STAT 443: Lab 1

Alex Grinius (20712972)

14 March, 2025

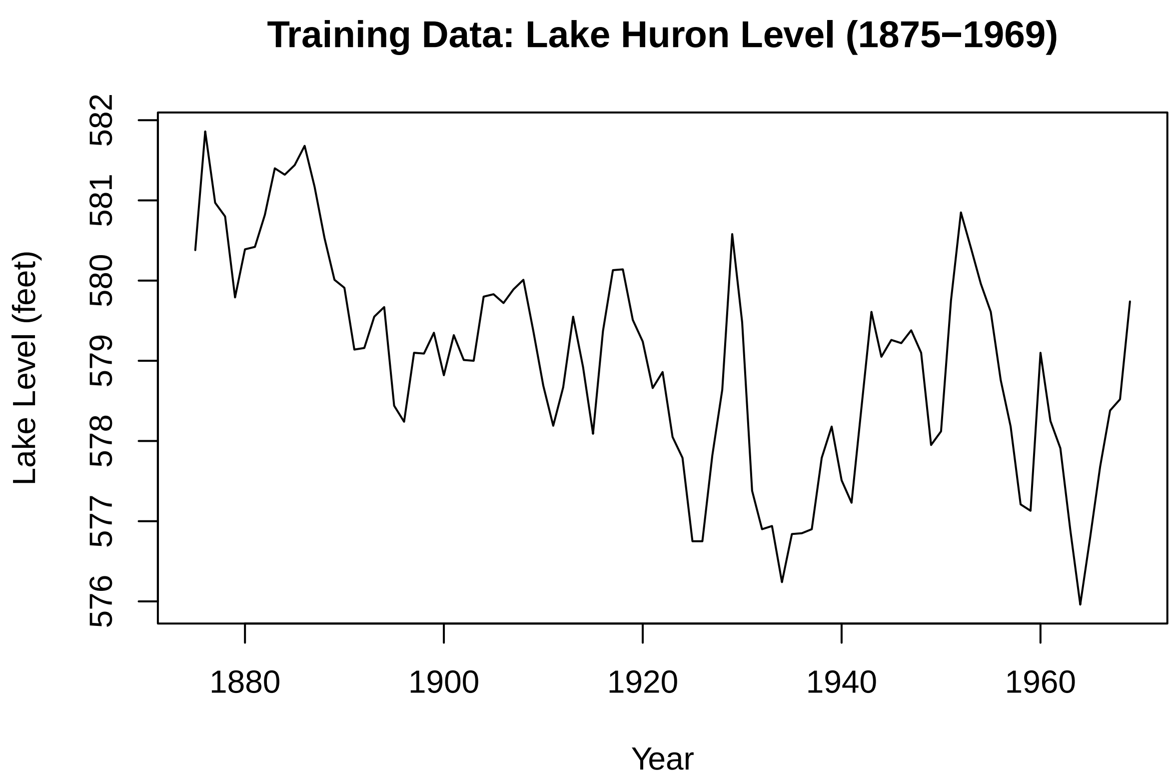
Question 1

(a)

