Time Series Lab 3: Monthly Wine Sales Australia

Billy Dang

2024-11-08

PSTAT 174/274 Fall 2024 – Lab Assignment 3

Today, we'll be working with a montly wine sales dataset from Australia. We'll start by reading in the file:

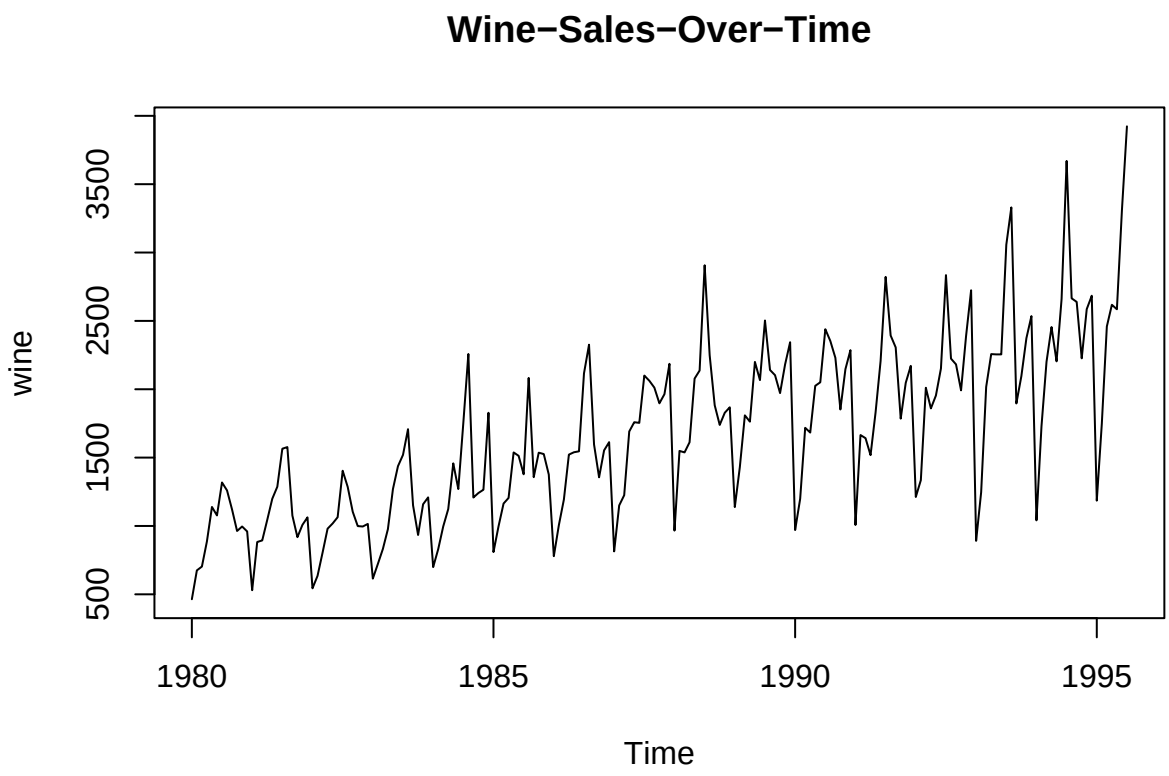
```
getwd()
```

```
## [1] "C:/Users/Billy Dang/Desktop/PSTAT174 - Time Series Analysis/Lab 3 - Monthly Wine Sales"
```

```
wine.csv = read.table("data/monthly-australian-wine-sales-th.csv", sep = ",", header = FALSE, skip = 1,

colnames(wine.csv) <- c('Month', 'Sales')
```

Now let's create a time-series object and plot it:

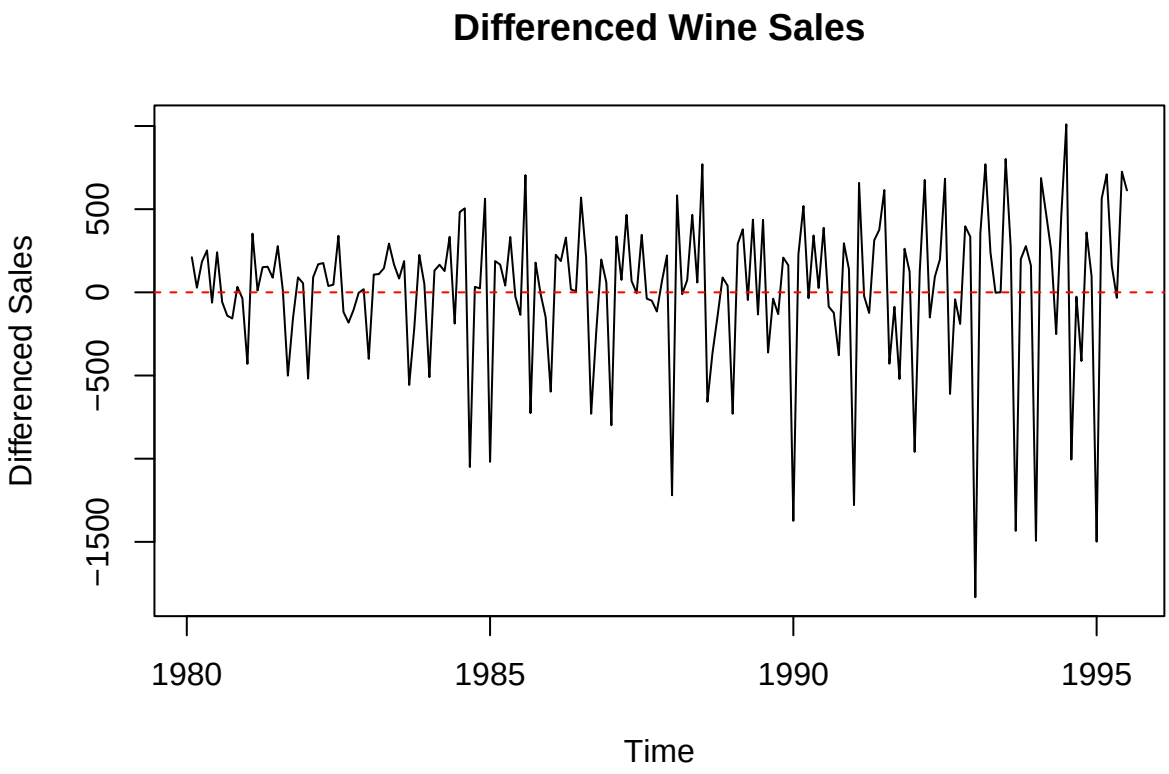
```
wine <- ts(wine.csv$Sales, start = c(1980, 1), frequency = 12)
ts.plot(wine, main = 'Wine-Sales-Over-Time')
```



