SARIMA Modeling for the S&P 500

Pstat 274 Final Project

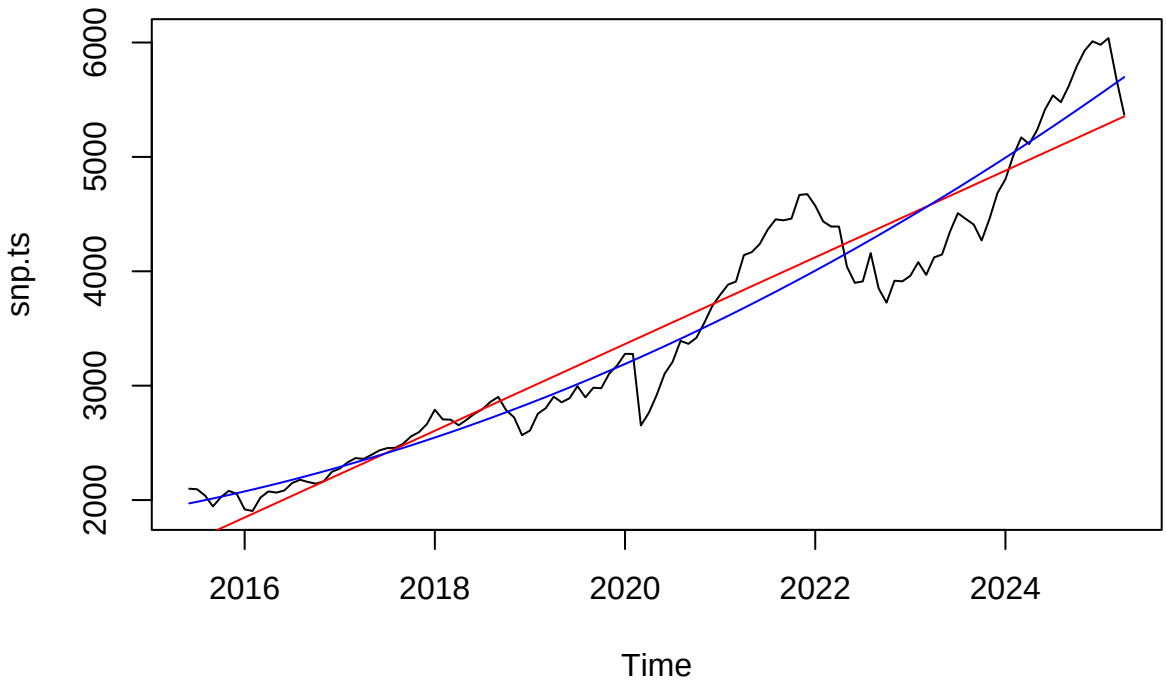
Jake Merry

06.06.2025

Setup Chuck:

```
knitr::opts_chunk$set(echo = TRUE)
library(MASS)
library(forecast)
```

```
snp <- read.csv("data/snpMonthly.csv")
snp.ts <- ts(snp[, 2], start=c(2015, 6), end=c(2025, 4), frequency=12)

index <- 1:length(snp.ts)

plot(snp.ts, type='l')
lines(tslm(snp.ts ~ trend)$fitted, col="red") # Linear trend
lines(tslm(snp.ts ~ trend + I(trend^2))$fitted, col="blue") # Quadratic trend
```

