

# DSA 8020 R Session 11: Time Series Analysis I

Whitney

## Contents

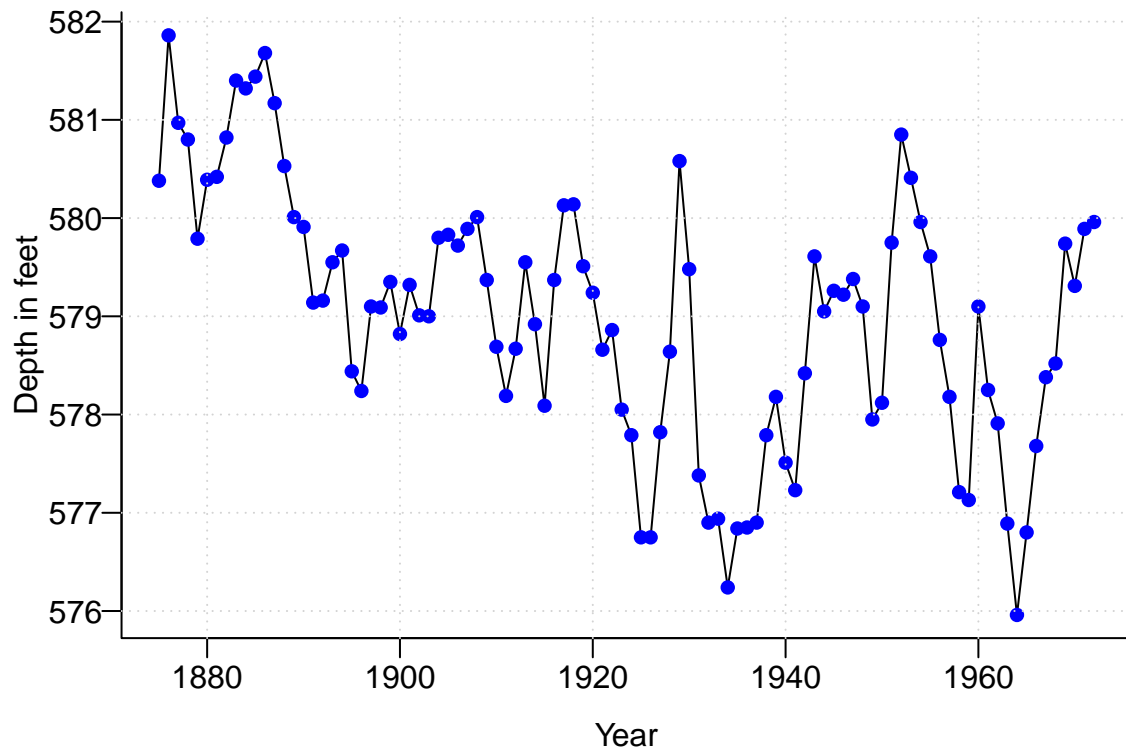
Time Series Data . . . . .	1
Lake Huron Time Series . . . . .	1
CO <sub>2</sub> Concentration . . . . .	2
U.S. monthly unemployment rates . . . . .	3
Time Series Models . . . . .	5
ARMA: Autocovariance Function (ACF) . . . . .	5
ARMA: Partial Autocorrelation Function (PACF) . . . . .	6
Lake Huron Case Study . . . . .	7
Plot the time series data . . . . .	7
Estimate the linear trend . . . . .	8
ARMA model fitting, selection, and diagnostics . . . . .	10
AR(2) . . . . .	13
ARMA(2, 1) . . . . .	14
Model selection using AIC . . . . .	16
AR(2) Fitting and Forecasting . . . . .	16

## Time Series Data

### Lake Huron Time Series

Annual measurements of the level of Lake Huron in feet

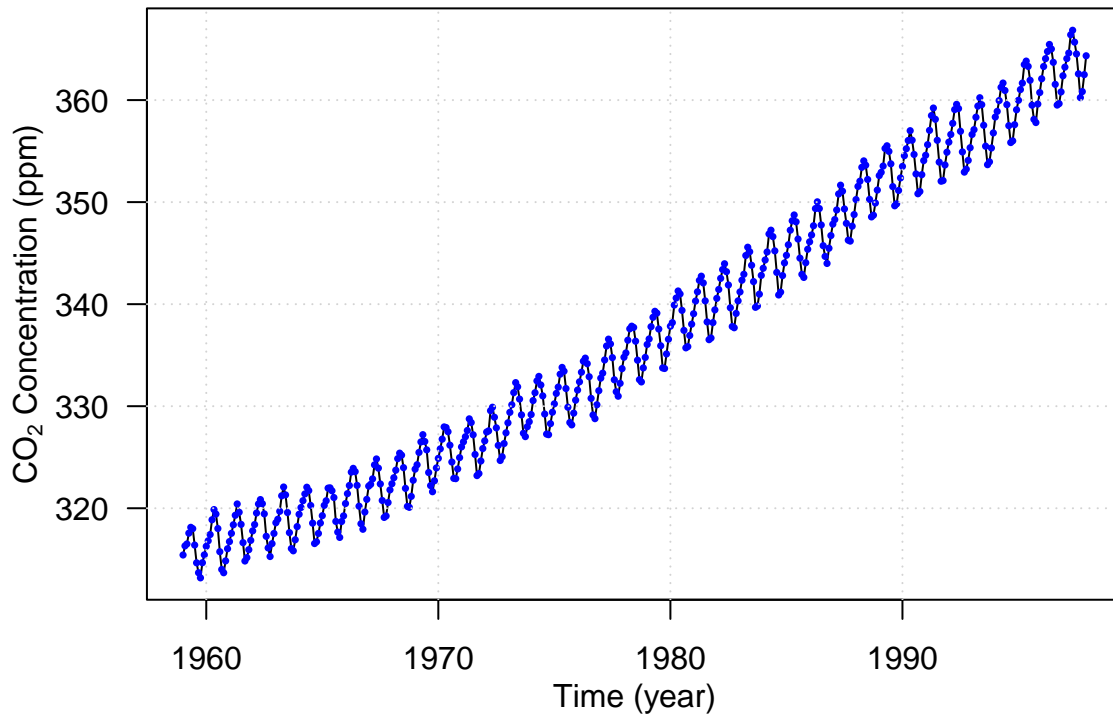
```
par(mar = c(3.2, 3.2, 0.5, 0.5), mgp = c(2, 0.5, 0), bty = "L")
data(LakeHuron)
plot(LakeHuron, ylab = "Depth in feet", xlab = "Year", las = 1)
points(LakeHuron, pch = 16, col = "blue")
grid()
```