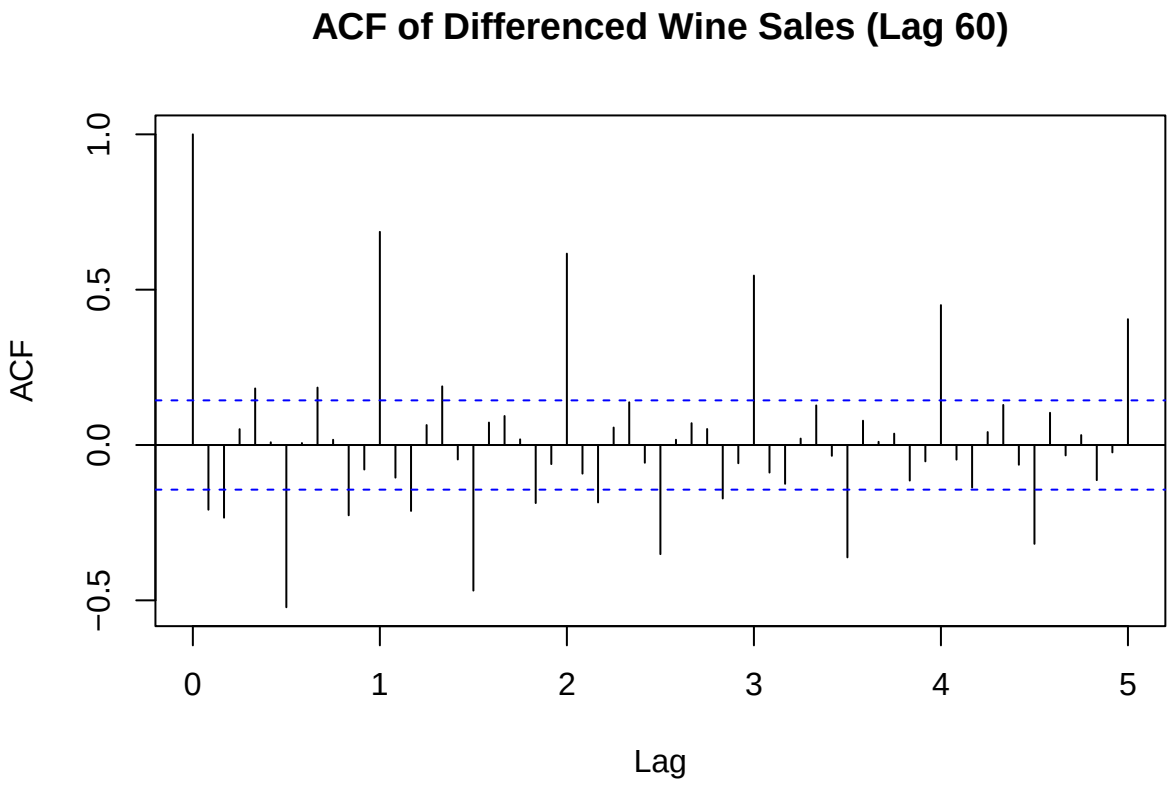
Questions | Part 1 Based on this time series, we can see that the data appears to be increasing over time, the data variability changes over time, and that there appears to be seasonality. Let's explore how to deal with this below:

First, we'll try a simple differencing and examine what it gives us. If X_t represents our wine sales, differencing a series creates a new time series $Y_t = \nabla X_t = X_t - X_{t-1}$. We'll create a difference series, plot that, and plot a line through 0 so we can see if we're close to de-trending.

```
diff_wine <- diff(wine)

plot(diff_wine, main='Differenced Wine Sales', ylab='Differenced Sales', xlab='Time')
abline(h = 0, col='red', lty=2)
```

