

Time Series Analysis

Midterm project

Helen Kafka

Part 1

1.

I am expecting a somewhat clear upward trend due to economic growth causing people to have more money to spend, inflation causing prices to rise, and population growth increasing demand for retail products. Also, there could be a seasonal trend (e.g. people spend more money during Christmas season), such that the retail sales are higher on special holidays. Because the Brexit was later than 2007, I am not expecting a big sink somewhere in the graph.

```
##
```

```
## Attache Paket: 'TSA'
```

```
## Die folgenden Objekte sind maskiert von 'package:stats':
```

```
##
```

```
##      acf, arima
```

```
## Das folgende Objekt ist maskiert 'package:utils':
```

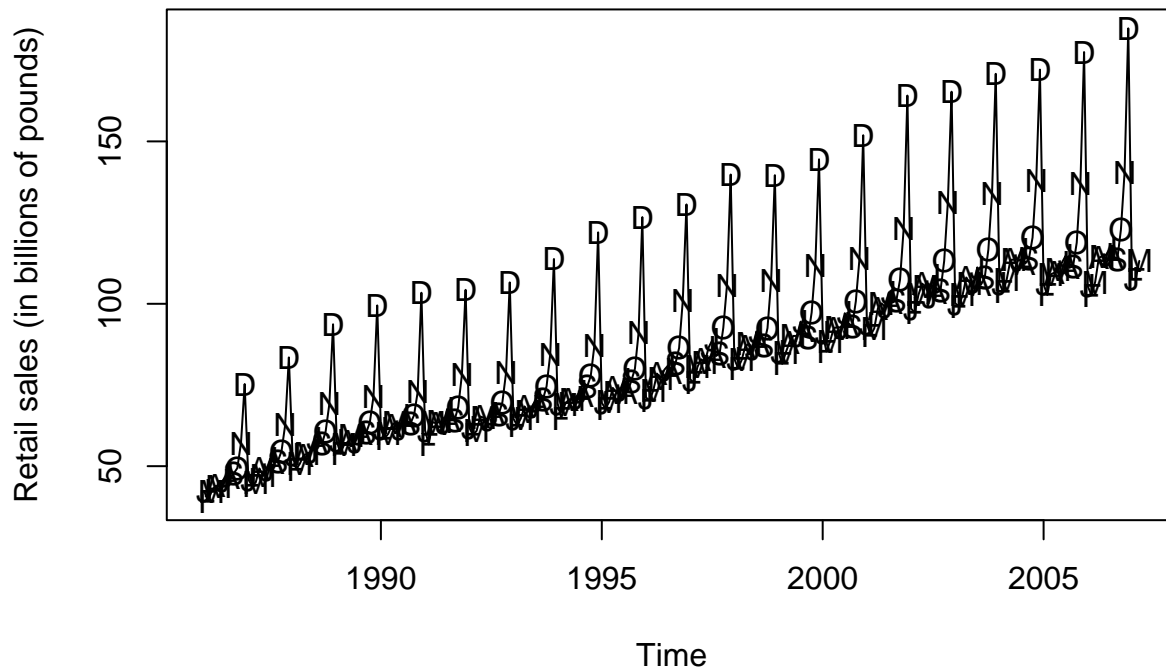
```
##
```

```
##      tar
```

```
data(retail)
```

```
plot(retail, xlab = expression("Time"), ylab = expression("Retail sales (in billions of pounds)"), main =  
points(retail, pch = as.vector(season(retail)))
```

Time Series Plot of UK Retail sales 1987–2007



When seeing the plot, I can confirm the upward trend. It is also obvious that there is a seasonal trend, that is described by a peak while Christmas season. This means that the sales are going up starting November each year and reach the peak in December before dropping again.

2.

I would try a seasonal ARIMA model, given the upward trend and the potential seasonality.

```
model <- forecast::auto.arima(retail, seasonal = TRUE)
```

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

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

summary(model)
```

```
## Series: retail
## ARIMA(1,0,4)(0,1,1)[12] with drift
##
## Coefficients:
##      ar1      ma1      ma2      ma3      ma4      sma1      drift
##      0.8469 -0.5939  0.0504  0.0878  0.1742 -0.0966  0.3083
## s.e.  0.0533  0.0788  0.0740  0.0772  0.0756  0.0631  0.0405
##
## sigma^2 = 3.413: log likelihood = -491.04
## AIC=998.07  AICc=998.69  BIC=1026.02
##
```