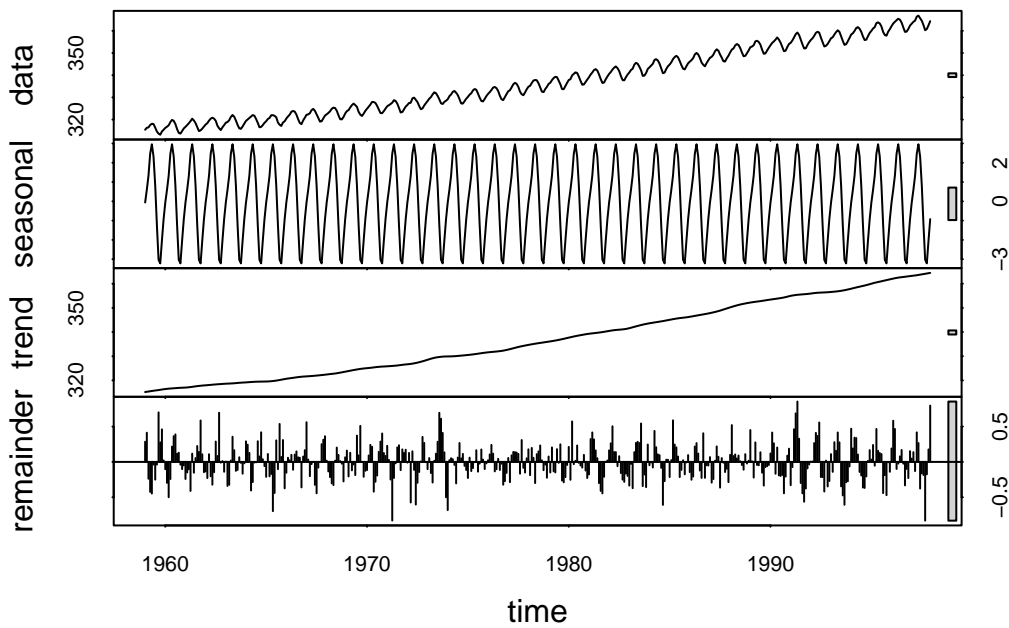
## CO<sub>2</sub> Concentration

Atmospheric concentrations of CO<sub>2</sub> are expressed in parts per million (ppm) and reported using the preliminary 1997 SIO manometric mole fraction scale.

```
data(co2)
par(mar = c(3.8, 4, 0.8, 0.6))
plot(co2, las = 1, xlab = "", ylab = "")
points(co2, pch = 16, col = "blue", cex = 0.5)
mtext("Time (year)", side = 1, line = 2)
mtext(expression(paste("CO"[2], " Concentration (ppm)")), side = 2, line = 2.5)
grid()
```



```
# Seasonal and Trend decomposition using Loess (STL)
par(mar = c(4, 3.6, 0.8, 0.6))
stl <- stl(co2, s.window = "periodic")
plot(stl, las = 1)
```



U.S. monthly unemployment rates

```
library(quantmod)
getSymbols("UNRATE", src = "FRED")
```

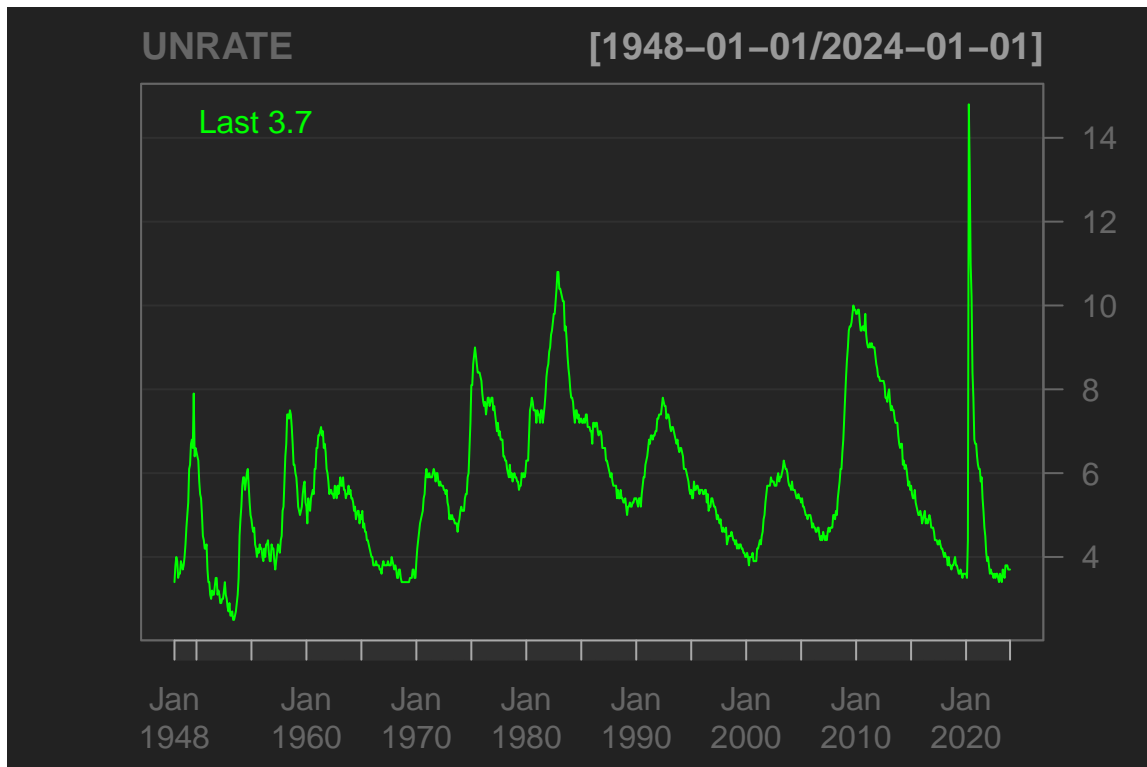
```
## [1] "UNRATE"
```

```
head(UNRATE); tail(UNRATE)
```

```
##           UNRATE
## 1948-01-01    3.4
## 1948-02-01    3.8
## 1948-03-01    4.0
## 1948-04-01    3.9
## 1948-05-01    3.5
## 1948-06-01    3.6
```

```
##           UNRATE
## 2023-08-01    3.8
## 2023-09-01    3.8
## 2023-10-01    3.8
## 2023-11-01    3.7
## 2023-12-01    3.7
## 2024-01-01    3.7
```