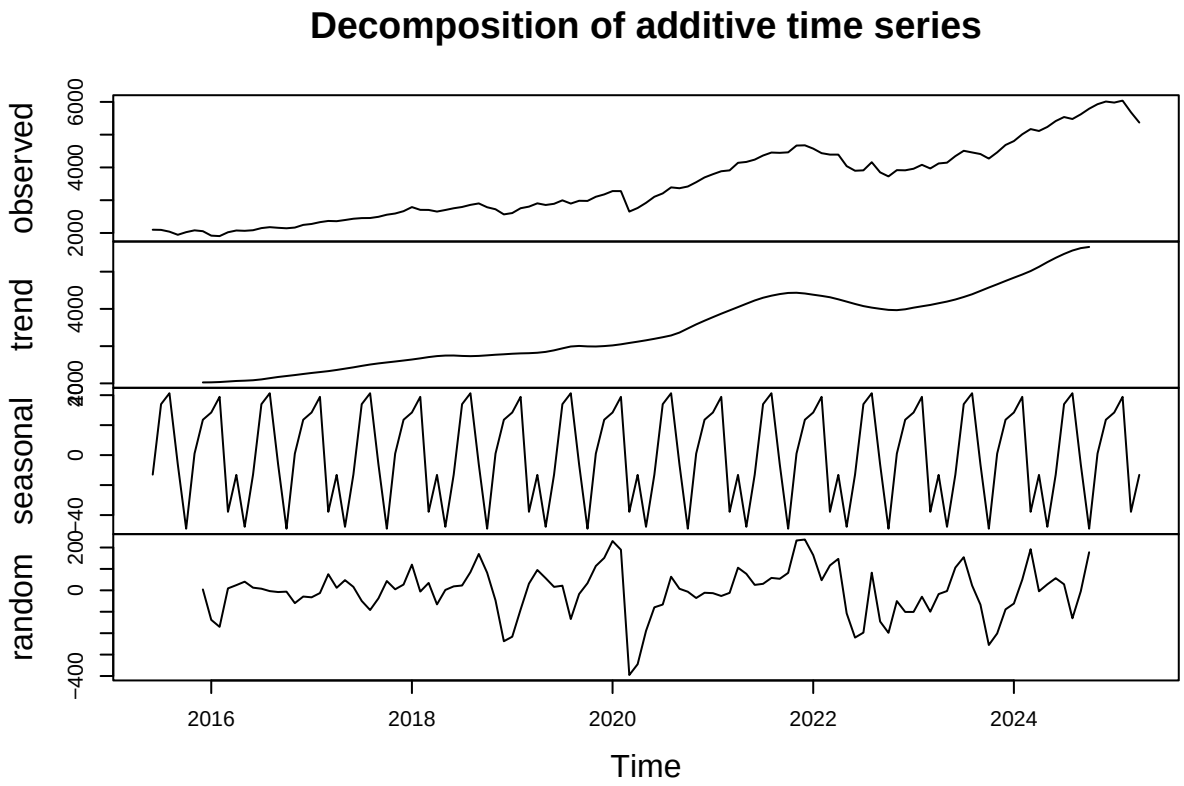



```
# Needs transformation to remove heteroskedasticity
# Difference to remove seasonality and trend
```

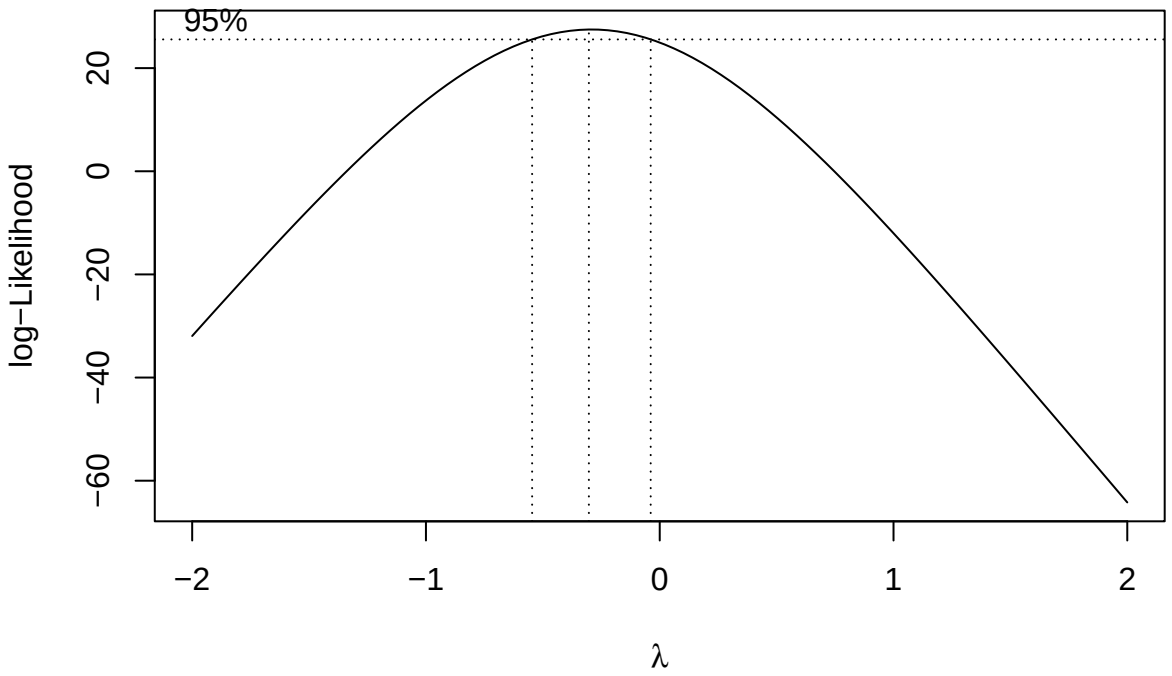
```
var(snp.ts)
```

```
## [1] 1301228
```

```r
plot(decompose(snp.ts))
```