```
## Training set error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.0003999953 1.777353 1.287341 -0.1007814 1.485438 0.3327206
##           ACF1
## Training set 0.003399519
```

Model type and coefficients

Looking at the summary, we are given an ARIMA model with $p = 1$, $d = 0$, $q = 4$, where q is the autoregressive order with the coefficient $ar_1 = 0.8469$, also known as the lag order. d is the number of differences we take, which is 0. The ARIMA model does not take a difference, hence, it indicates that the data is already stationary. And q is the order of the moving average process, where the coefficients are $ma_1 = -0.5939$, $ma_2 = 0.0504$, $ma_3 = 0.0878$, $ma_4 = 0.1742$. All in all, we have a first order degree autoregressive model with zero order of differencing and a fourth order moving average model. That was for the non-seasonal part. For the seasonal part, we have $p = 0$, $d = 1$, $q = 1$, where p , d and q are exactly as above. “[12]” indicates the number of seasons. Here we have 12 months per season.

Fit

There are some values given to rate the model fit, which are

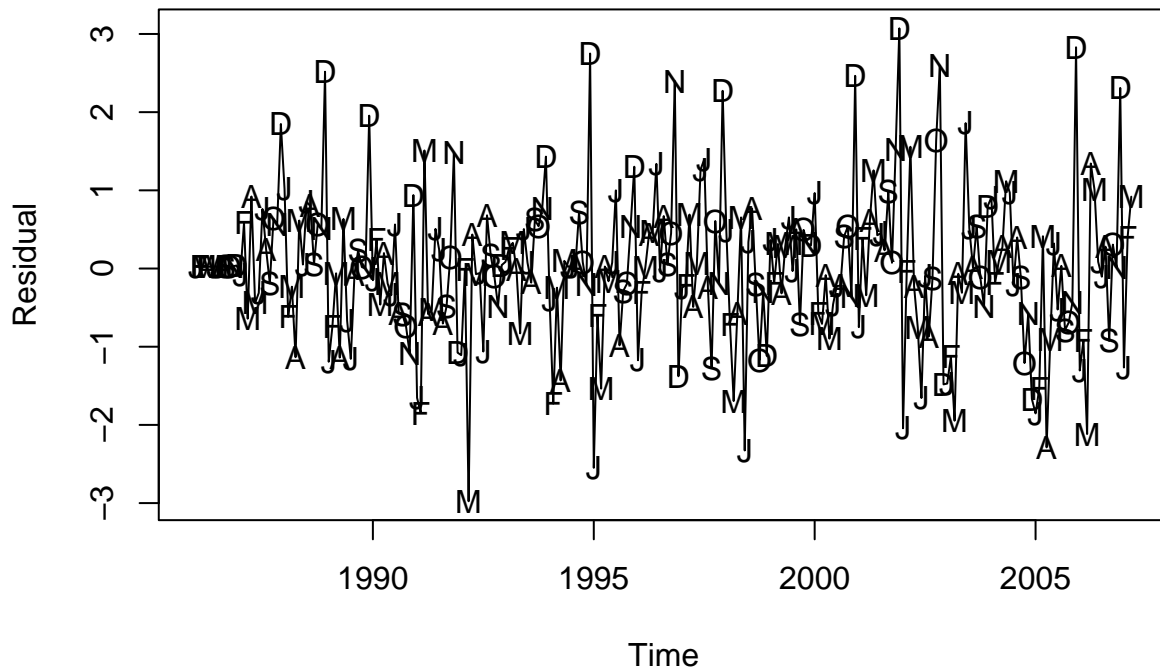
$$\begin{aligned}\sigma^2 &= 3.413, \\ \log \text{likelihood} &= -491.04, \\ AIC &= 998.07, \\ AICc &= 998.69, \\ BIC &= 1026.02.\end{aligned}$$

Here, we can just look at AIC and BIC which suggest a reasonable choice based on model complexity vs fit. The other values are for forecasting reasons.

3.