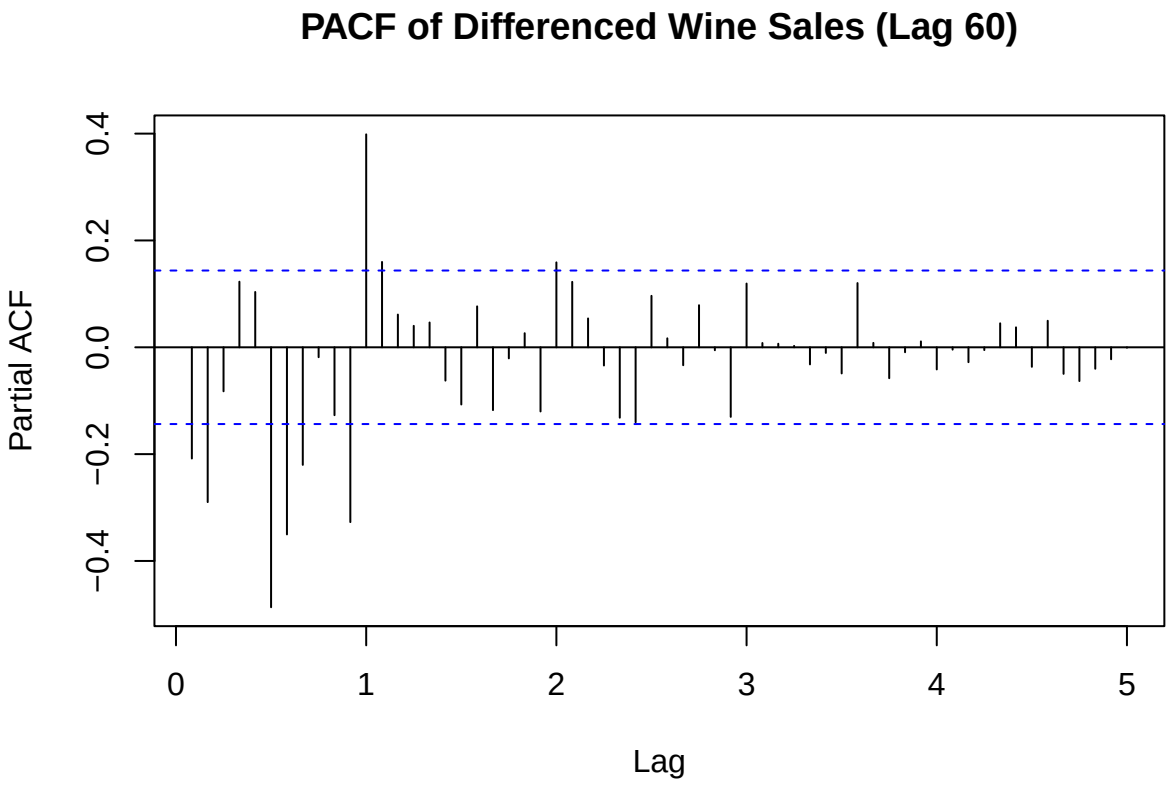



```
acf(diff_wine, lag.max = 60, main='ACF of Differenced Wine Sales (Lag 60)')
```



Based on the ACF plot of the difference wine sales time series above, we can make a couple observations. First, we see significant correlation values within lags across the entire time series. These lags also reveal a seasonal pattern, with the presence of spikes at regular intervals (i.e. around 12, 24, etc.). This repeating pattern suggests that while differencing by 1 step removed the trend, it did not eliminate the 12-month seasonality that is inherently built into the wine sales data.

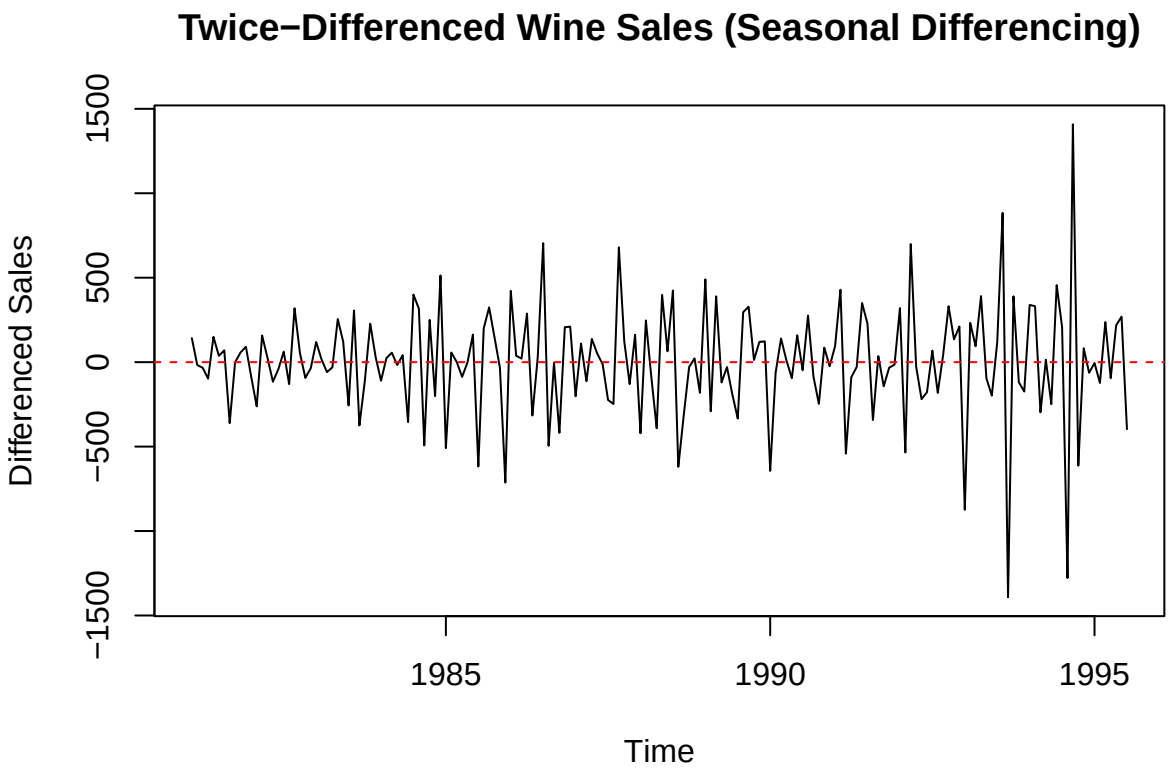
```
pacf(diff_wine, lag.max = 60, main='PACF of Differenced Wine Sales (Lag 60)')
```