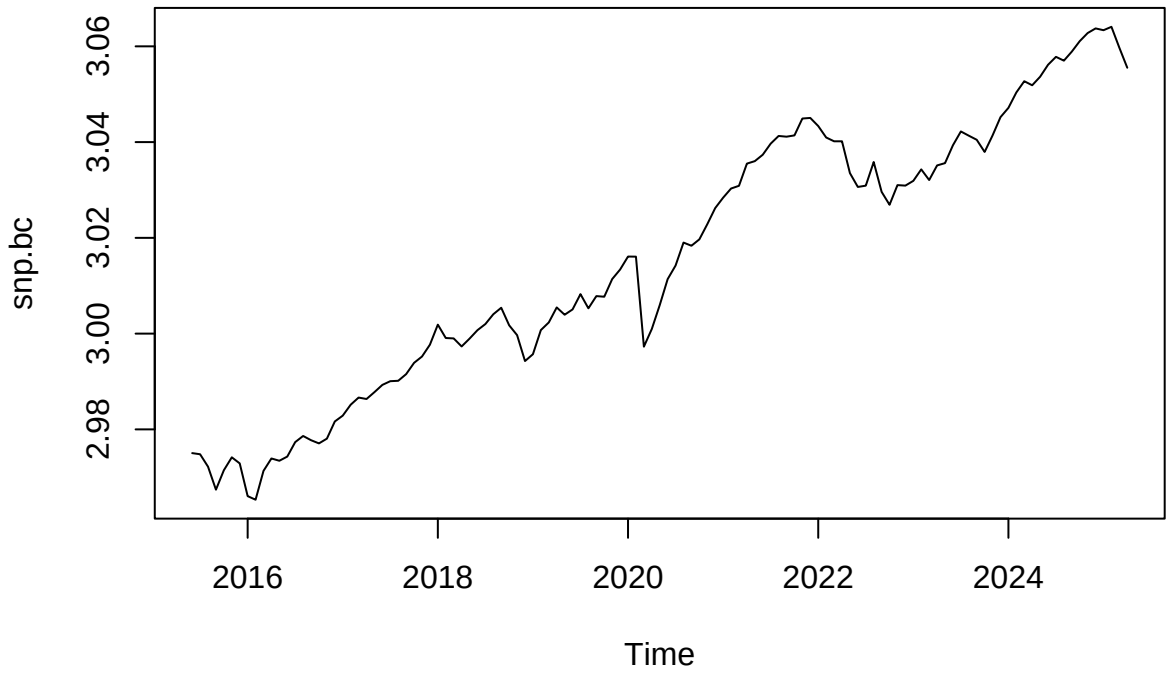

```r
# Box-Cox transformation
index <- 1:length(snp.ts)
bcTransform <- boxcox(snp.ts ~ index, plotit=TRUE)
```




```r
lambda <- bcTransform$x[which(bcTransform$y == max(bcTransform$y))]
lambda
```

```
## [1] -0.3030303
```

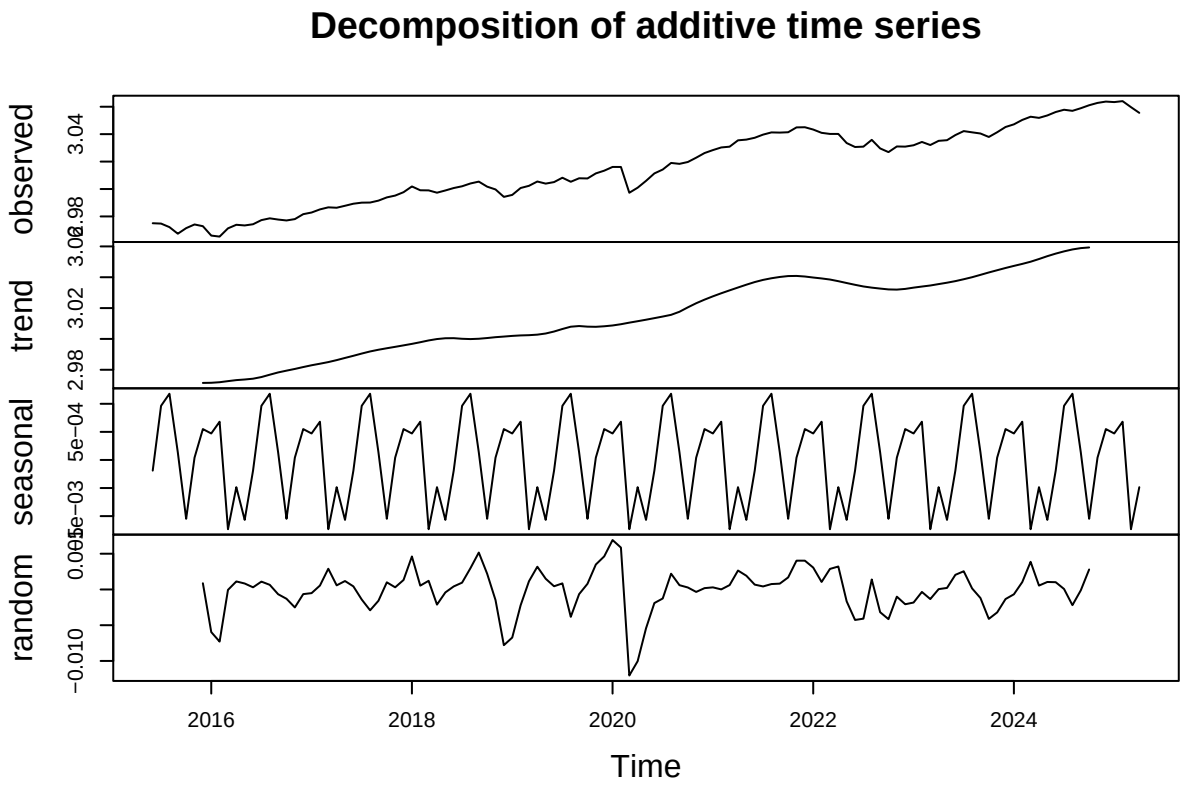
```
snp.bc <- (1/lambda)*(snp.ts^lambda - 1)

plot.ts(snp.bc)
```