```
res <- ts(rstandard(model), start = c(1986, 1), end = c(2007, 3), frequency = 12)
plot(res, xlab = "Time", ylab = "Residual", main = "Time Series Plot of Residuals")
points(y = res, x = time(res), pch = as.vector(season(res)))
```

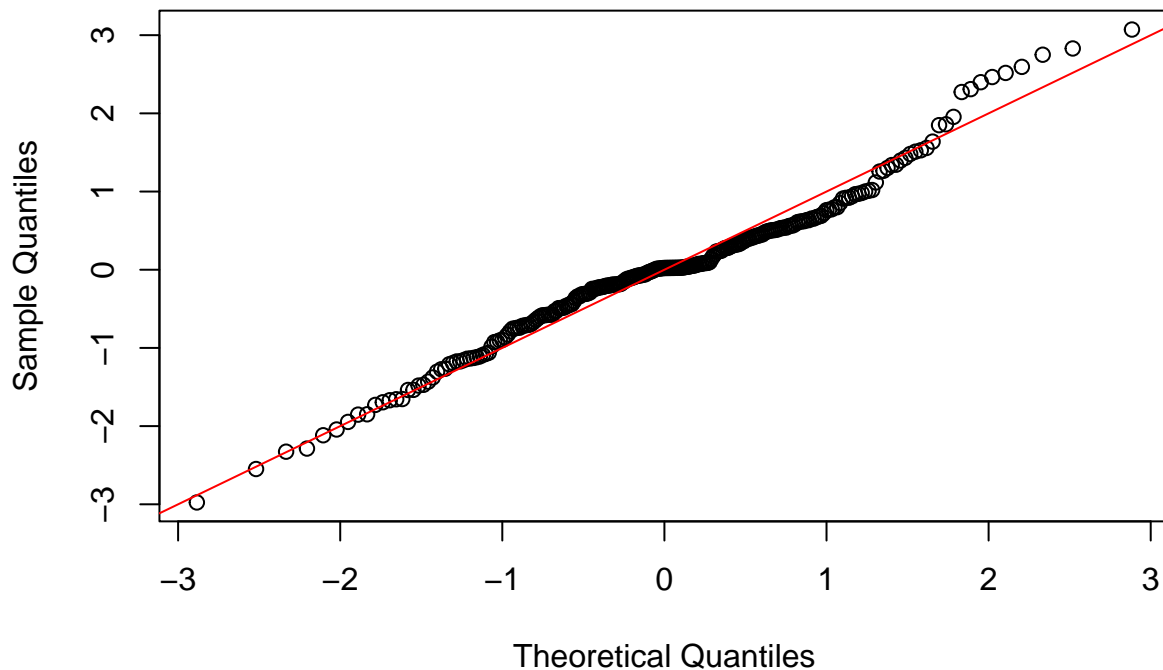
Time Series Plot of Residuals



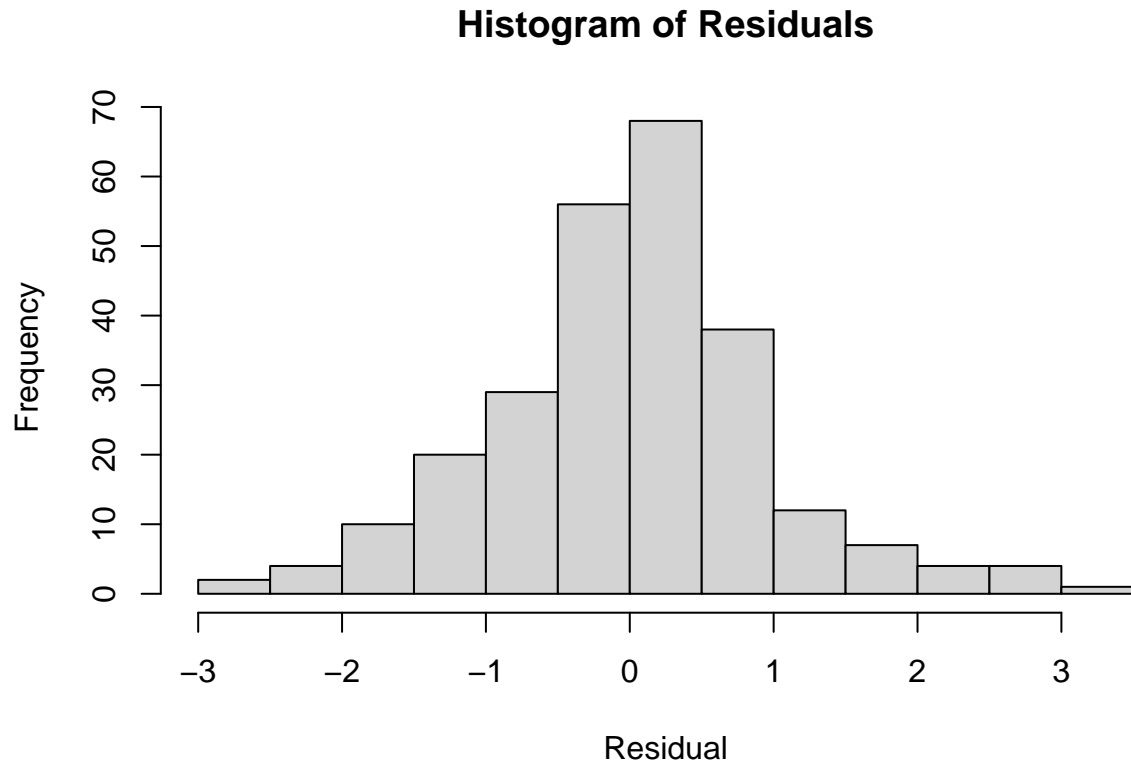
The plotting signs “D” and “N” are often displayed as peaks. Hence, the plot, given the plotting signs, somewhat reflects the seasonal trend.

```
qqnorm(res, main = "Normal Probability Plot vs \n Residuals Fit")
abline(a = 0, b = 1, col = "red")
```

Normal Probability Plot vs Residuals Fit



```
hist(res, xlab = "Residual", main = "Histogram of Residuals")
```



The QQ-Plot and the histogram indicate that we have somewhat standard-normal distributed residuals, which indicates that we have a good model fit.

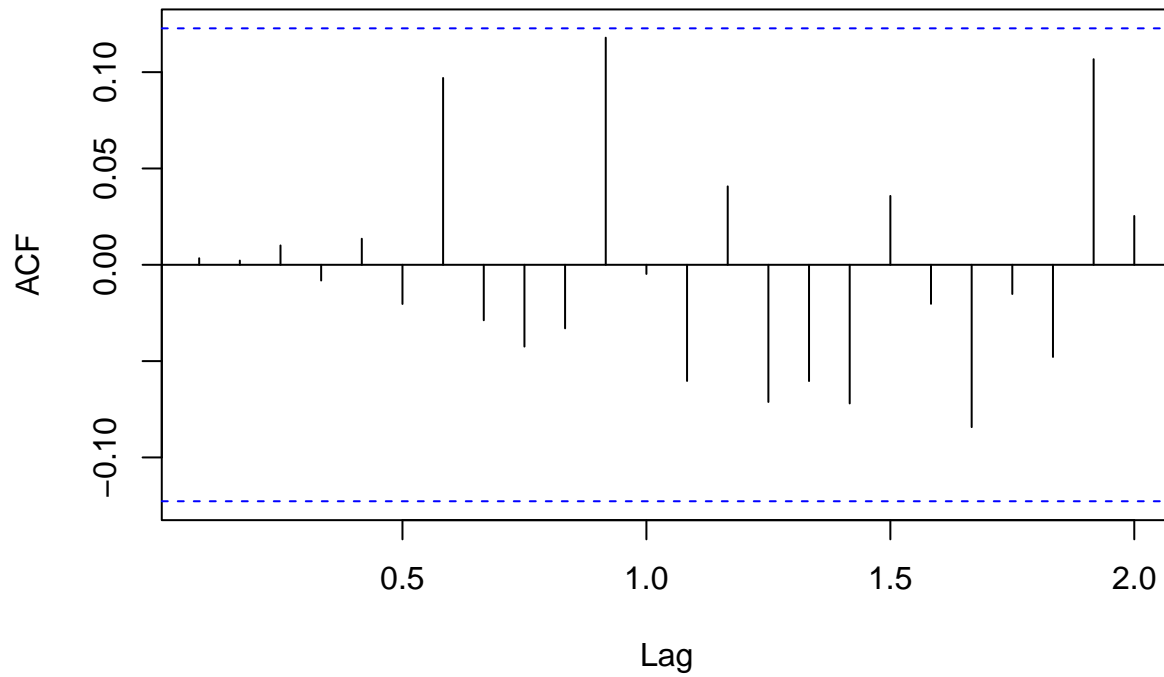
```
shapiro.test(res)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  res  
## W = 0.97597, p-value = 0.0002613
```

The p-value of the Shapiro-Wilk test is pretty low, especially much lower than 0.05, which also indicates that the residuals are normally distributed.