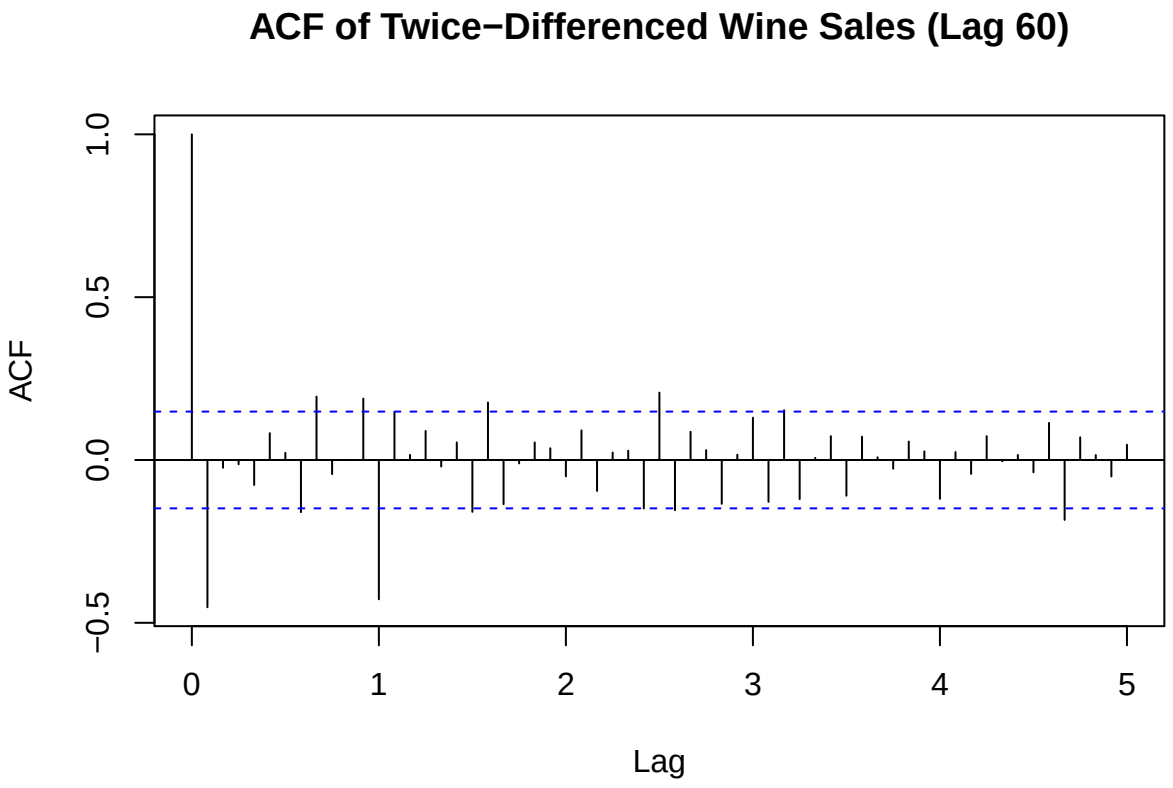
Moving onto our PACF plot, we see a couple of initial significant lags that suggest short-term dependencies in the data. However, beyond the first ~12 lags, the PACF gradually diminishes and no longer holds significance based on the confidence bands.

As we have reason to believe that there is a 12-month cycle to sales, we'll difference the series again, but for 12 steps this time. We'll then plot the time-series, the ACF, and the PACF.

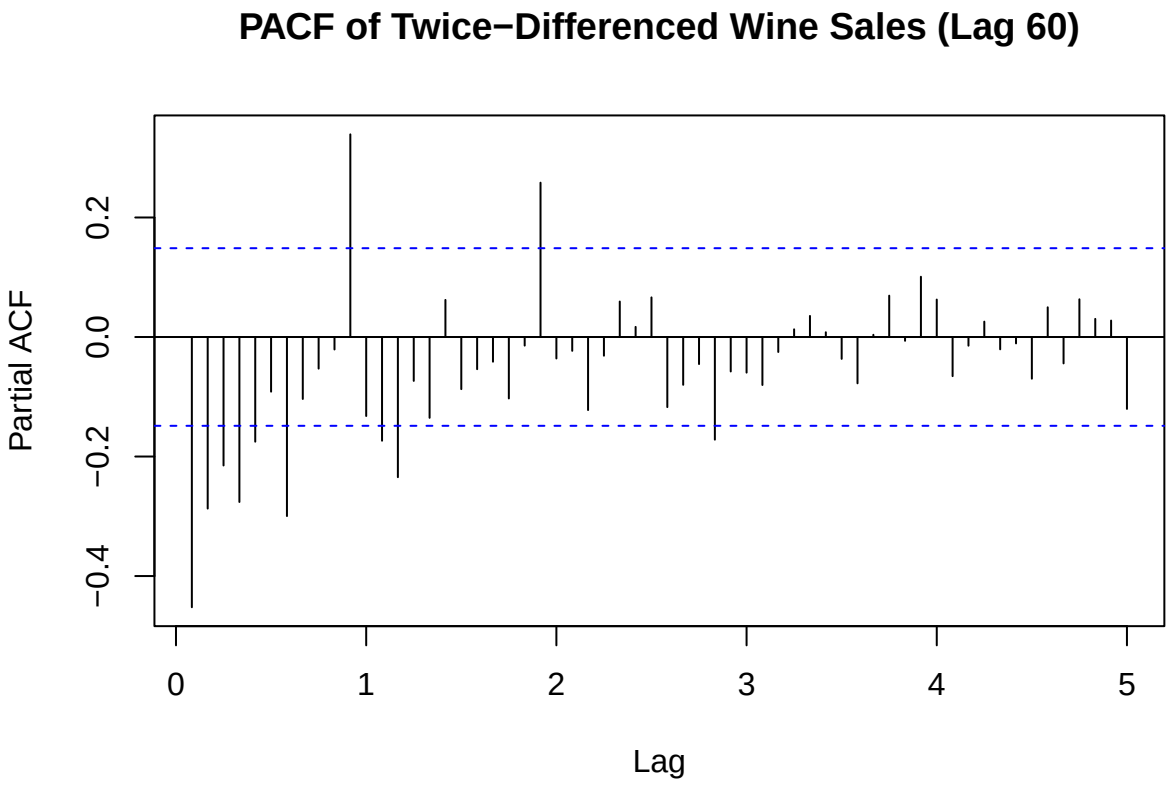
```
#Second differencing (seasonal differencing with lag of 12)
diff_wine_12 <- diff(diff_wine, lag = 12)

plot(diff_wine_12, main='Twice-Differenced Wine Sales (Seasonal Differencing)', ylab='Differenced Sales
abline(h = 0, col='red', lty=2)
```




```
acf(diff_wine_12, lag.max = 60, main='ACF of Twice-Differenced Wine Sales (Lag 60)')
```




```
pacf(diff_wine_12, lag.max = 60, main='PACF of Twice-Differenced Wine Sales (Lag 60)')
```