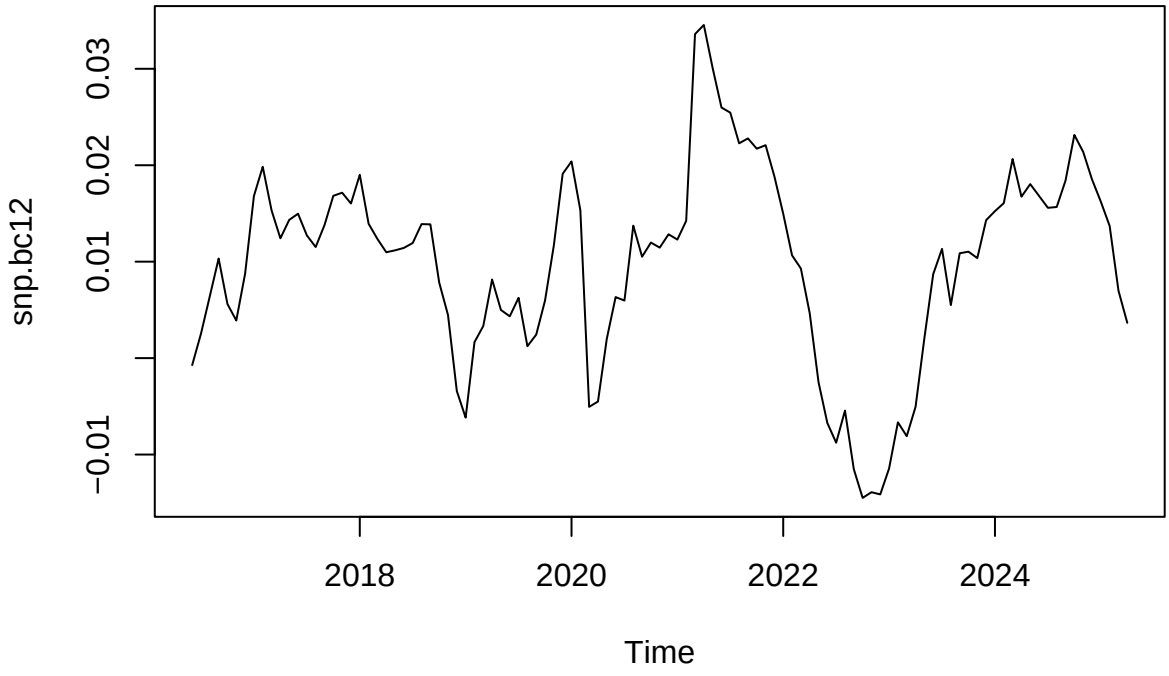

```
var(snp.bc)
```

```
## [1] 0.0007807271
```

```
plot(decompose(snp.bc))
```