```
acf(res, main = "Autocorrelation Plot of Residuals")
```

Autocorrelation Plot of Residuals

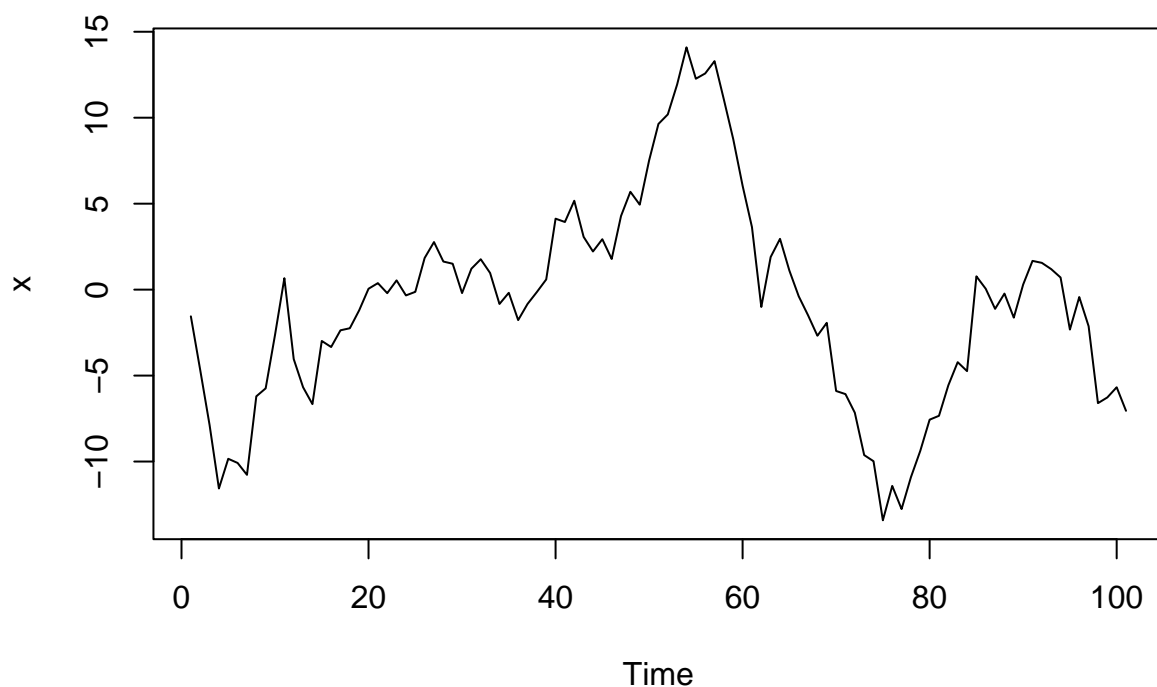


The autocorrelation plot shows no significant lags. Especially, the lags are in absolute value smaller than 0.1. There are positive and negative lags, such as movement around 0 and no seasonality to see.

Part 2

