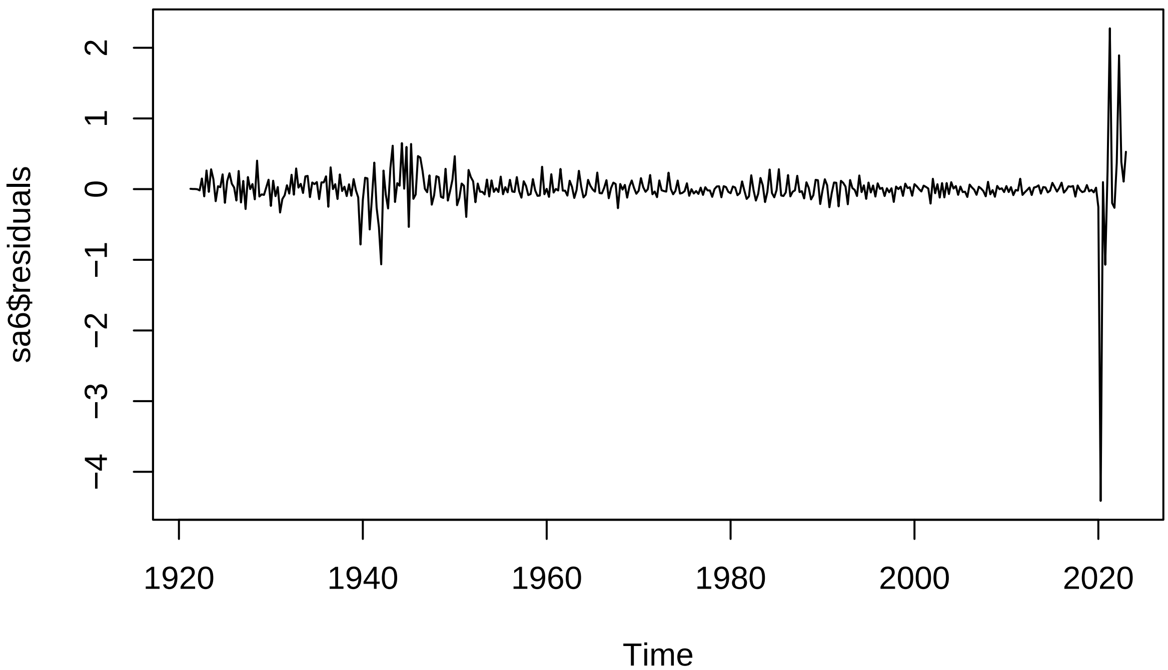



```
tsdiag(sa4)
```

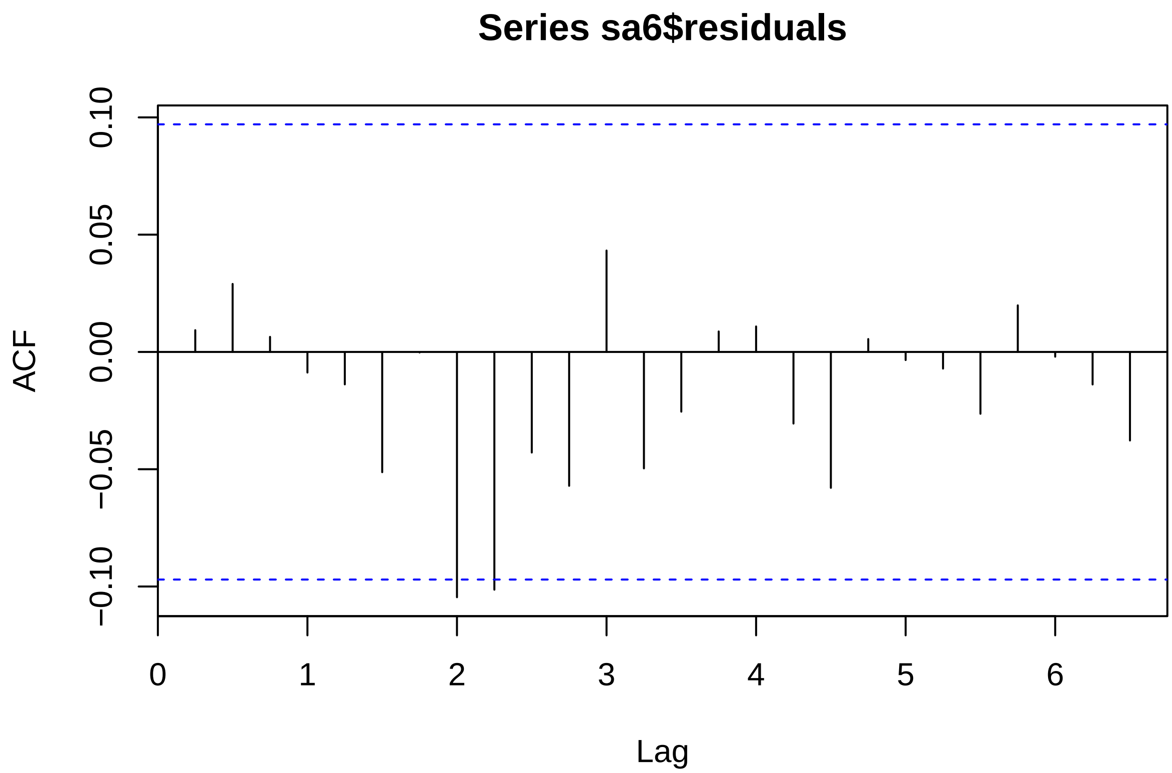


The ACF shows a MA(5) and the PACF shows an AR(5)