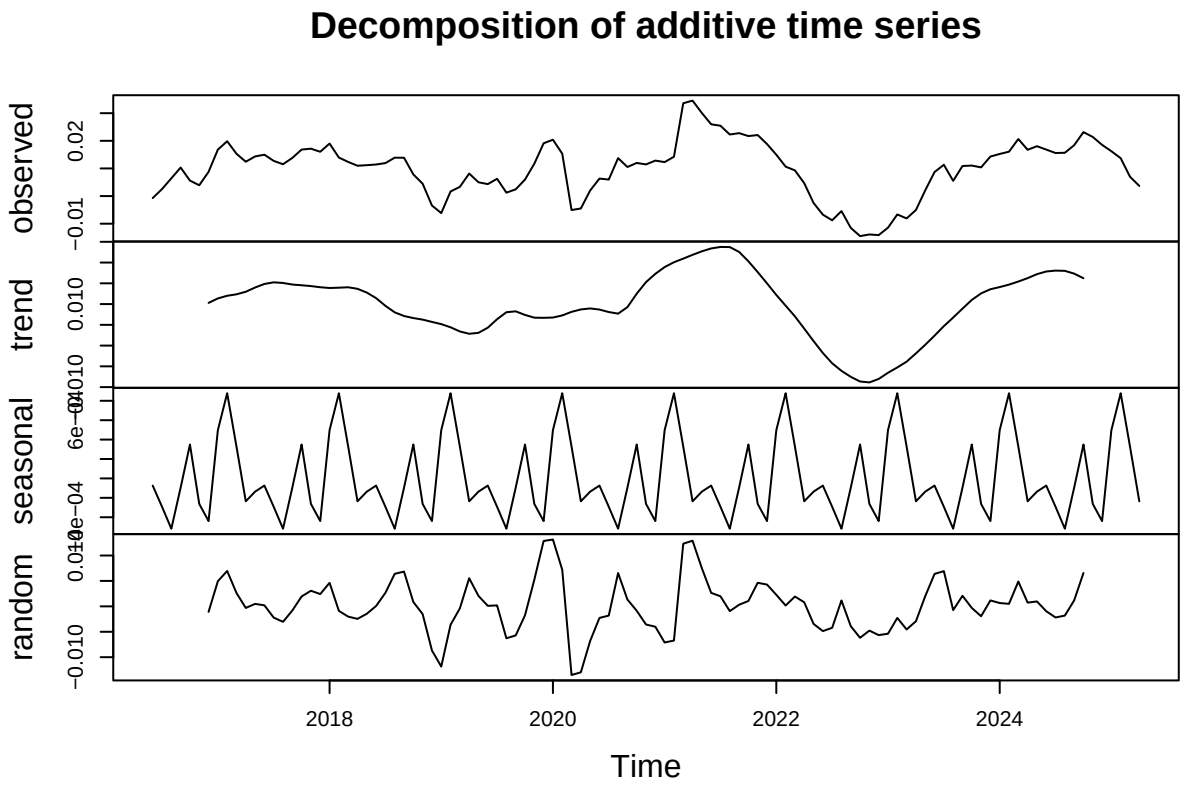

```
# Differencing at lag 12
snp.bc12 <- diff(snp.bc, 12)
plot(snp.bc12)
```



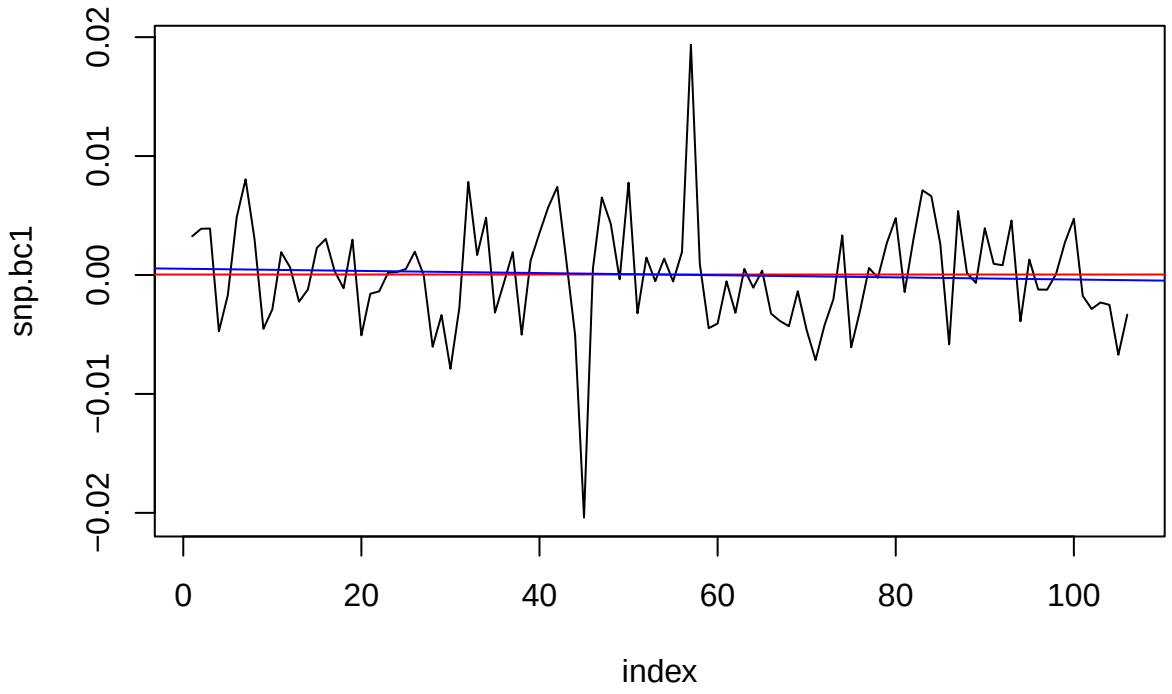

```
var(snp.bc12)
```

```
## [1] 0.0001014235
```