```
series <- as.ts(read.delim(file = "./MidtermPt2.txt", header = TRUE, sep = " "))  
plot(series, main = "Original Series")
```

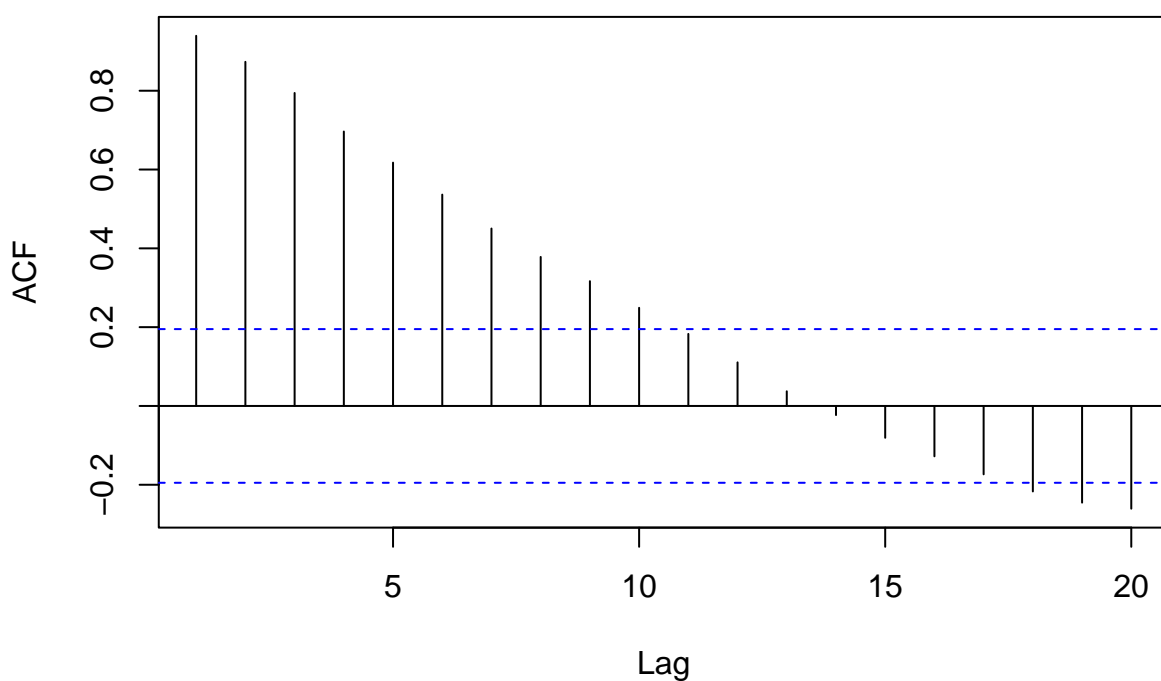
Original Series



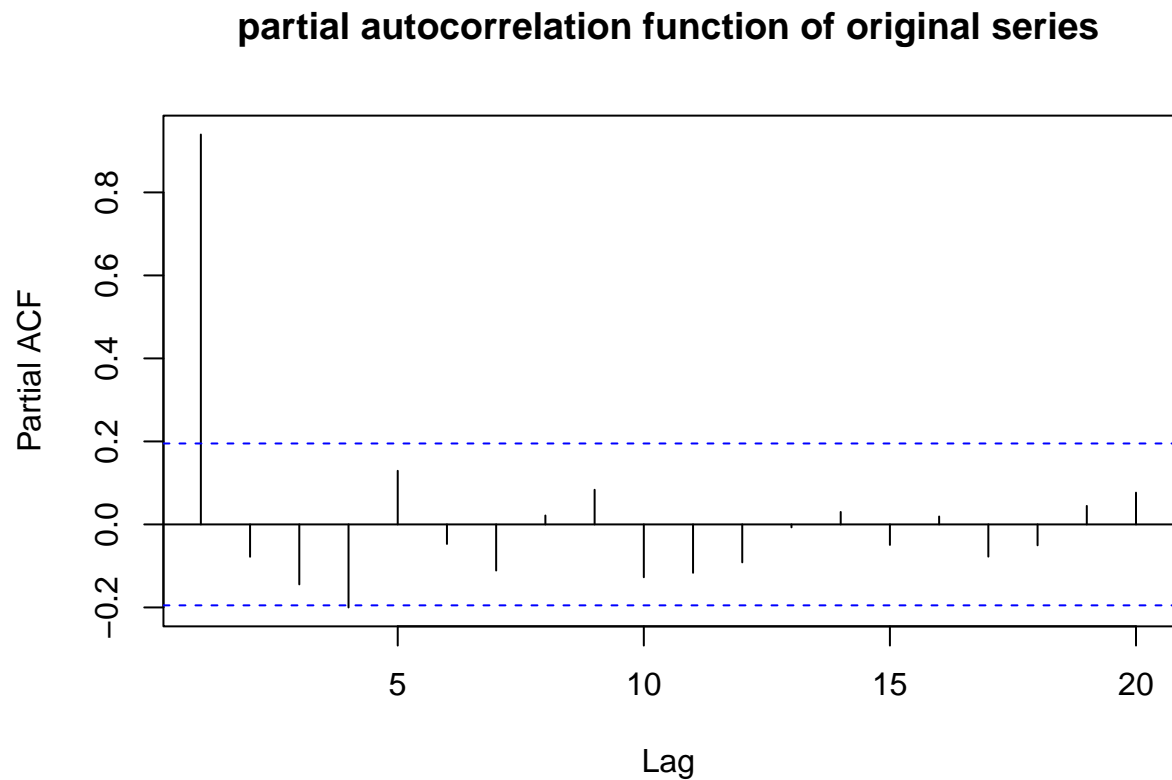
It is hard to determine from this graph whether the series is stationary or not. It seems to move around 0 over time but the variance also seems to increase over time. Hence, we will have a look at the ACF and PACF of the series and hopefully these give us a better insight.