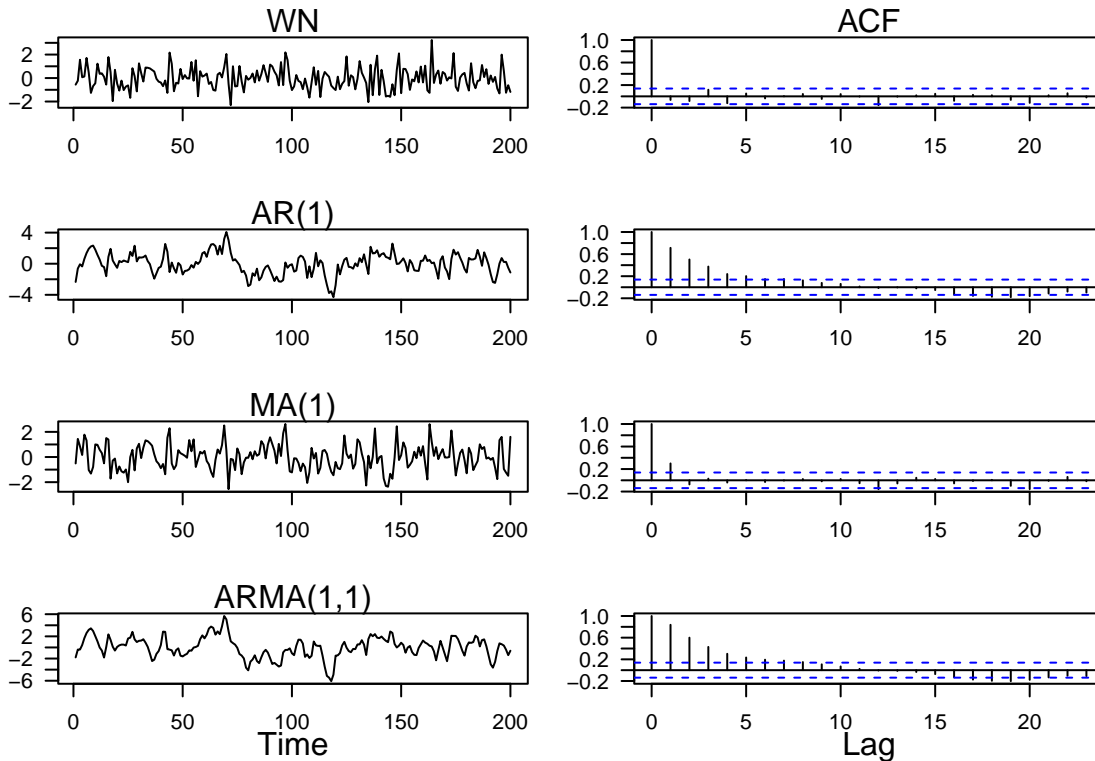
```
chartSeries(UNRATE)
```



## Time Series Models

### ARMA: Autocovariance Function (ACF)

```
set.seed(123)
n = 200
WN <- rnorm(n)
par(mfrow = c(4, 2), mar = c(3.6, 3.6, 1.2, 0.6))
plot(1:n, WN, type = "l", las = 1, xlab = "", ylab = "")
mtext("WN")
acf(WN, xlab = "", ylab = "", main = "", las = 1)
mtext("ACF")
# AR(1) phi = 0.8
set.seed(123)
AR <- arima.sim(n = n, model = list(ar = 0.8))
plot(1:n, AR, type = "l", las = 1, xlab = "", ylab = "")
mtext("AR(1)")
acf(AR, xlab = "", ylab = "", main = "", las = 1)
# MA(1) theta = 0.5
set.seed(123)
MA <- arima.sim(n = n, model = list(ma = 0.5))
plot(1:n, MA, type = "l", las = 1, xlab = "", ylab = "")
mtext("MA(1)")
acf(MA, xlab = "", ylab = "", main = "", las = 1)
# ARMA(1, 1) phi = 0.8, theta = 0.5
set.seed(123)
ARMA <- arima.sim(n = n, model = list(ar = 0.8, ma = 0.5))
plot(1:n, ARMA, type = "l", las = 1, xlab = "", ylab = "")
mtext("ARMA(1,1)")
mtext("Time", side = 1, line = 2)
acf(ARMA, xlab = "", ylab = "", main = "", las = 1)
mtext("Lag", side = 1, line = 2)
```