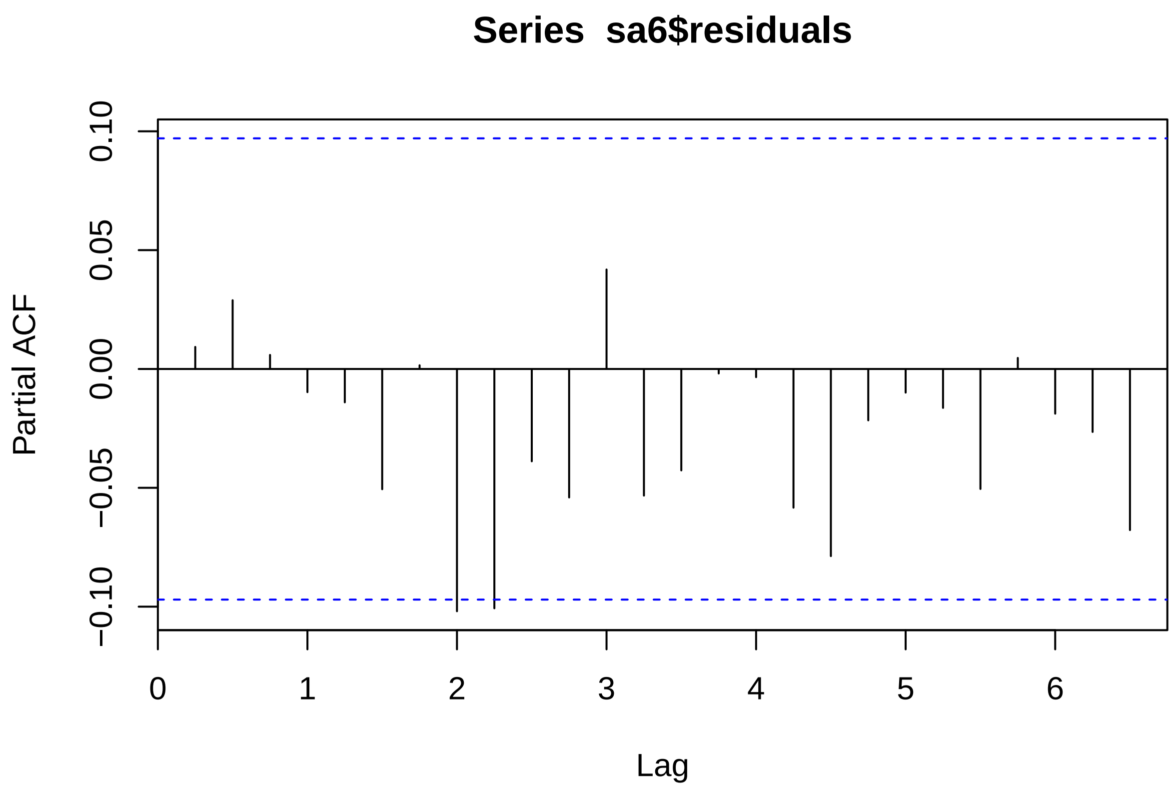
```
sa6 <- arima(logvis, order=c(5,1,5), seasonal = list(order=c(0,1,1), period=4))
ts.plot(sa6$residuals)
```