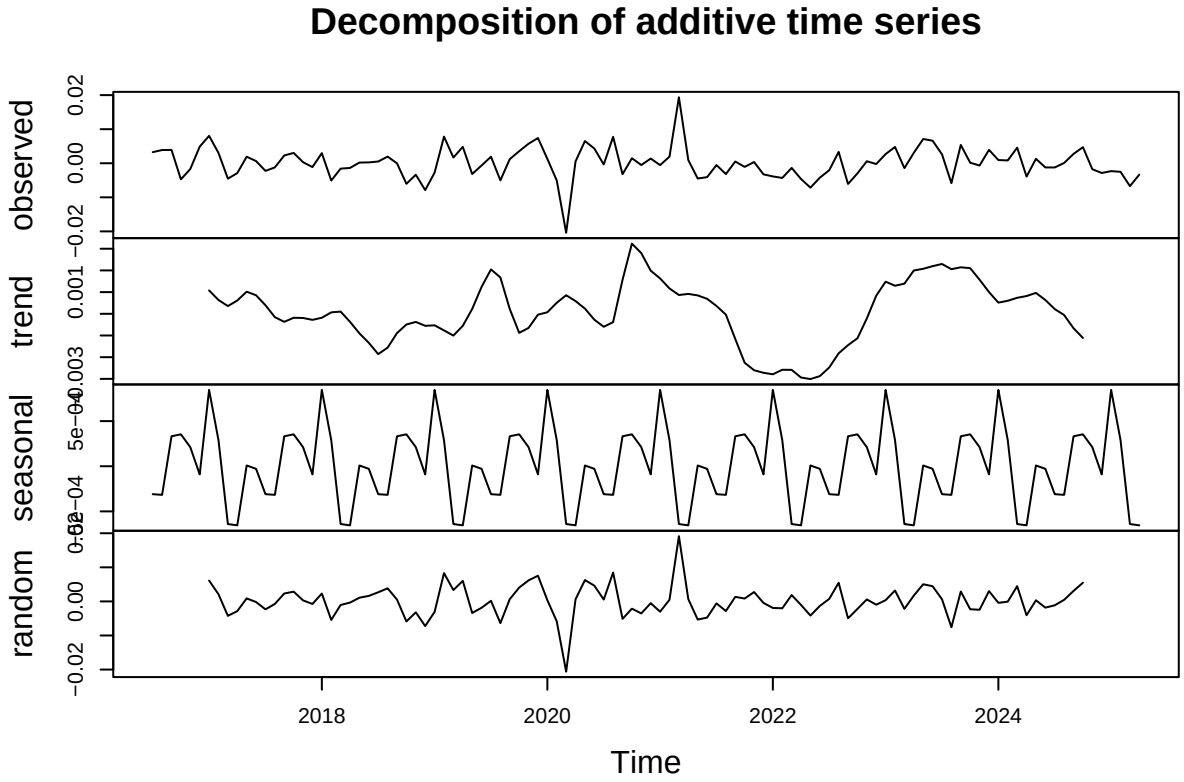
```r
plot(decompose(snp.bc12))
```




```r
# Differencing at lag 1
snp.bc1 <- diff(snp.bc12, 1)
index <- 1:length(snp.bc1)
plot(index, snp.bc1, type='l')
abline(h=mean(snp.bc1), col="red")
abline(lm(snp.bc1 ~ index), col="blue")
```




```
plot(decompose(snp.bc1))
```