```
acf <- acf(series, main = "autocorrelation function of original series")
```

autocorrelation function of original series



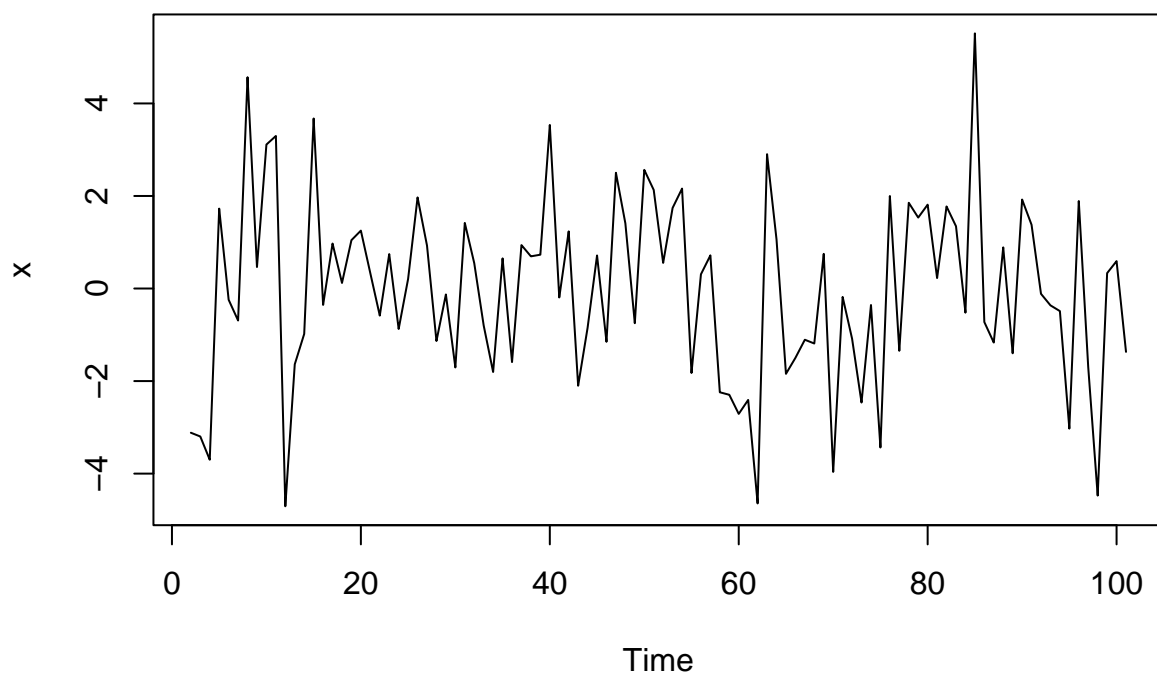
```
pacf <- pacf(series, main = "partial autocorrelation function of original series")
```



We can see that the ACF decreases over time and, hence, it depends on time. Thus, the series is not stationary. We consider first differencing to transform the series such that it is stationary.

```
d_series <- diff(series)
plot(d_series, main = "Differenced Series")
```

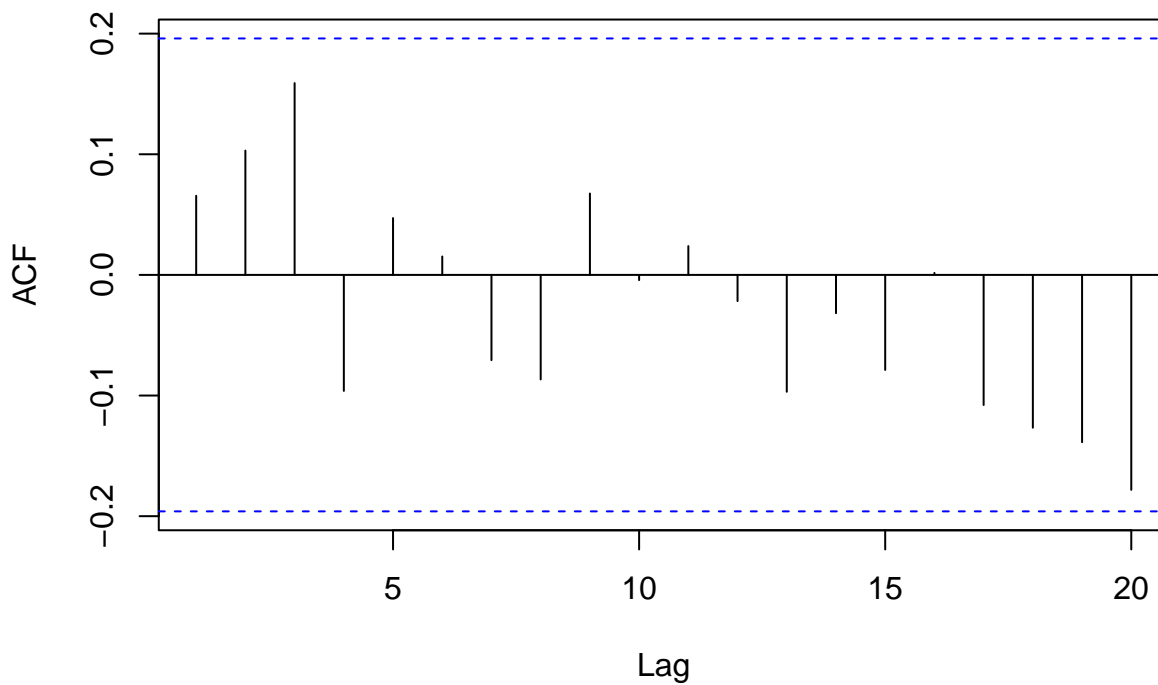

Differenced Series



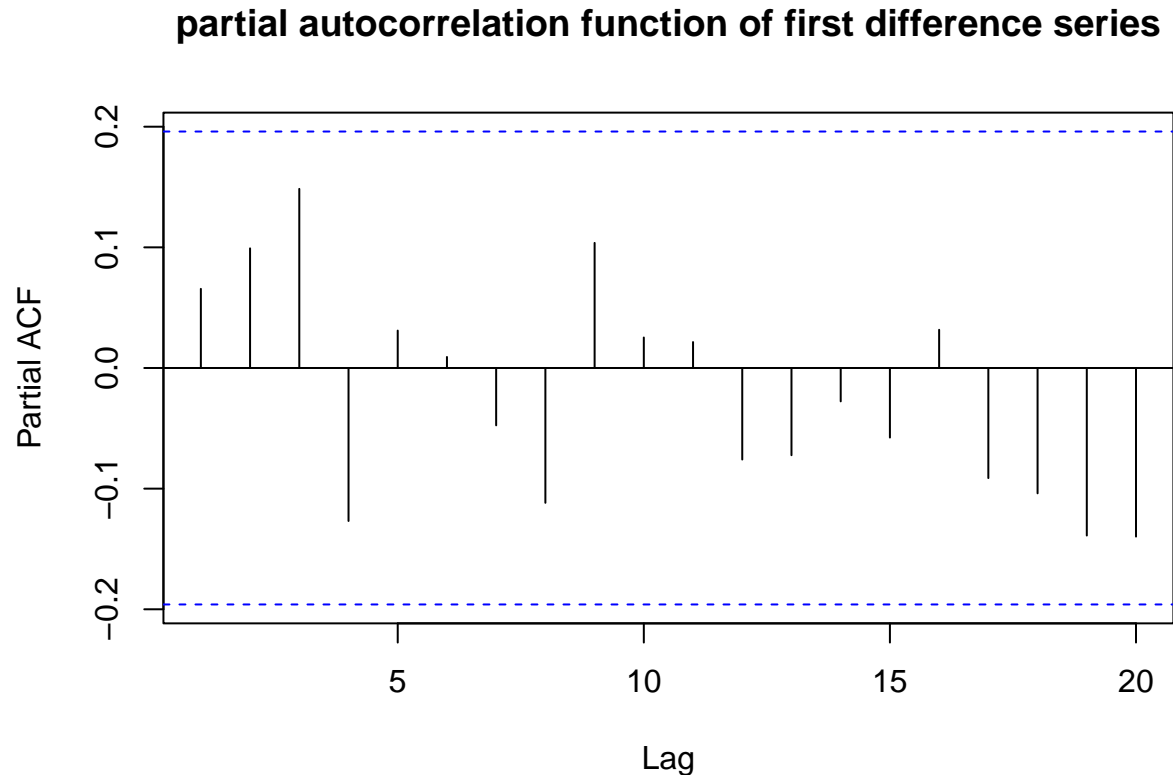
This plot seems constant around 0 over time and it is also evenly spread over time. Hence, it looks stationary. We will check our assumption with the ACF and PACF plots:

```
acf <- acf(d_series, main = "autocorrelation function of first difference series")
```

autocorrelation function of first difference series



```
pacf <- pacf(d_series, main = "partial autocorrelation function of first difference series")
```



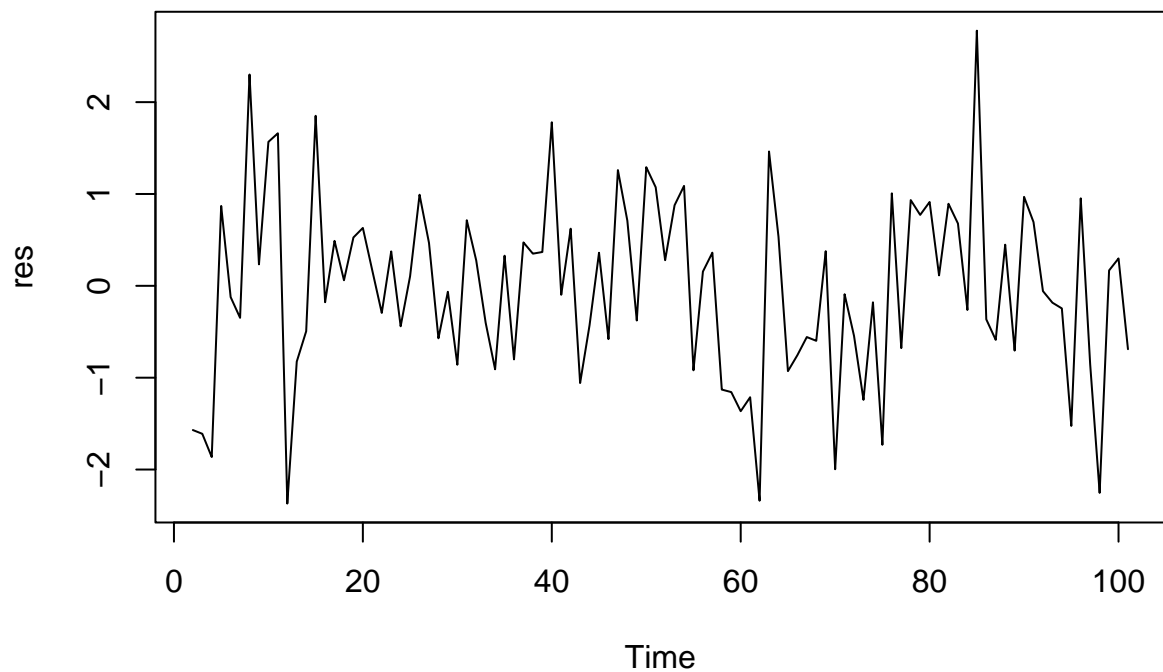
Both plots have a small magnitude (in absolute value of 0.2) and it seems that they do not change over time that much. There is no seasonal/periodic pattern to see. That indicates that the model of first differences is stationary.

```
modelseries <- forecast::auto.arima(d_series, seasonal = TRUE)
summary(modelseries)
```