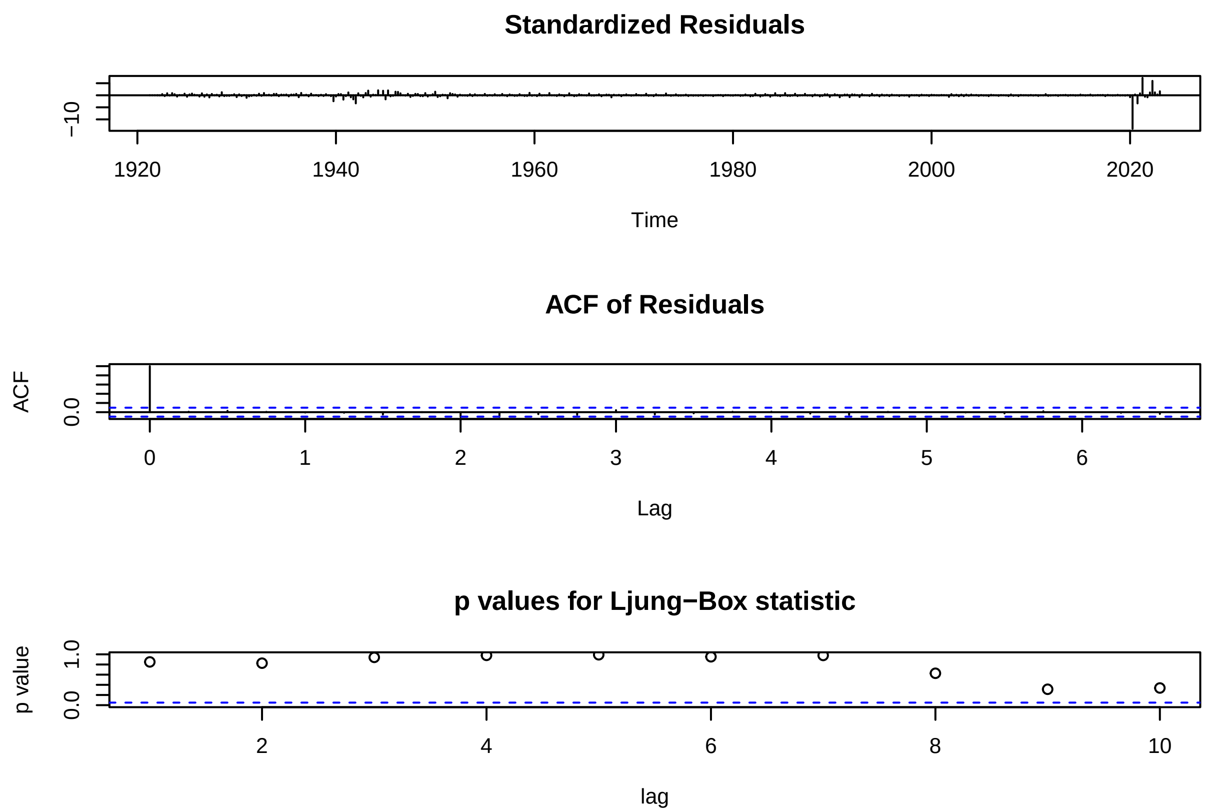

```
acf(sa6$residuals)
```




```
pacf(sa6$residuals)
```


Series sa6$residuals
Partial ACF
0.10
0.05
0.00
−0.05
−0.10
0
1
2
3
4
5
6
Lag

```
tsdiag(sa6)
```



The residuals appear to have constant variance in the standardized residuals plot and there are no significant spikes in the acf after the first there spikes also look like white noise. all p-values are significant in the Ljung-box statistic, this means our model fits well. All of this suggests the residuals are white noise