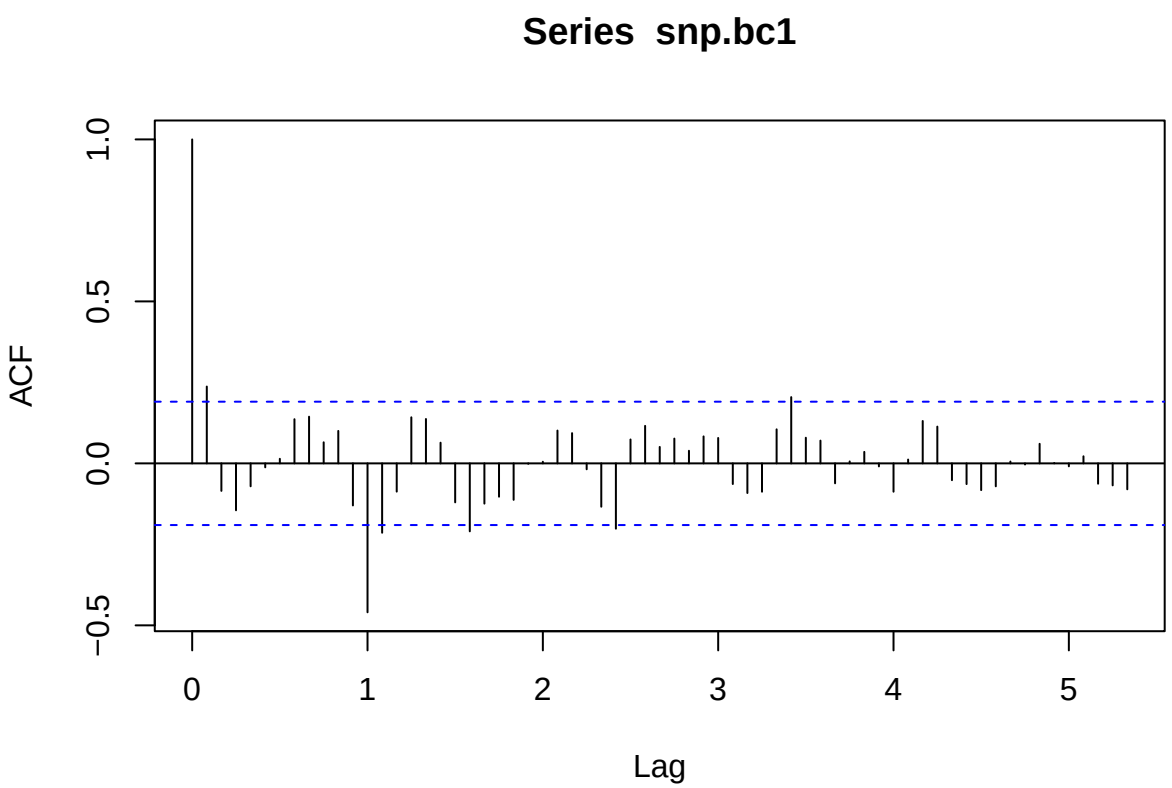

```
mean(snp.bc1)
```

```
## [1] 4.137866e-05
```

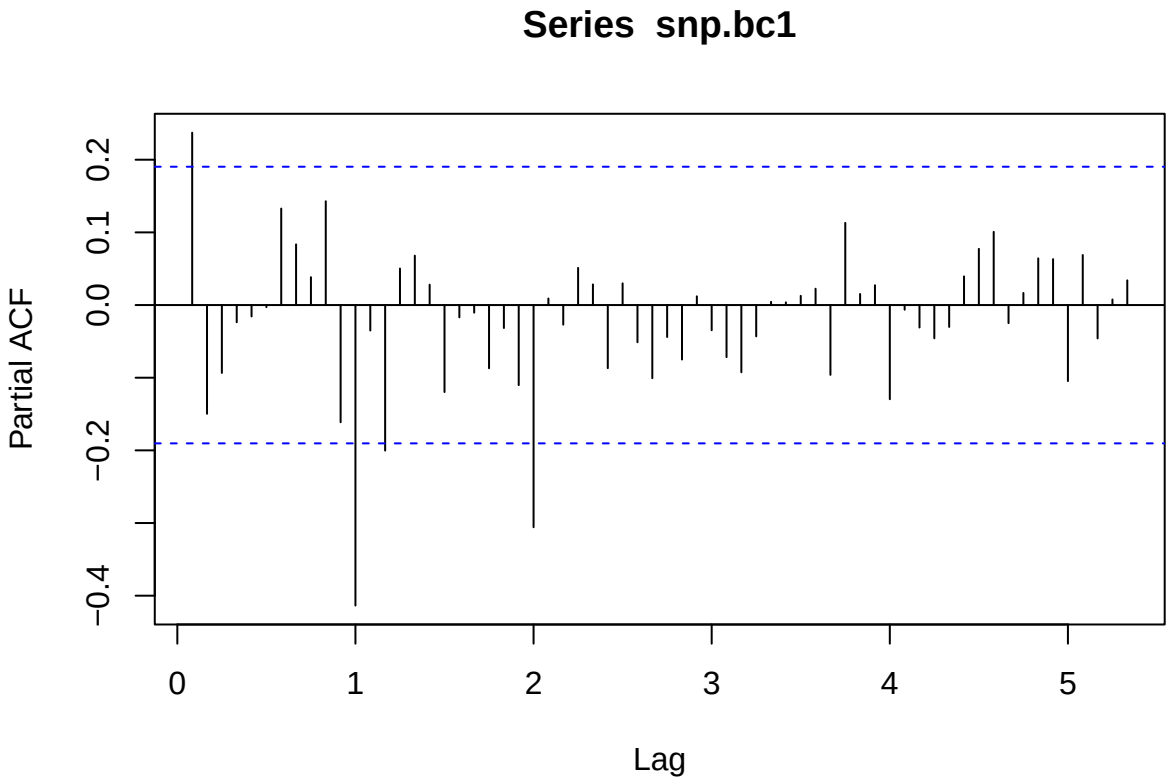
```r
var(snp.bc1)
```

```
## [1] 2.083074e-05
```

```r
acf(snp.bc1, lag.max=64)
```