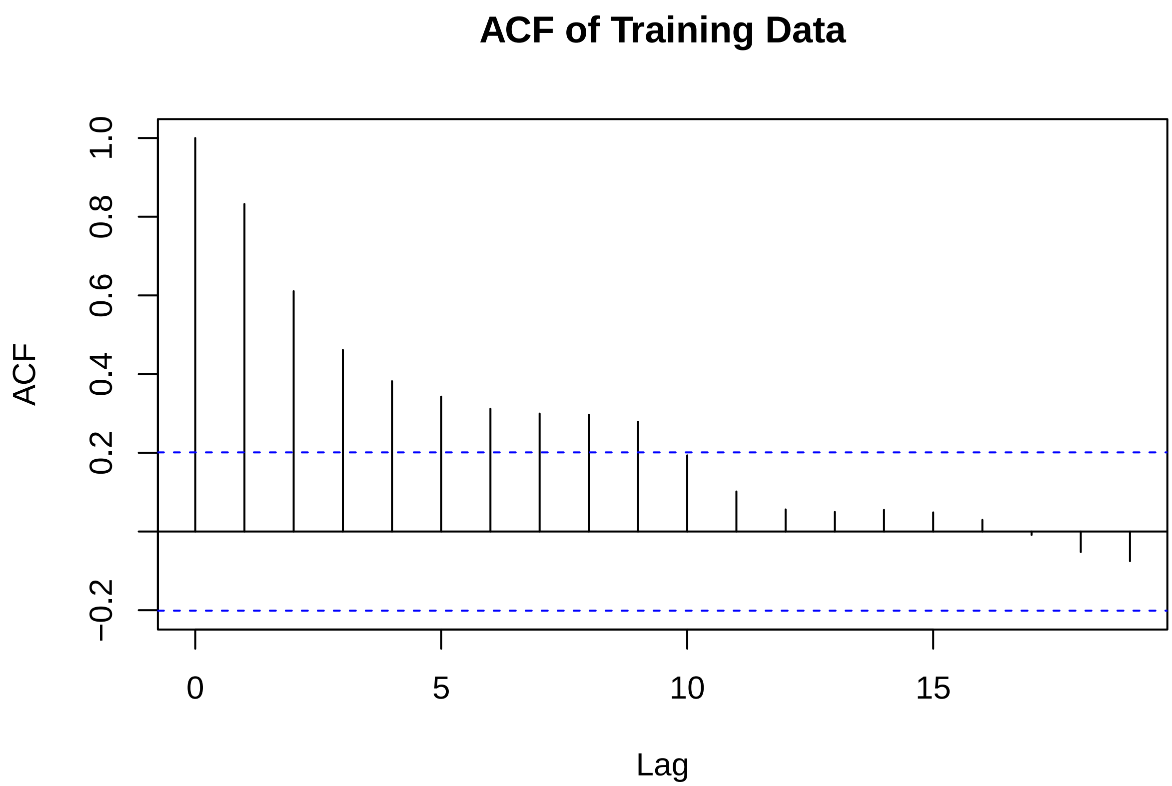
```
data <- LakeHuron
training_data <- window(LakeHuron, start = 1875, end = 1969)
test_data <- window(LakeHuron, start = 1970, end = 1972)
```

(b)

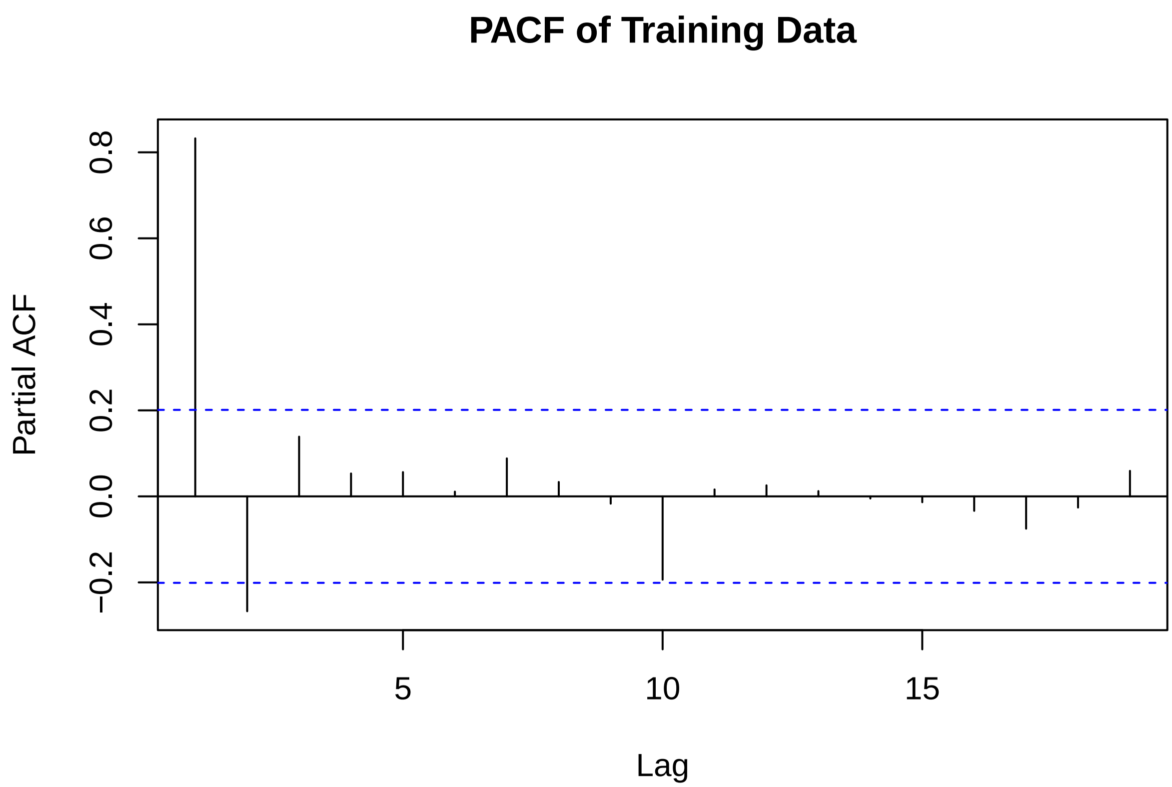
```
plot(training_data, main = "Training Data: Lake Huron Level (1875-1969)", ylab = "Lake Level (feet)", xl
```



```r
acf(training_data, main = "ACF of Training Data")
```




```r
pacf(training_data, main = "PACF of Training Data")
```