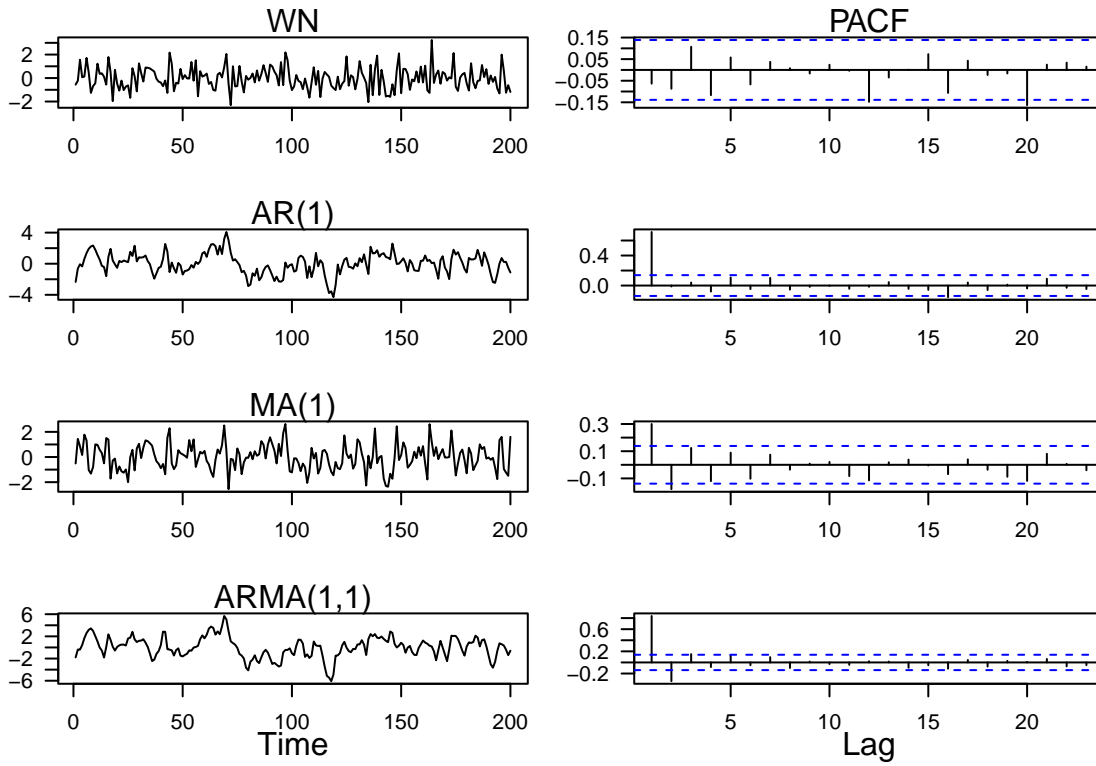


#### ARMA: Partial Autocorrelation Function (PACF)

```

par(mfrow = c(4, 2), mar = c(3.6, 3.6, 1.2, 0.6))
plot(1:n, WN, type = "l", las = 1, xlab = "", ylab = "")
mtext("WN")
pacf(WN, xlab = "", ylab = "", main = "", las = 1)
mtext("PACF")
# AR(1) phi = 0.8
set.seed(123)
AR <- arima.sim(n = n, model = list(ar = 0.8))
plot(1:n, AR, type = "l", las = 1, xlab = "", ylab = "")
mtext("AR(1)")
pacf(AR, xlab = "", ylab = "", main = "", las = 1)
# MA(1) theta = 0.5
set.seed(123)
MA <- arima.sim(n = n, model = list(ma = 0.5))
plot(1:n, MA, type = "l", las = 1, xlab = "", ylab = "")
mtext("MA(1)")
pacf(MA, xlab = "", ylab = "", main = "", las = 1)
# ARMA(1, 1) phi = 0.8, theta = 0.5
set.seed(123)
ARMA <- arima.sim(n = n, model = list(ar = 0.8, ma = 0.5))
plot(1:n, ARMA, type = "l", las = 1, xlab = "", ylab = "")
mtext("ARMA(1,1)")
mtext("Time", side = 1, line = 2)
pacf(ARMA, xlab = "", ylab = "", main = "", las = 1)
mtext("Lag", side = 1, line = 2)

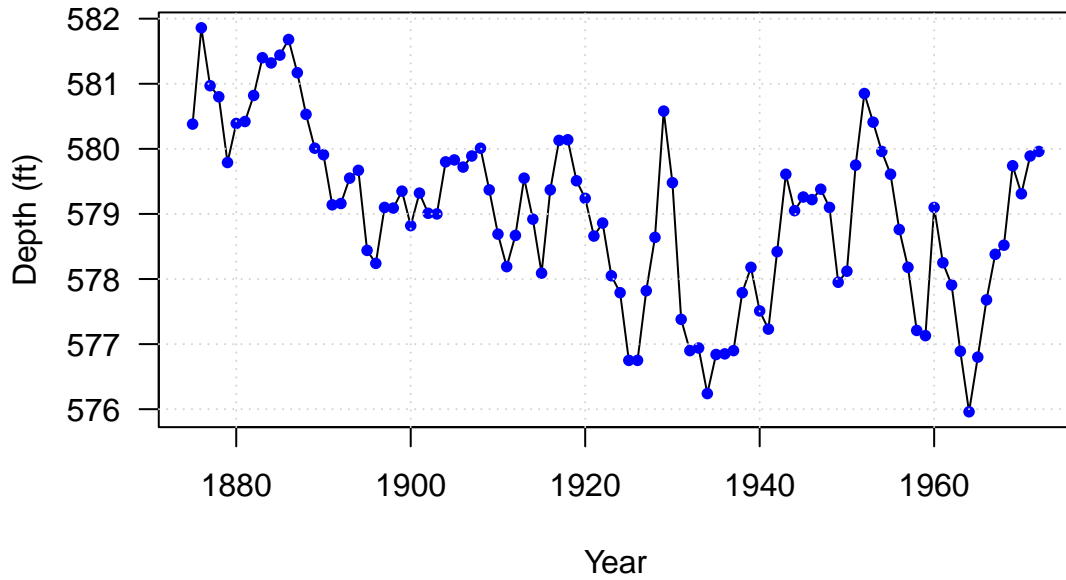
```



## Lake Huron Case Study

Plot the time series data

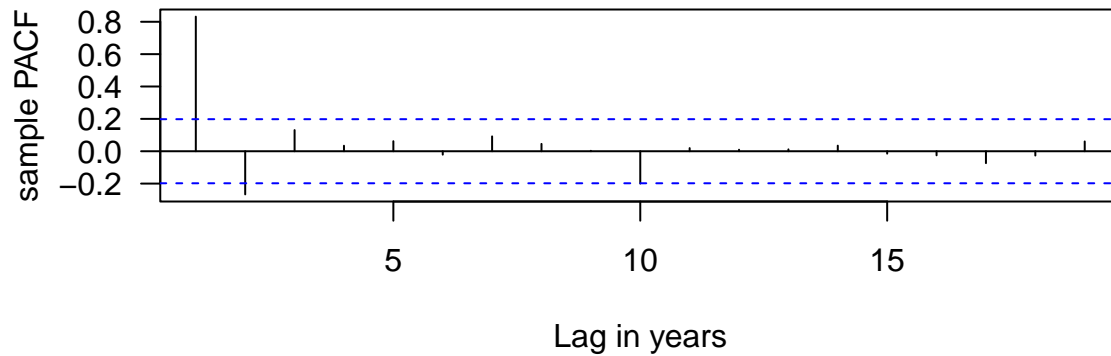
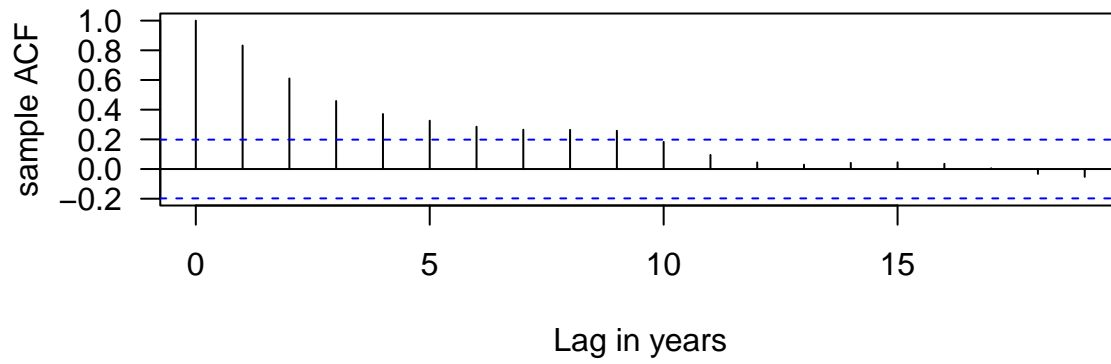
```
## Let us create a 'years' variable.
years <- time(LakeHuron)
## Plot time series
plot(LakeHuron, ylab = "Depth (ft)", xlab = "Year", las = 1)
points(LakeHuron, col = "blue", pch = 16, cex = 0.8)
grid()
```



### Calcu-

late ACF and PACF

```
par(mfrow = c(2, 1), mar = c(4, 4, 1, 1), las = 1)
acf(LakeHuron, xlab="Lag in years", ylab = "sample ACF", main = "")
pacf(LakeHuron, xlab="Lag in years", ylab = "sample PACF", main = "")
```