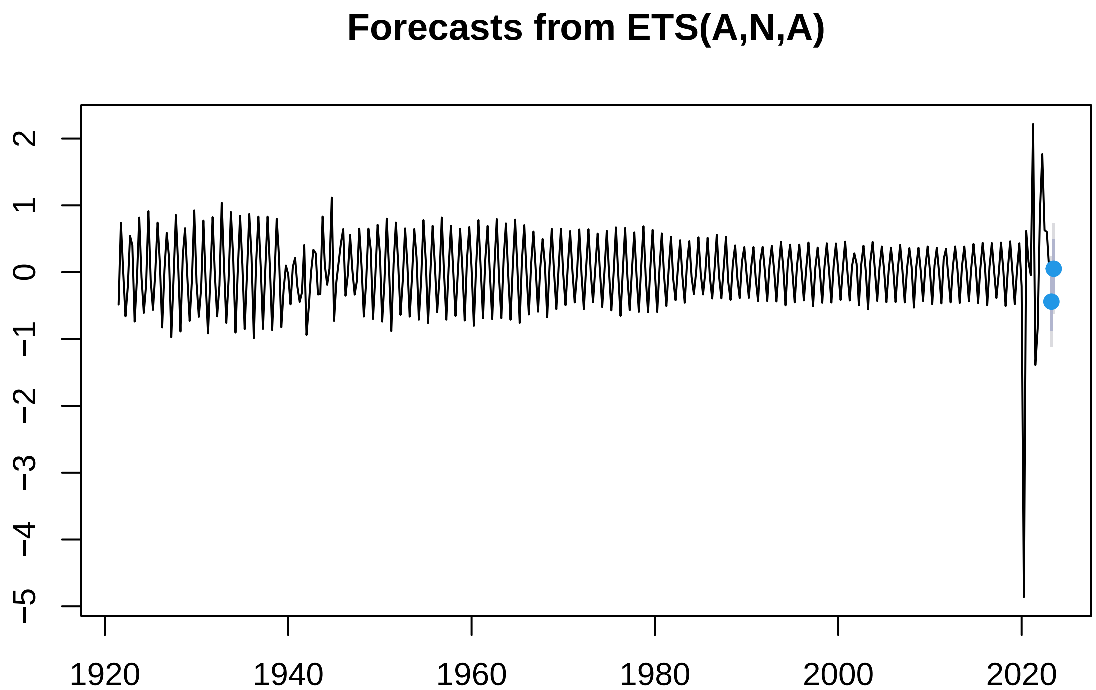
The equation for out model is $\phi(B)Y_t = \theta(B)\Theta(B^4)W_t$

where $\phi(B) = (1 + 0.278B + 0.881B^2 + 0.564B^3 + 0.564B^4 + 0.581B^5)$

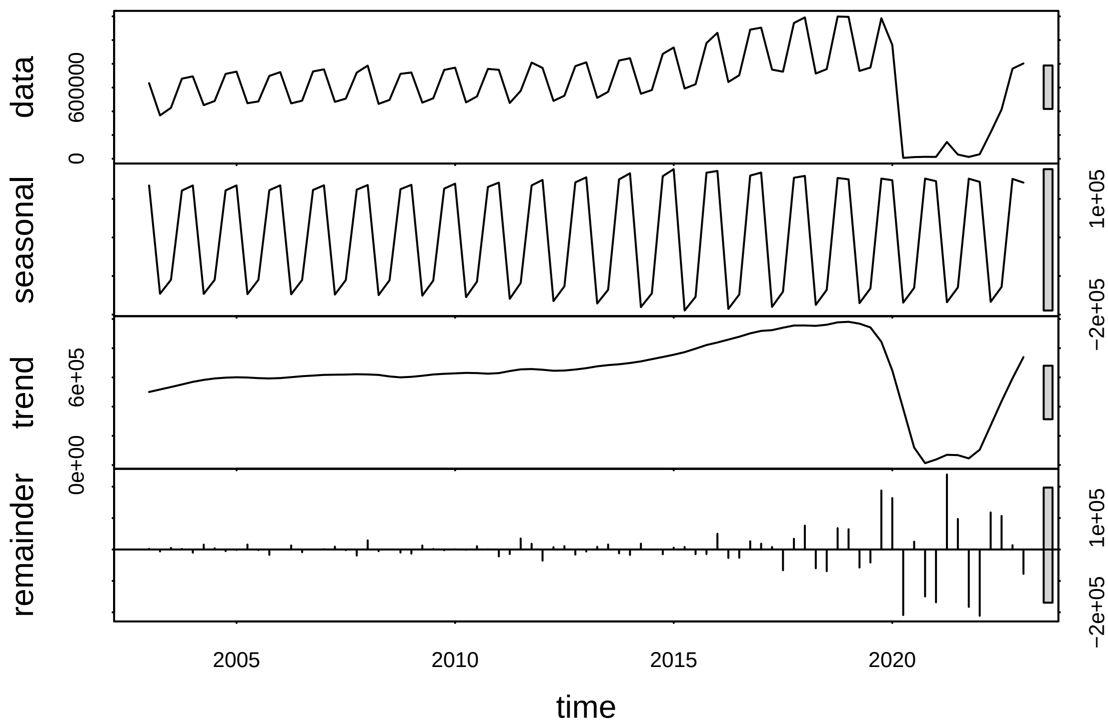
$\theta(B) = (1 + 0.186B + 0.663B^2 + 0.456B^3 + 0.242B^4 + 0.683B^5)$