From the ACF we see that significance cuts off at q = 9 (looking closely, 10 is below the line). The PACF cuts off at p = 2, with again that weird spike at 10, not quite hitting significance. Thus I am moving forward with an ARMA(2,9) model.

(c)

```
# this is where your R code goes
```

Question 2

(a)

```
model1 <- arima(training_data, order = c(2,0 ,9))
model2 <- arima(training_data, order = c(2,0 ,10))
print(model1)
```

```
##
## Call:
## arima(x = training_data, order = c(2, 0, 9))
##
## Coefficients:
##          ar1      ar2      ma1      ma2     ma3     ma4     ma5      ma6
##       1.4143  -0.6254  -0.3609  -0.1276  0.1141  0.0790  0.1743  -0.0343
## s.e.  0.4397   0.4920   0.4535   0.1231  0.2553  0.1972  0.1735   0.1302
##          ma7      ma8      ma9  intercept
```

```
##       0.0929  0.0517  0.3471   579.0064
## s.e.  0.1563  0.1549  0.1586     0.4090
##
## sigma^2 estimated as 0.4307:  log likelihood = -96.6,  aic = 219.2
```

```r
print(model2)
```

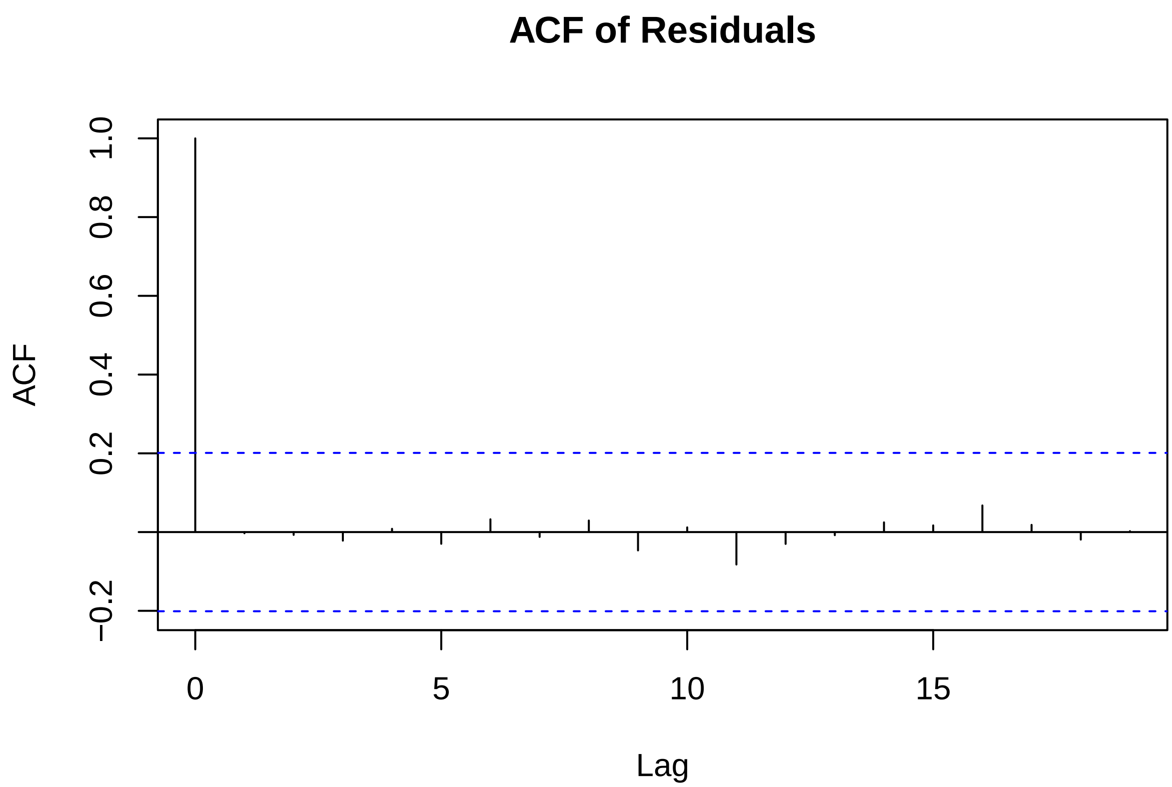
```
##
## Call:
## arima(x = training_data, order = c(2, 0, 10))
##
## Coefficients:
##          ar1     ar2     ma1     ma2      ma3      ma4     ma5     ma6     ma7
##       0.1017  0.4833  0.9592  0.1555  -0.0913  -0.0675  0.0236  0.0793  0.0262
## s.e.  0.2585  0.2196  0.2715  0.2554   0.2142   0.1691  0.1463  0.1393  0.1683
##          ma8     ma9    ma10  intercept
##       0.1073  0.3130  0.3761   579.0455
## s.e.  0.1861  0.1853  0.1560     0.4473
##
## sigma^2 estimated as 0.4386:  log likelihood = -97.35,  aic = 222.71
```