Estimate the linear trend

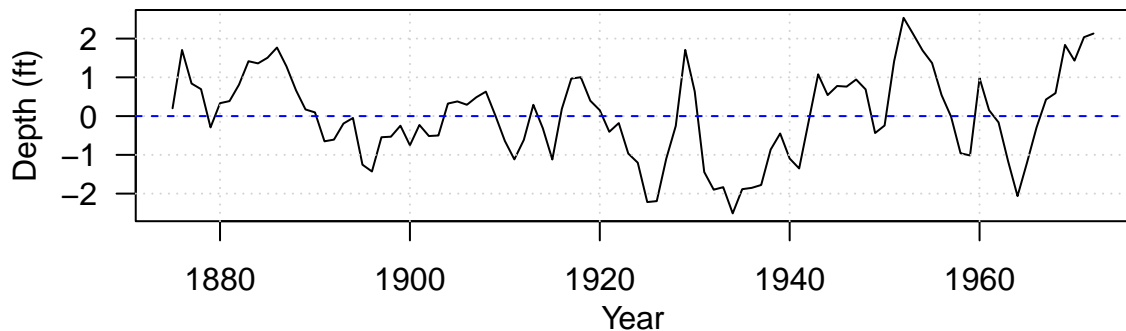
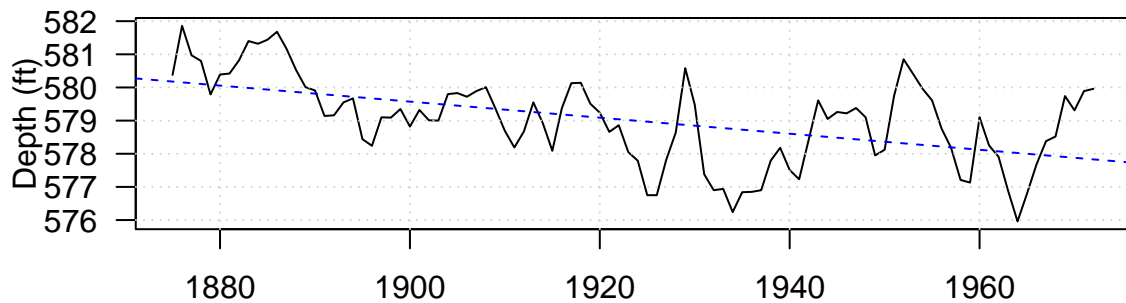
```
lm <- lm(LakeHuron ~ years)
par(mfrow = c(2, 1), mar = c(3.5, 3.5, 1, 0.6))
```



```

plot(LakeHuron, ylab = "", xlab = "", las = 1); grid()
abline(lm, col = "blue", lty = 2)
mtext("Depth (ft)", 2, line = 2.4)
deTrend <- resid(lm)
plot(1875:1972, deTrend, type = "l", ylab = "", xlab = "", las = 1); grid()
abline(h = 0, col = "blue", lty = 2)
mtext("Year", 1, line = 2)
mtext("Depth (ft)", 2, line = 2.4)

```



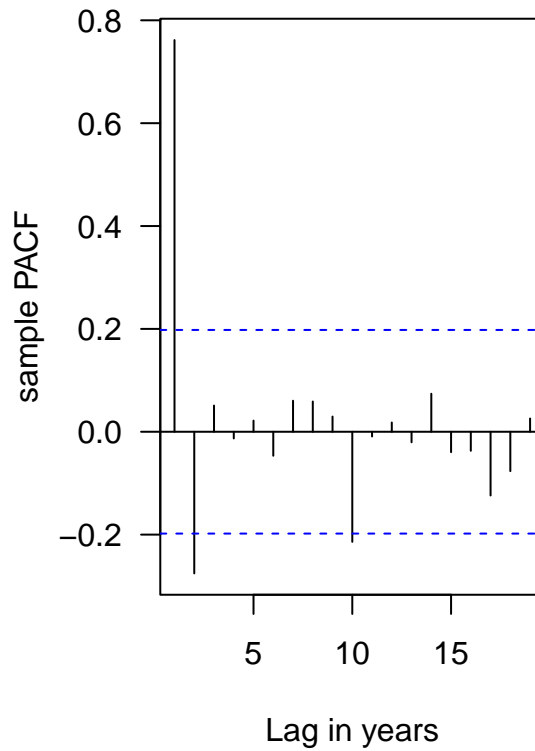
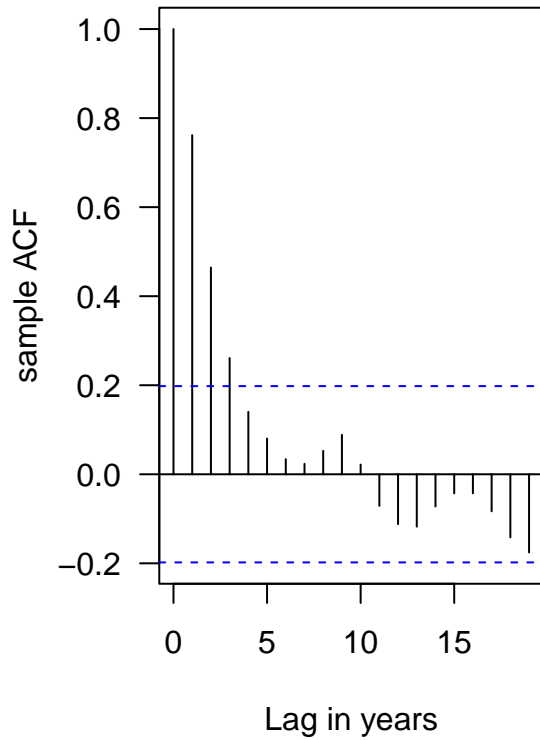
Recompute ACF and PACF for the detrended time series

##

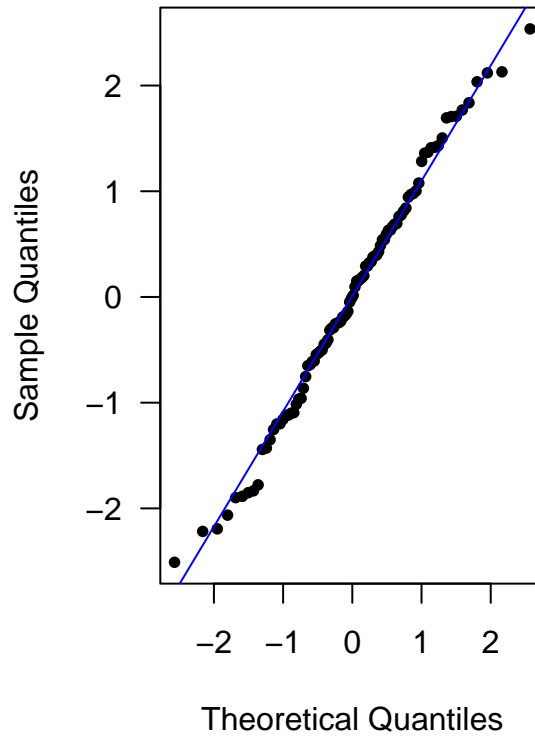
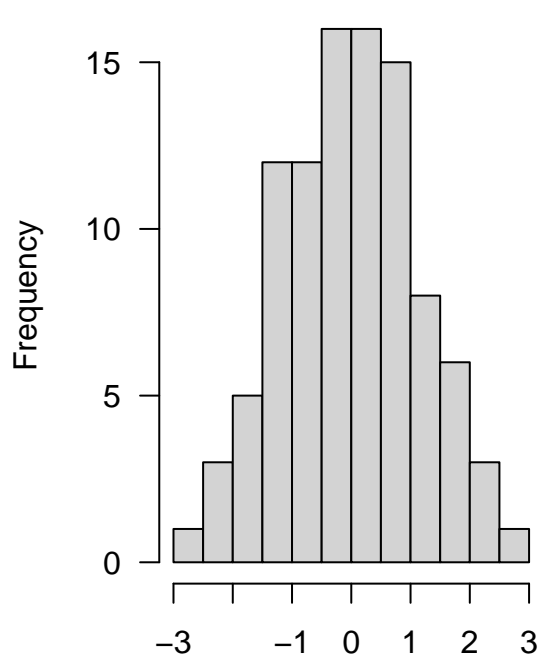
```

par(mfrow = c(1, 2), mar = c(4, 4, 1, 1), las = 1)
acf(deTrend, xlab = "Lag in years", ylab = "sample ACF", main = "")
pacf(deTrend, xlab = "Lag in years", ylab = "sample PACF", main = "")