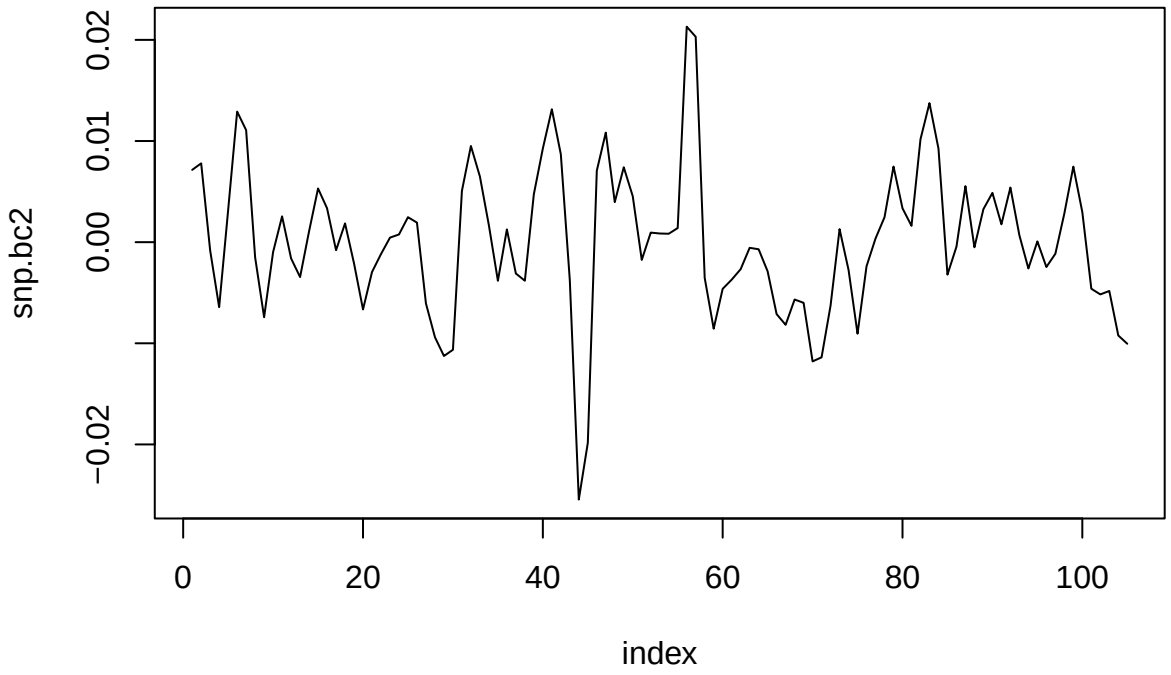

```r
pacf(snp.bc1, lag.max=64)
```



P = 2, Q = 1
p = 1, q = 0 or 1 (probably 1)

So, after analyzing the sample ACF and PACF graphs for our data we attempt to fit the data to a SARIMA$(1,1,1) \times (2,1,1)_{12}$ model.

```
# Differencing at lag 2
snp.bc2 <- diff(snp.bc12, 2)
index <- 1:length(snp.bc2)
plot(index, snp.bc2, type='l')
```




```
#abline(h=mean(snp.bc2), col="red")
#abline(lm(snp.bc2 ~ index), col="blue")
```

```
mean(snp.bc2)
```

```
## [1] 8.424316e-05
```

```
var(snp.bc2)
```

```
## [1] 5.183049e-05
```