$\Theta(B^4) = (1 - 0.926B^4)$

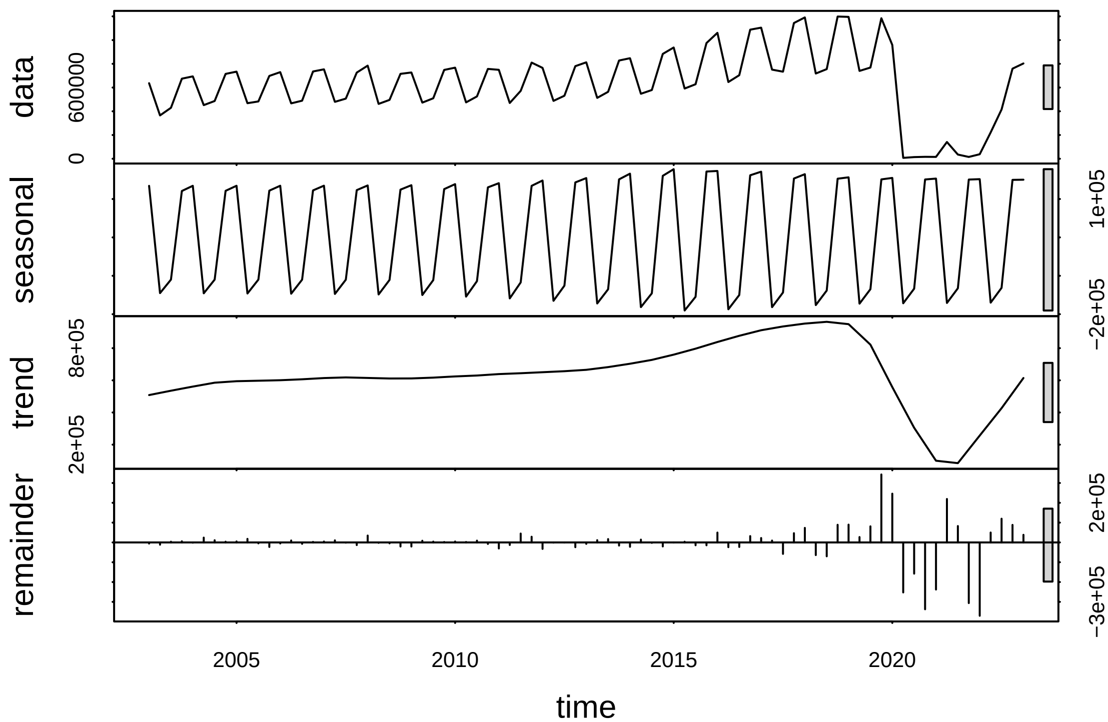
```
forecast_val <- forecast(logdvis, h=2)
plot(forecast_val)
```




```
l20<- tail(ts.visitor, 81)

stl_result <- stl(l20, s.window = 15)
plot(stl_result)
```




```
stl_result2 <- stl(l20, s.window = 15, t.window=12)
plot(stl_result2)
```



the seasonal pattern shows an increase in the first and fourth quarters of the year. This is because it is summer in the first and fourth quaters, school holidays and Christmas time so there are many tourists visiting New Zealand as there are a lot of warm weather activities in New Zealand. Many New Zealanders living abroad are also visiting family around Christmas time which falls in the fourth quarter.