```



```
# Histogram and QQ plot
hist(deTrend, 15, main = "", xlab = "", las = 1)
qqnorm(deTrend, main = "", pch = 16, cex = 0.8); qqline(deTrend, col = "blue")
```



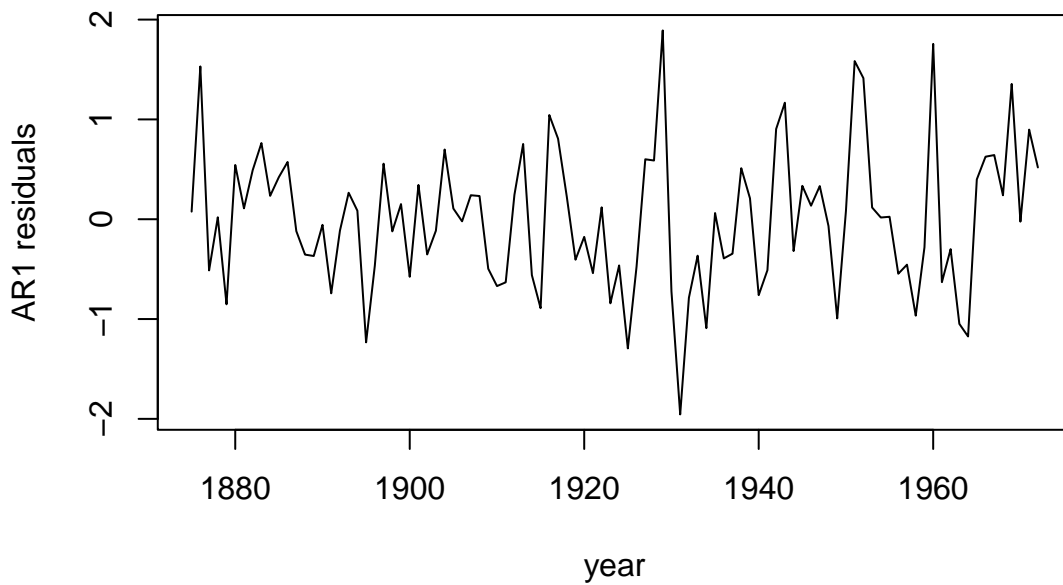
ARMA model fitting, selection, and diagnostics

```
(ar1.model <- arima(deTrend, order = c(1, 0, 0), method = "ML"))
```

## AR(1)

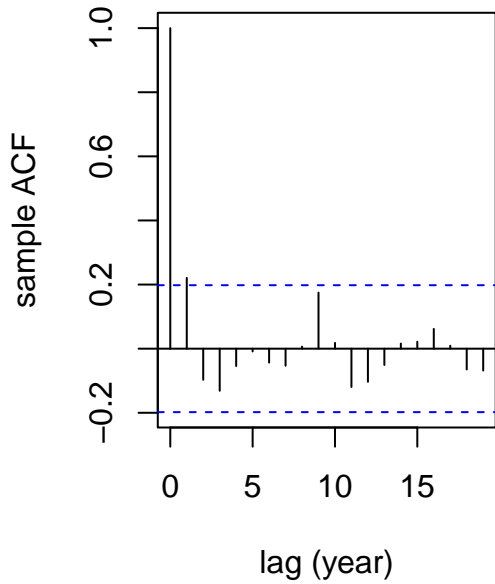
```
##  
## Call:  
## arima(x = deTrend, order = c(1, 0, 0), method = "ML")  
##  
## Coefficients:  
##      ar1  intercept  
##  0.7829    0.0799  
## s.e. 0.0634    0.3179  
##  
## sigma^2 estimated as 0.4972:  log likelihood = -105.29,  aic = 216.58
```

```
ar1.resids <- resid(ar1.model)  
plot(1875:1972, ar1.resids, type = "l", xlab = "year", ylab = "AR1 residuals")
```

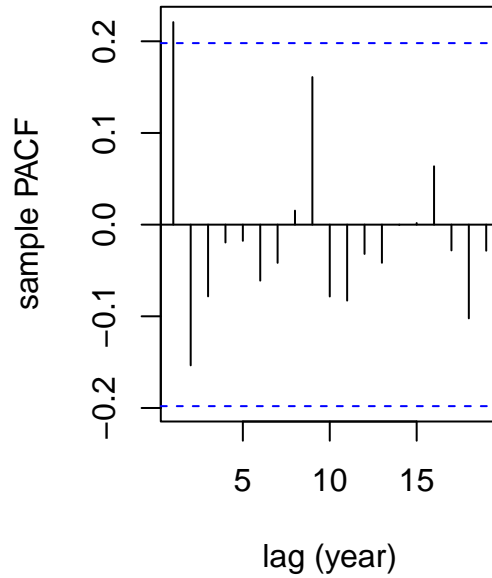


```
## Sample ACF and PACF of the residuals  
par(mfrow = c(1, 2))  
acf(ar1.resids, ylab = "sample ACF", xlab = "lag (year)")  
pacf(ar1.resids, ylab = "sample PACF", xlab = "lag (year)")
```

Series ar1.resids

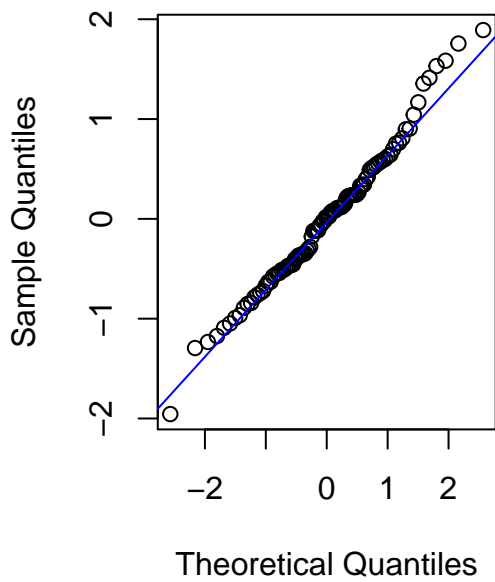


Series ar1.resids



```
## Normal Q-Q plot for the residuals
qqnorm(ar1.resids, main = ""); qqline(ar1.resids, col = "blue")
## Test for time dependence for the residuals
Box.test(ar1.resids, lag = 5, fitdf = 1, type = "Ljung-Box")
```

```
##
## Box-Ljung test
##
## data: ar1.resids
## X-squared = 7.9867, df = 4, p-value = 0.09207
```