Questions | Part 2

```
library(astsa)
library(forecast)
```

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

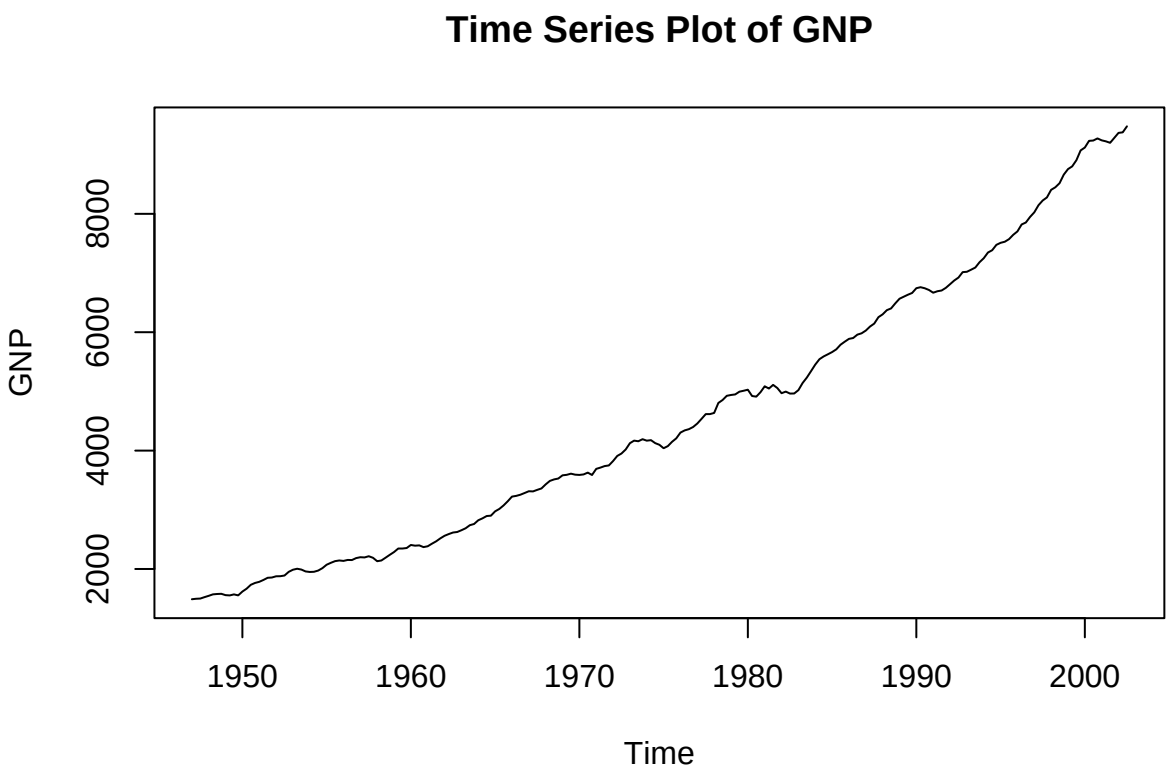
```
##
## Attaching package: 'forecast'
```

```
## The following object is masked from 'package:astsa':
##
##     gas
```

For this problem, we will be looking at the procedure and steps for fitting ARIMA(p, d, q) models to time series data. We consider the gnp data from the astsa package.

First, produce a time series plot of the gnp data using the function ts.plot. Examine if the data is stationary, if there is any evidence of trends or seasonality, and propose any steps to obtain stationary data.

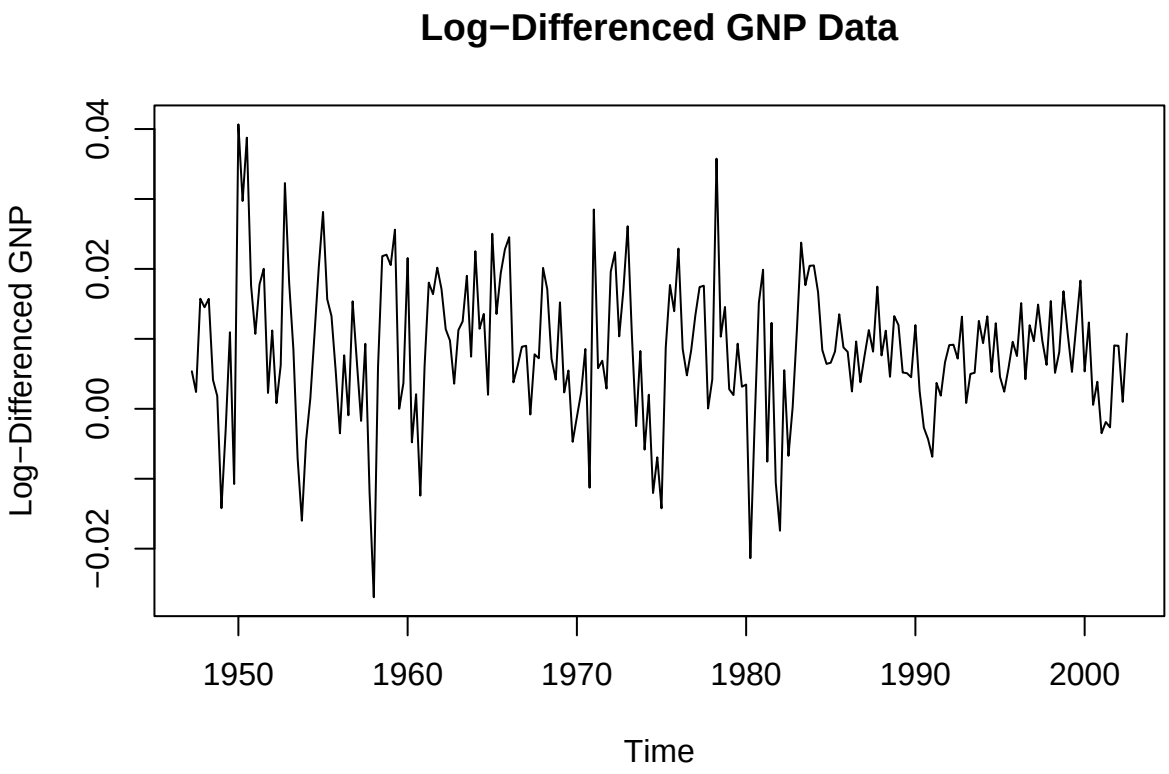
```
ts.plot(gnp, main = "Time Series Plot of GNP", ylab = "GNP", xlab = "Time")
```



