DSA 8020 R Session 11: Time Series Analysis

Whitney

April 07, 2022

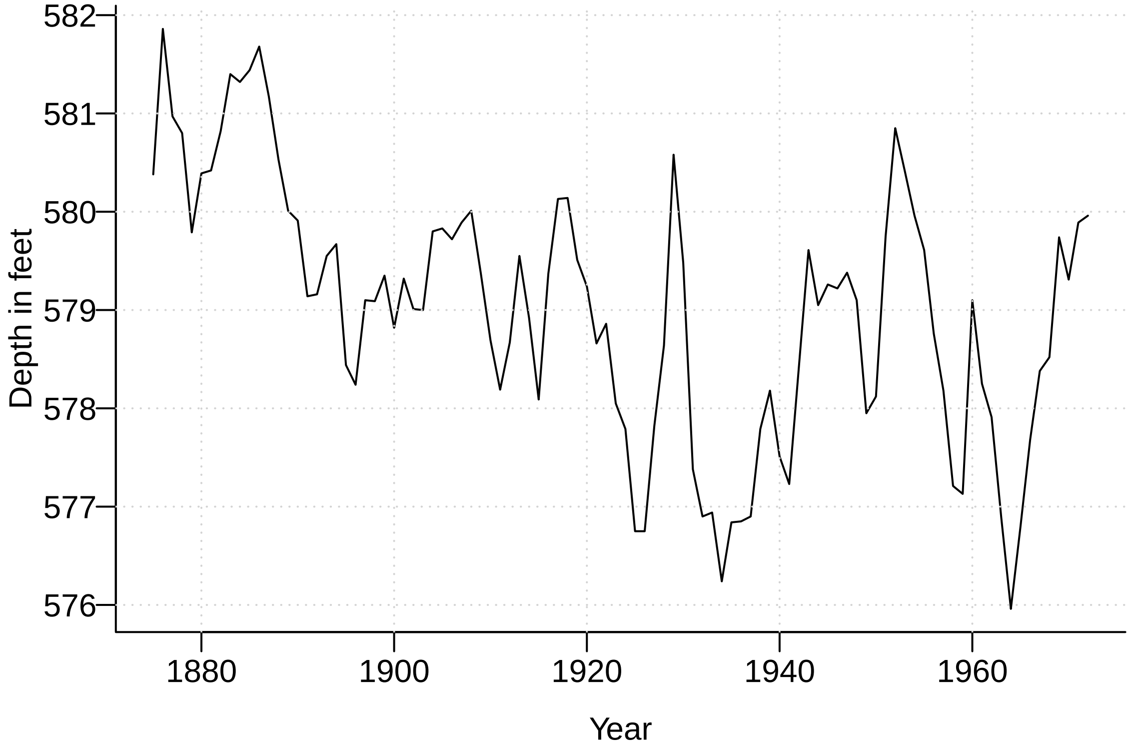
Contents

Time Series Data

Lake Huron Time Series

Annual measurements of the level of Lake Huron in feet

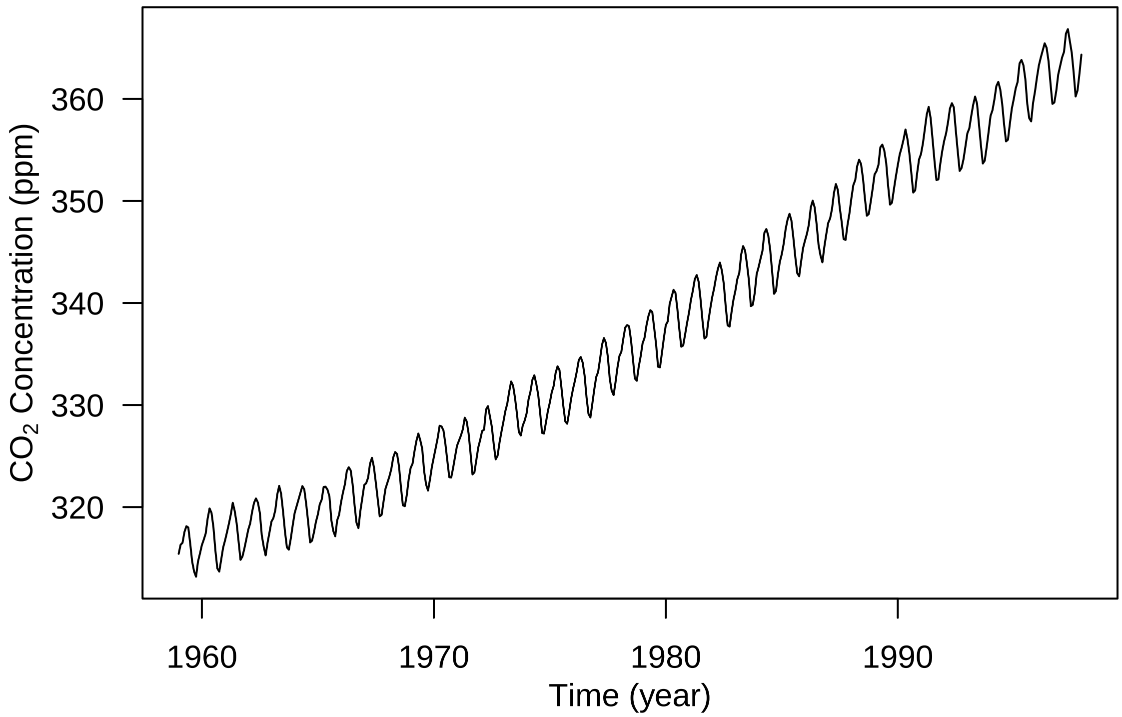
```
par(mar = c(3.2, 3.2, 0.5, 0.5), mgp = c(2, 0.5, 0), bty = "L")
data(LakeHuron)
plot(LakeHuron, ylab = "Depth in feet", xlab = "Year", las = 1)
grid()
```