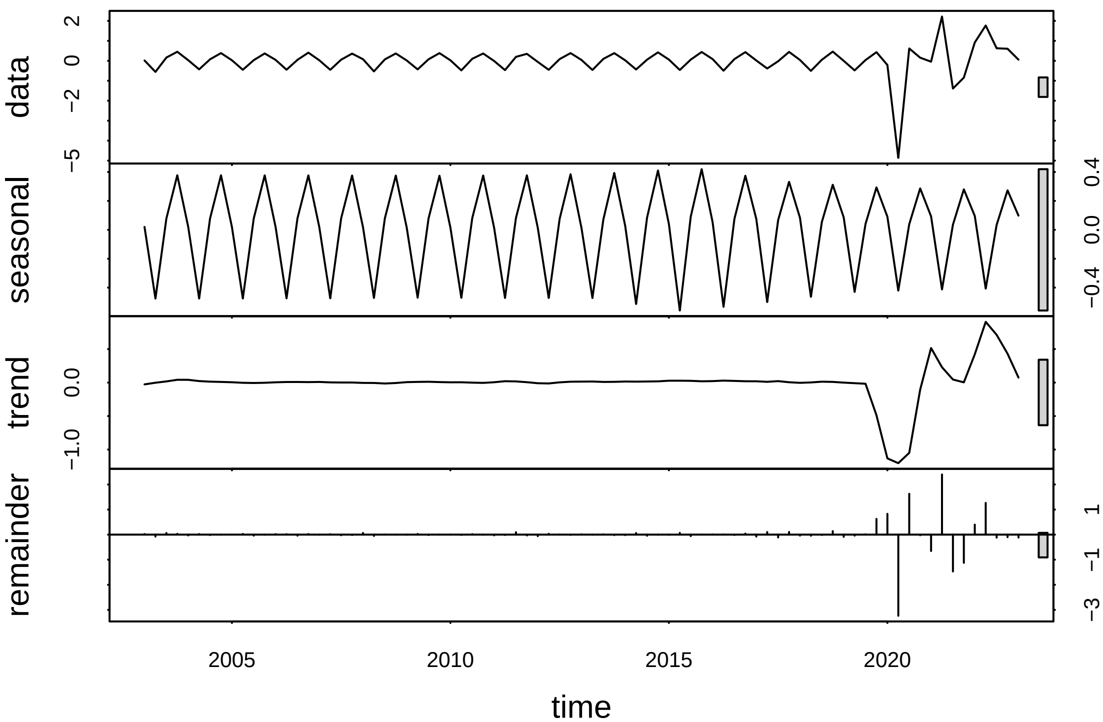
The trend is slowly increases until 2020 Q1 when covid happend causing a trough in the trend. This is because New Zealand was in lock down and visitors were not allowed into New Zealand. There is a sharp increase starting at 2022 Q2 when the boarders were reopened.

The remainder shows the residuals before Q4 2019 are white noise, meaning the values are closely following the trend and seasonal pattern. Q1 2020 when covid occured. There is a sharp drop from when lock down was announced in the remainder as the actual visitor arrivals were lower than expected. There is a slight peak in remainder when the boarders were briefly opened in 2021 followed by another sharp decrease due to the second lock down.

to calculate the default window we do 1.5 * the period (4) and then add 1 1.5 * 4 + 1 = 7

We want the t.window to be odd as the trend is smoothed based on the t.window value. We want the value to be odd as it will center on the point we are interested in and take the average of the points eg when t.window = 5 we take the average of the previous 2 and next 2 points. If t.window was even we wont take the same amount of previous values as future values. When we set t.window to 12 it is even. The increased window is smoother as it is calculating the trend using the average of more data points so it is influenced less by extreme points.

```r
logl20 <- tail(logdvis, 81)

stl_result2 <- stl(logl20, s.window = 15)
plot(stl_result2)
```



The stl of the logged appears to be slightly different. The trend appears relatively flat in the logged series whereas is was slightly increasing in the original. The trough in the trend caused by COVID did not last as long in the logged series but the increase once the boarders were opened is more sporadic in the logged series. The seasonal component appears to have a sharper peak at Q1 when compared to the original series. The remainder appears to be slightly more stable in the logged series and COVID appears to have less of an effect on the remainder of the logged series.

The default value for t.window will still be 7 as the period has not changed and therefore will remain 7

After looking at both decomposition I would remove the values around COVID time which are 2020 Q2 - 2021 Q2 and 2021 Q3 to 2022 Q2. This is cause the consequences from COVID were unpredictable and we do not see pandemic outbreaks to the magnitude of COVID often so we dont want our predictions to be influenced by COVID.