The data is *NOT* stationary as we see a clear upward trend over time which makes sense for a value that measure the total value of all goods and serviced produced by a country's residents and business over time. While the data does not show any signs of seasonality, the claim of not stationary holds true. For the following data, we can obtain stationary by taking the logarithm / difference of the time series.

Next we'll take the log differences of the data and produce a second time series plot. We'll examine what impact this has on the time series and whether or not the new time series now appears stationary.

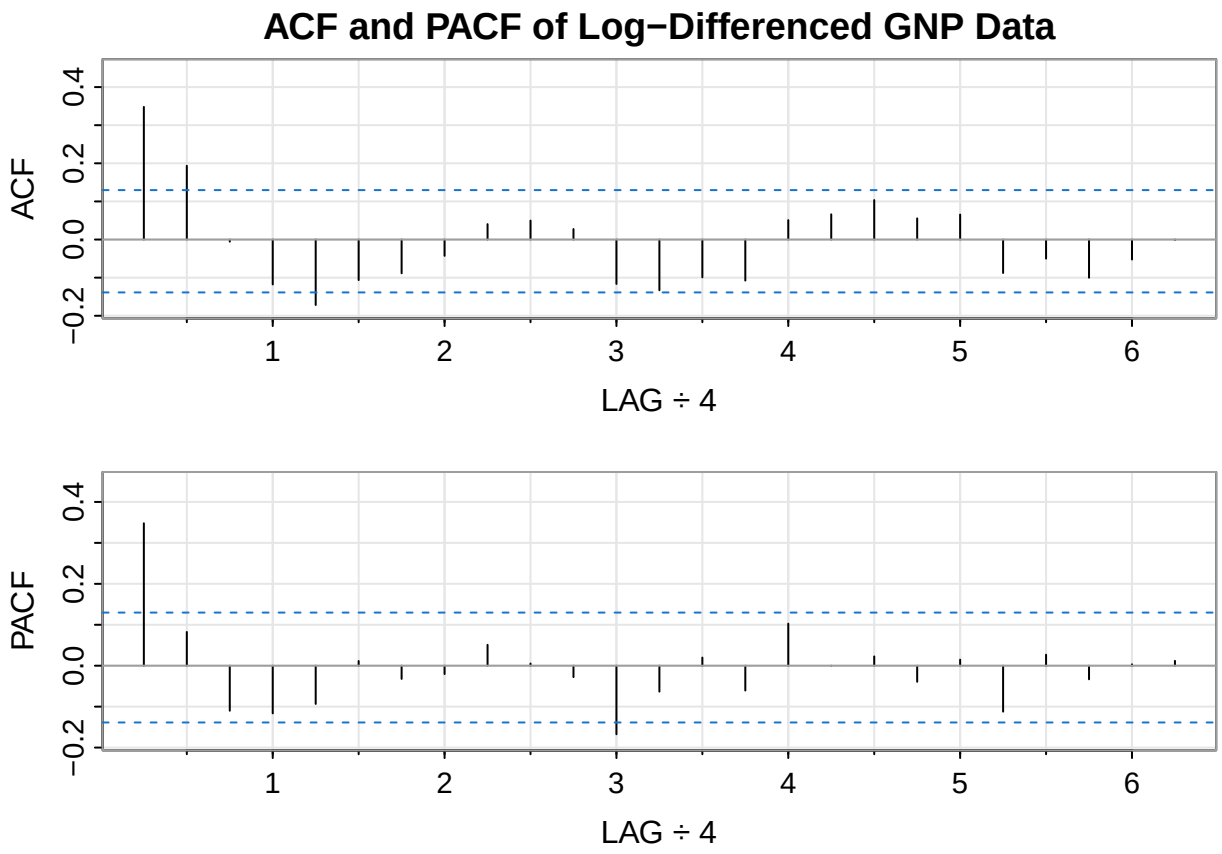
```
gnp_log <- log(gnp)
gnp_log_diff <- diff(gnp_log)

ts.plot(gnp_log_diff, main = "Log-Differenced GNP Data", ylab = "Log-Differenced GNP", xlab = "Time")
```



Based on the graph above, the log-difference transformation has seems to stabilized the time series around a constant mean by removing the upward-trend seen in the previous plot. Additionally, the series' variance seems to have been stabilized a bit as well.