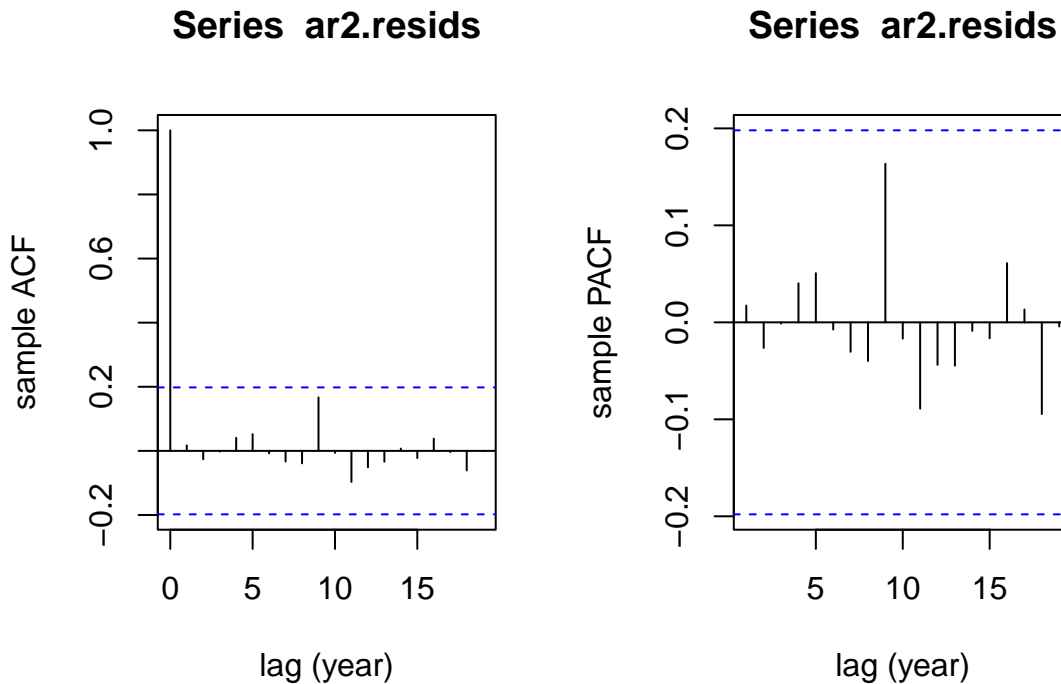


```
(ar2.model <- arima(deTrend, order = c(2, 0, 0), method = "ML"))
```

## AR(2)

```
##  
## Call:  
## arima(x = deTrend, order = c(2, 0, 0), method = "ML")  
##  
## Coefficients:  
##          ar1          ar2  intercept  
##      1.0047  -0.2919    0.0197  
## s.e.  0.0977   0.1004    0.2350  
##  
## sigma^2 estimated as 0.4571:  log likelihood = -101.25,  aic = 210.5
```

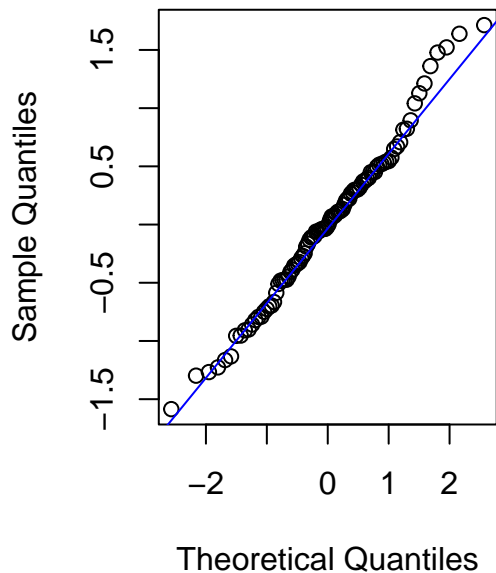
```
## calculate the residuals  
ar2.resids <- resid(ar2.model)  
## Sample ACF and PACF of the residuals  
par(mfrow = c(1, 2))  
acf(ar2.resids, ylab = "sample ACF", xlab = "lag (year)")  
pacf(ar2.resids, ylab = "sample PACF", xlab = "lag (year)")
```



```
## Test for time dependence for the residuals  
Box.test(ar2.resids, lag = 5, fitdf = 2, type = "Ljung-Box")
```

```
##  
## Box-Ljung test  
##  
## data:  ar2.resids  
## X-squared = 0.55962, df = 3, p-value = 0.9056
```

```
## Normal Q-Q plot for the residuals
qqnorm(ar2.resids, main = ""); qqline(ar2.resids, col = "blue")
```



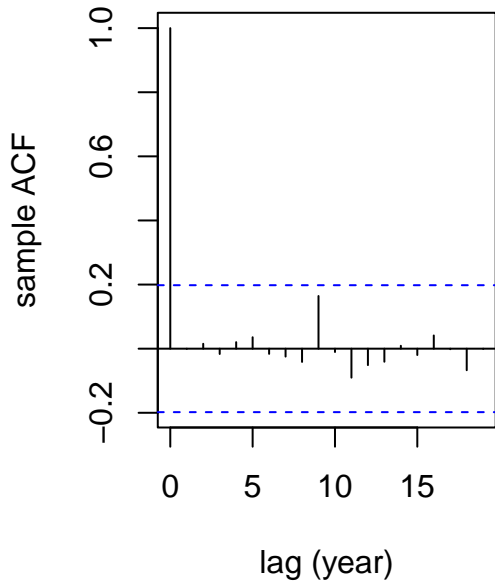
```
(arma21.model <- arima(deTrend, order = c(2, 0, 1)))
```

### ARMA(2, 1)

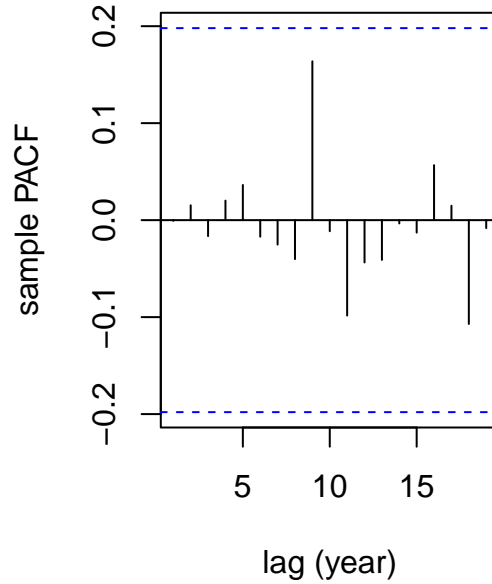
```
##
## Call:
## arima(x = deTrend, order = c(2, 0, 1))
##
## Coefficients:
##      ar1      ar2      ma1  intercept
##  0.8374 -0.1622  0.1846   0.0245
## s.e.  0.3180  0.2621  0.3180   0.2452
##
## sigma^2 estimated as 0.4556:  log likelihood = -101.09,  aic = 212.18
```

```
## calculate the residuals
arma21.resids <- resid(arma21.model)
## Sample ACF and PACF of the residuals
par(mfrow=c(1,2))
acf(arma21.resids, ylab = "sample ACF", xlab = "lag (year)")
pacf(arma21.resids, ylab = "sample PACF", xlab = "lag (year)")
```