Here I verify that the AIC is lower for the model with q = 9 than it is for q = 10

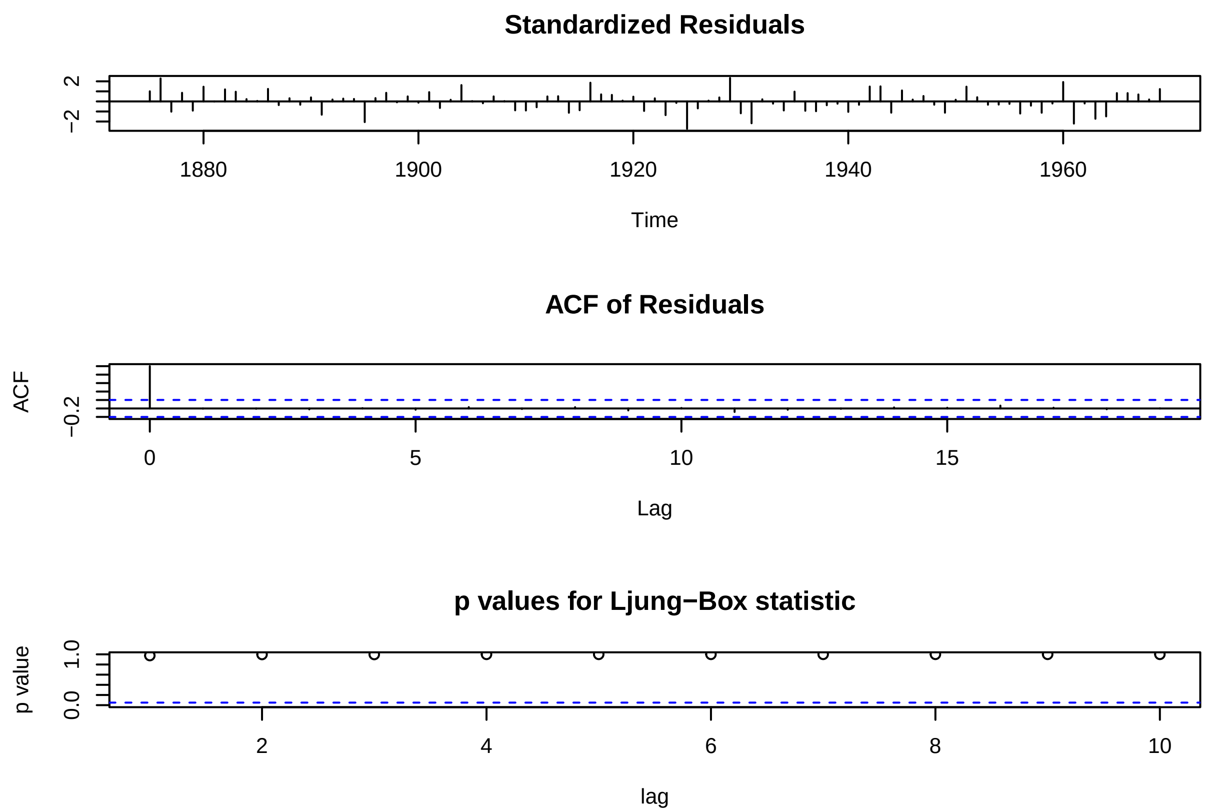
Question 3

(a)

```r
acf(residuals(model1), main = "ACF of Residuals")
```




```r
tsdiag(model1)
```



The ACF shows now significant corelations after lag = 0, and the Ljung-Box plots don't show a pattern in the residuals. This indicates a decent fit that we can move forward with.

(b)

```
# this is where your R code goes
```

Question 4

(a)

```
forecast <- predict(model1, n.ahead = 3)
```

(b)

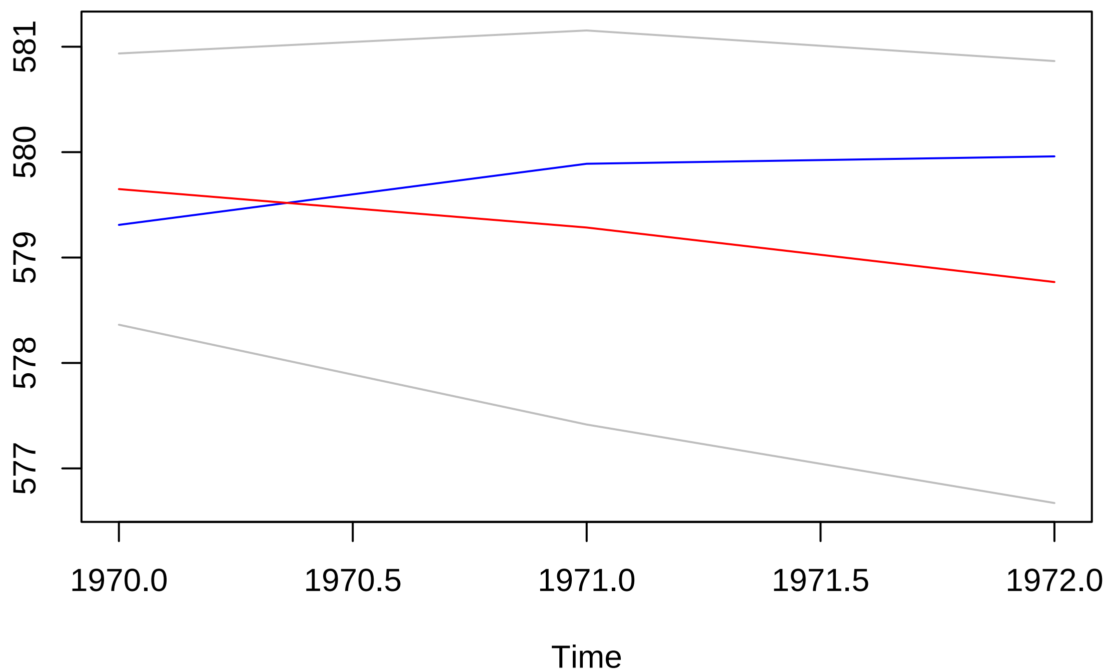
```
comparison <- data.frame(
  Year = 1970:1972,
  Forecast = forecast$pred,
  True_Value = test_data,
  Lower_Bound = forecast$pred - 1.96 * forecast$se,
  Upper_Bound = forecast$pred + 1.96 * forecast$se
)
print(comparison)
```

```
##   Year Forecast True_Value Lower_Bound Upper_Bound
## 1 1970 579.6495     579.31    578.3629    580.9361
## 2 1971 579.2851     579.89    577.4160    581.1542
## 3 1972 578.7678     579.96    576.6716    580.8640
```

Question 5

(a)

```
ts.plot(comparison$True_Value, comparison$Forecast, comparison$Lower_Bound, comparison$Upper_Bound, col
```



On this gtraph the blue is the actual value, the red is the forcast prediction and the two grey lines are the 95% confidence intervals. As we can see, the model predicts a continuous decline even though the true value goes up from 1970 to 71 & 72. The 95% confidence intervals still contain the true values.