Now, we'll produce an ACF plot and a PACF plot of the transformed data, and examine the results.

```
acf2(gnp_log_diff, main = "ACF and PACF of Log-Differenced GNP Data")
```




```
##      [,1] [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8] [,9] [,10] [,11] [,12] [,13]
## ACF  0.35 0.19 -0.01 -0.12 -0.17 -0.11 -0.09 -0.04 0.04  0.05  0.03 -0.12 -0.13
## PACF 0.35 0.08 -0.11 -0.12 -0.09  0.01 -0.03 -0.02 0.05  0.01 -0.03 -0.17 -0.06
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF  -0.10 -0.11  0.05  0.07  0.10  0.06  0.07 -0.09 -0.05 -0.10 -0.05  0.00
## PACF  0.02 -0.06  0.10  0.00  0.02 -0.04  0.01 -0.11  0.03 -0.03  0.00  0.01
```

Based on the ACF plot, we see a strong initial spike at lag 1, which gradually decays to zero and holds no sustained significance after lag 2 (except once at the 5th lag). The PACF also shows a significant spike at lag 1, but drops off sharply after lag 1.

```
ma.model <- Arima(gnp_log_diff, order = c(0, 0, 2))

summary(ma.model)
```

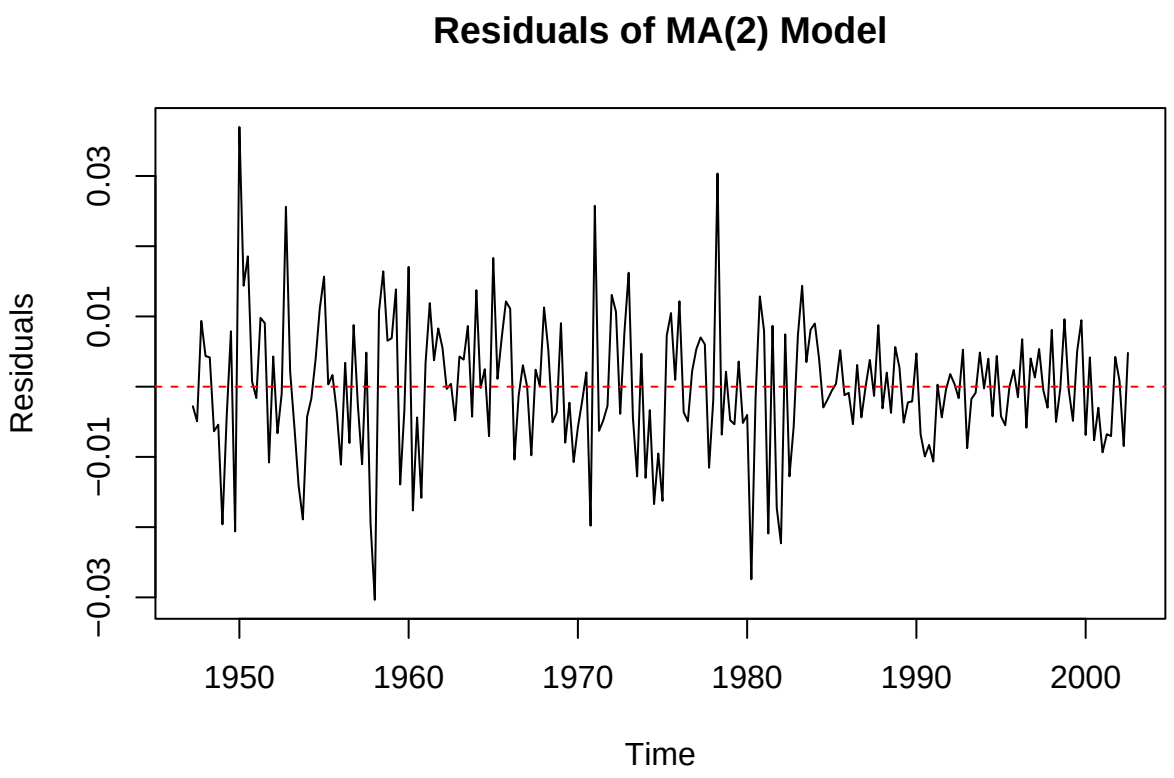
```
## Series: gnp_log_diff
## ARIMA(0,0,2) with non-zero mean
##
## Coefficients:
##          ma1     ma2    mean
##       0.3028  0.2035  0.0083
## s.e.  0.0654  0.0644  0.0010
##
## sigma^2 = 9.041e-05:  log likelihood = 719.96
## AIC=-1431.93   AICc=-1431.75   BIC=-1418.32
```

```
##
## Training set error measures:
##                       ME       RMSE         MAE  MPE MAPE      MASE       ACF1
## Training set 9.940243e-06 0.00944414 0.007108452 -Inf  Inf 0.6185504 0.01725908
```

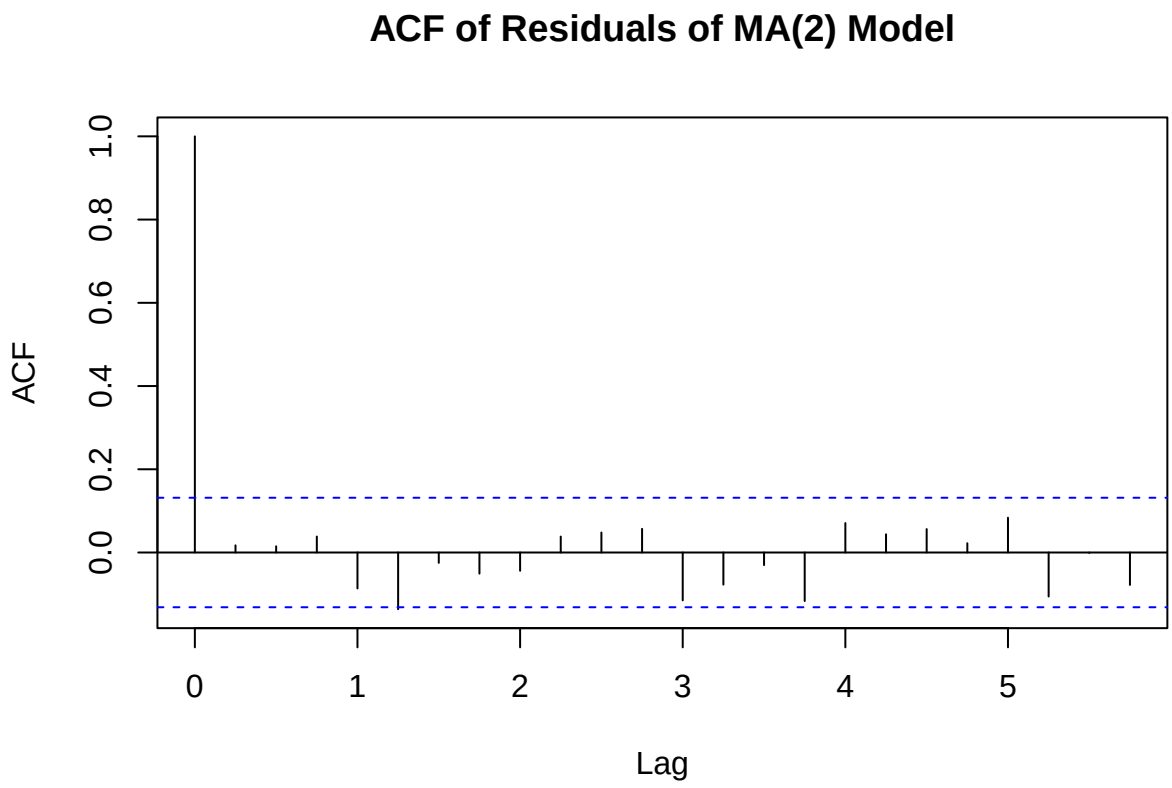
After fitting out MA(2) model, we see θ_1 is .3028 with a SE of .0654, and θ_2 is .2035 with a SE of .0644.