Series arma21.resids

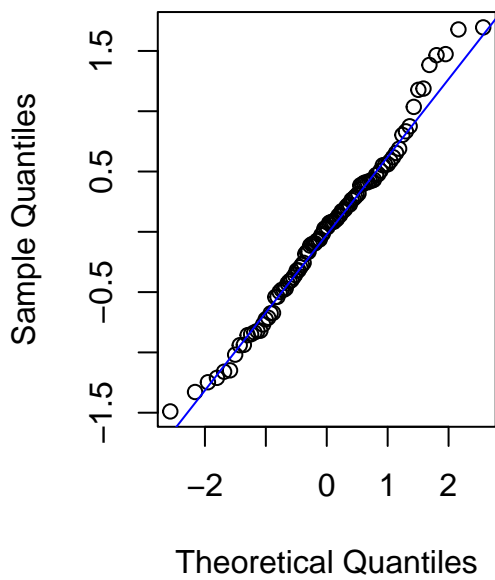


Series arma21.resids



```
## Normal Q-Q plot for the residuals
qqnorm(arma21.resids, main = ""); qqline(arma21.resids, col = "blue")
## Test
Box.test(arma21.resids, lag = 5, fitdf = 3, type = "Ljung-Box")
```

```
##
## Box-Ljung test
##
## data: arma21.resids
## X-squared = 0.2297, df = 2, p-value = 0.8915
```



```
AIC(ar1.model); AIC(ar2.model); AIC(arma21.model)
```

## Model selection using AIC

```
## [1] 216.5835
```

```
## [1] 210.5032
```

```
## [1] 212.1784
```

```
library(forecast)
auto.arima(deTrend, trace = T)
```

```
##
## ARIMA(2,0,2) with non-zero mean : 215.0455
## ARIMA(0,0,0) with non-zero mean : 304.222
## ARIMA(1,0,0) with non-zero mean : 216.8388
## ARIMA(0,0,1) with non-zero mean : 235.4585
## ARIMA(0,0,0) with zero mean      : 302.1373
## ARIMA(1,0,2) with non-zero mean : 212.7747
## ARIMA(0,0,2) with non-zero mean : 218.2478
## ARIMA(1,0,1) with non-zero mean : 210.9477
## ARIMA(2,0,1) with non-zero mean : 212.8306
## ARIMA(2,0,0) with non-zero mean : 210.9333
## ARIMA(3,0,0) with non-zero mean : 212.7787
## ARIMA(3,0,1) with non-zero mean : Inf
## ARIMA(2,0,0) with zero mean     : 208.7655
## ARIMA(1,0,0) with zero mean     : 214.7735
## ARIMA(3,0,0) with zero mean     : 210.569
## ARIMA(2,0,1) with zero mean     : 210.6186
## ARIMA(1,0,1) with zero mean     : 208.7891
## ARIMA(3,0,1) with zero mean     : Inf
##
## Best model: ARIMA(2,0,0) with zero mean
```

```
## Series: deTrend
## ARIMA(2,0,0) with zero mean
##
## Coefficients:
##          ar1      ar2
##          1.0050 -0.2925
## s.e.    0.0976  0.1002
##
## sigma^2 = 0.4667: log likelihood = -101.26
## AIC=208.51  AICc=208.77  BIC=216.27
```

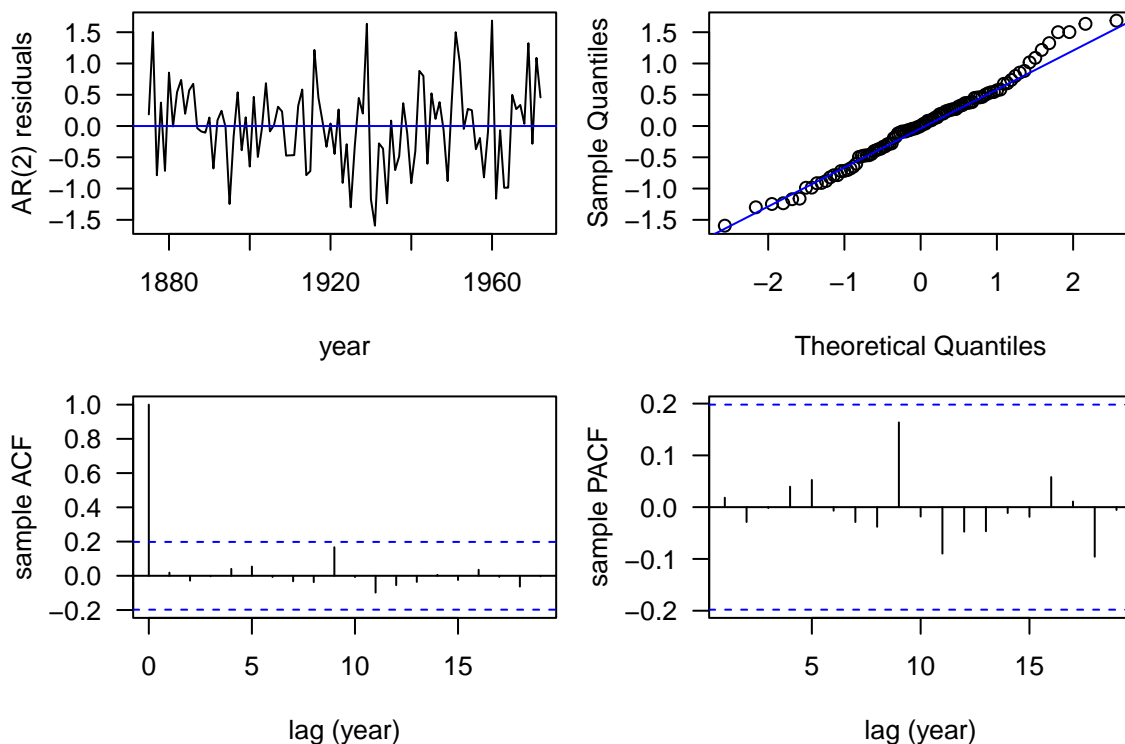
## AR(2) Fitting and Forecasting



```
(fit <- Arima(LakeHuron, order = c(2, 0, 0), include.drift = T))
```

```
## Series: LakeHuron
## ARIMA(2,0,0) with drift
##
## Coefficients:
##          ar1      ar2  intercept    drift
##          1.0048 -0.2913  580.0915 -0.0216
## s.e.      0.0976  0.1004    0.4636  0.0081
##
## sigma^2 = 0.476: log likelihood = -101.2
## AIC=212.4  AICc=213.05  BIC=225.32
```

```
par(mfrow = c(2, 2), mar = c(4.1, 4, 1, 0.8), las = 1)
res <- fit$residuals
plot(res, type = "l", xlab = "year", ylab = "AR(2) residuals", las = 1)
abline(h = 0, col = "blue")
qqnorm(res, main = ""); qqline(res, col = "blue")
acf(res, ylab = "sample ACF", xlab = "lag (year)")
pacf(res, ylab = "sample PACF", xlab = "lag (year)")
```



```
# 10-year-ahead forecasts
autoplot(forecast(fit, h = 10, level = c(50, 95)))
```

Forecasts from ARIMA(2,0,0) with drift