```
## Series: d_series
## ARIMA(0,0,0) with zero mean
##
## sigma^2 = 3.94: log likelihood = -210.45
## AIC=422.91 AICc=422.95 BIC=425.51
##
## Training set error measures:
##           ME      RMSE      MAE MPE MAPE      MASE      ACF1
## Training set -0.05495792 1.984958 1.589698 100 100 0.7621968 0.06560549
```

An ARIMA model with $d = p = q = 0$ indicates that the transformed model we got by taking first differences is stationary.

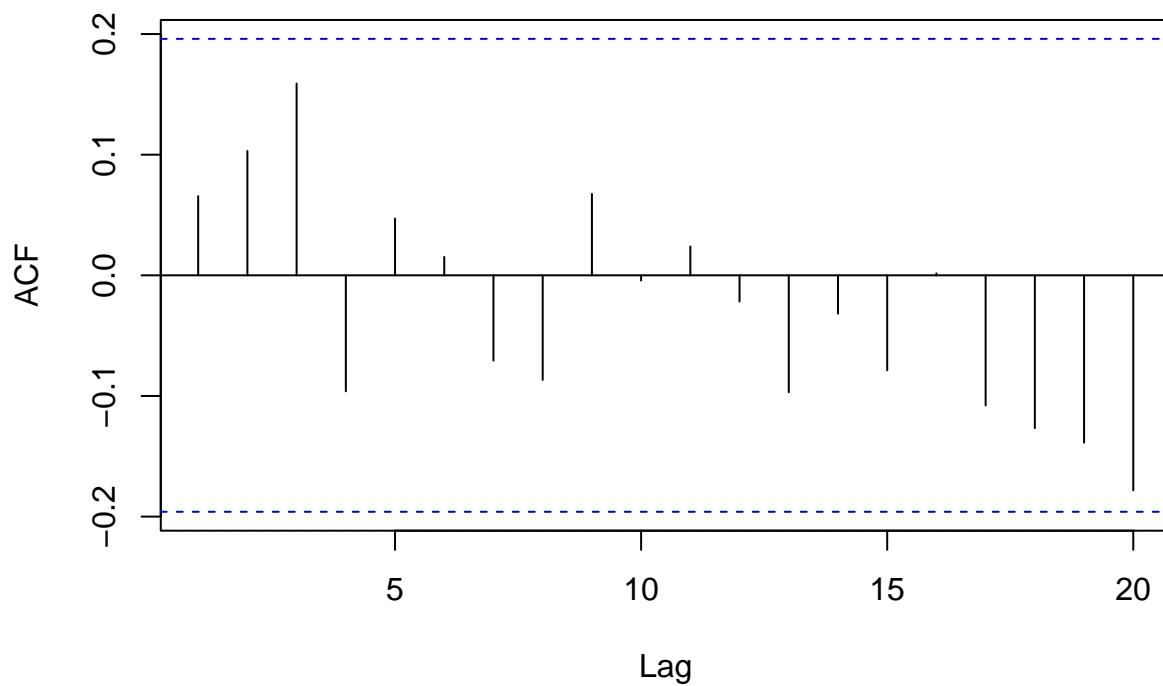
```
res <- as.ts(rstandard(modelseries))
plot(res)
```



The residuals of the transformed series stay around 0 and the variance is evenly spread.

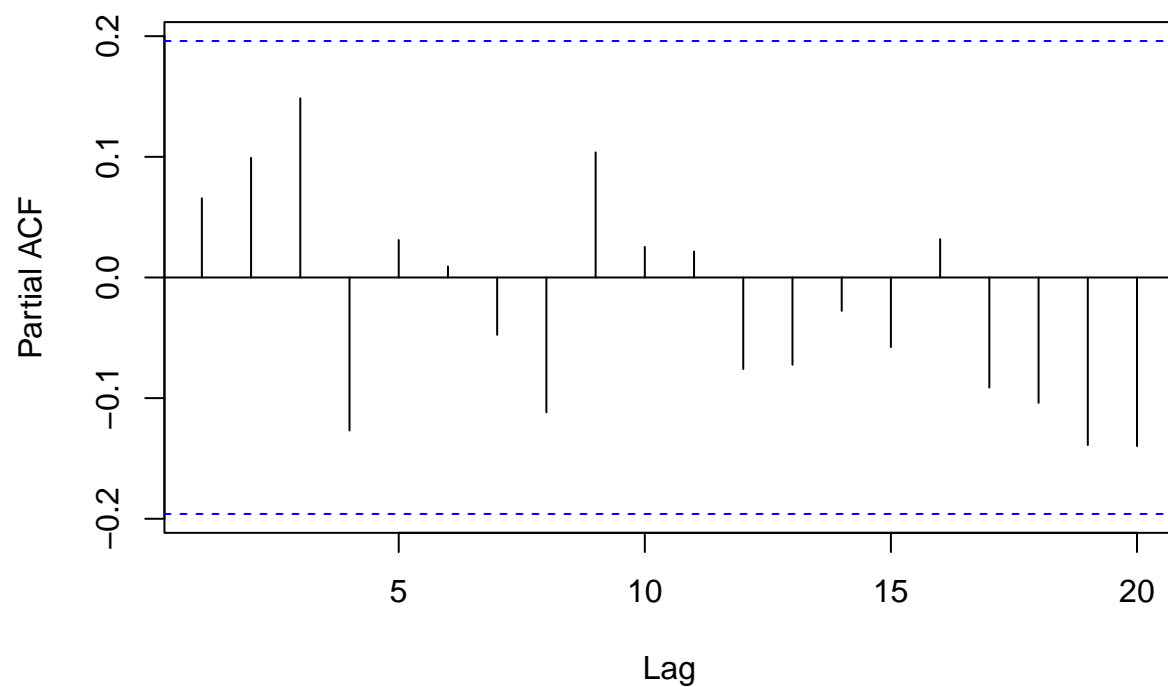
```
acf_res <- acf(res, main = "autocorrelation function of first difference series")
```

autocorrelation function of first difference series



```
pacf_res <- pacf(res, main = "autocorrelation function of first difference series")
```

autocorrelation function of first difference series



Looking at the ACF and PACf plot, we can see that the magnitude is again pretty small (smaller than 0.2 in absolute value) and there is no seasonal/periodic pattern to see.