We'll evaluate the fit of our model by producing a plot of the residuals as well as an ACF plot.

```
plot(ma.model$residuals, main = "Residuals of MA(2) Model", ylab = "Residuals", xlab = "Time")
abline(h = 0, col = "red", lty = 2)
```



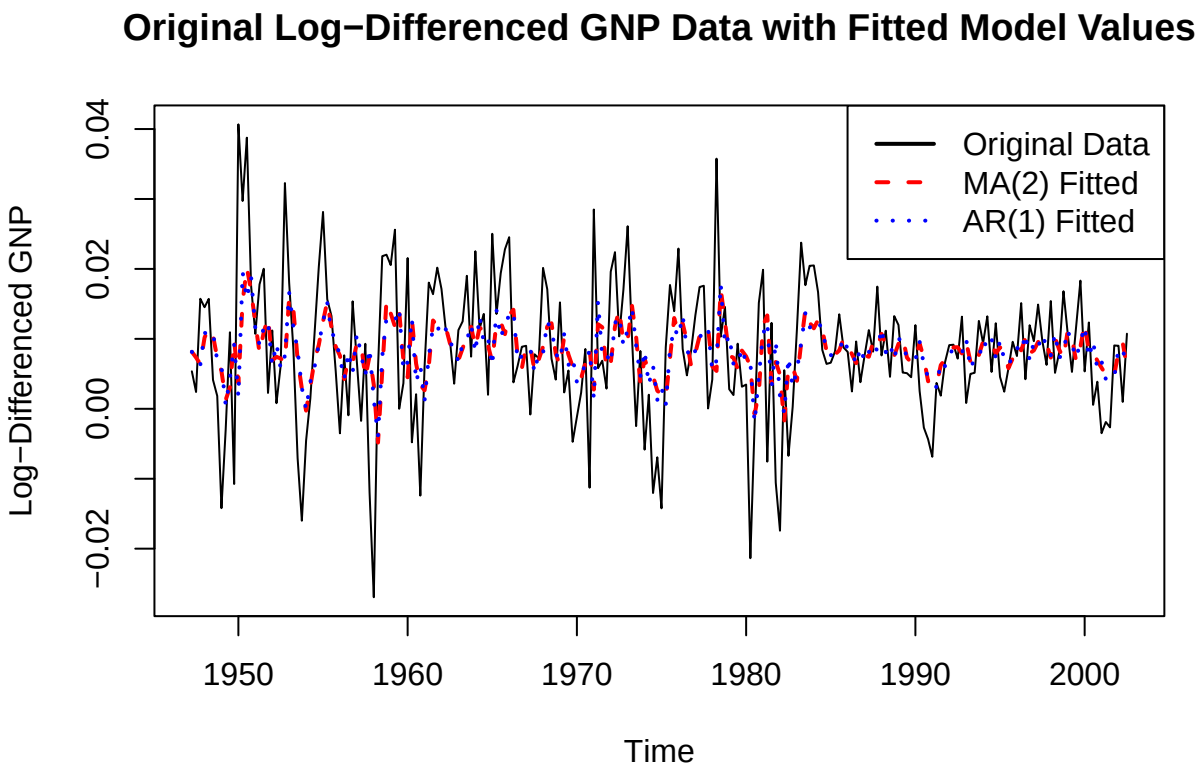

```
acf(ma.model$residuals, main = "ACF of Residuals of MA(2) Model")
```



The residual plot and ACF of the residuals indicate that the MA(2) model is a good fit for the log-differenced GNP data. This is shown by how the residuals fluctuated around zero with any clear patterns or trends over time, indicated that the model has likely captured most of the structure in the data. Additionally, the absence of significant auto correlations within the residuals indicates it behaves like white noise, which means the model has captured the main patterns in the data well and mainly leaves behind white noise.

Finally, we'll produce a plot of the original transformed data series with another fitted model values over-layed in different colors.

```
ar.model <- Arima(gnp_log_diff, order = c(1, 0, 0))


plot(gnp_log_diff, main = "Original Log-Differenced GNP Data with Fitted Model Values",
     ylab = "Log-Differenced GNP", xlab = "Time", col = "black", type = "l")


lines(ma.model$fitted, col = "red", lwd = 2, lty = 2)
lines(ar.model$fitted, col = "blue", lwd = 2, lty = 3)

legend("topright", legend = c("Original Data", "MA(2) Fitted", "AR(1) Fitted"),
       col = c("black", "red", "blue"), lty = c(1, 2, 3), lwd = 2)
```


Original Log−Differenced GNP Data with Fitted Model Values
Log−Differenced GNP
−0.02
0.00
0.02
0.04
1950
1960
1970
1980
1990
2000
Time
Original Data
MA(2) Fitted
AR(1) Fitted