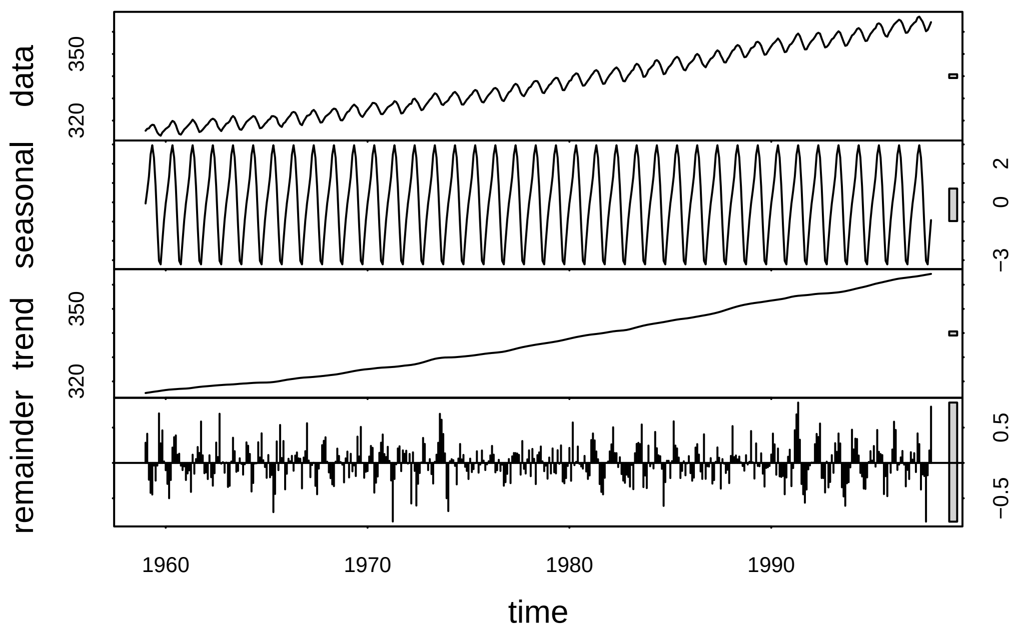
CO_2 Concentration

Atmospheric concentrations of CO2 are expressed in parts per million (ppm) and reported in 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 = "")
mtext("Time (year)", side = 1, line = 2)
mtext(expression(paste("CO"[2], " Concentration (ppm)")), side = 2, line = 2.5)
```




```r
# 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

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

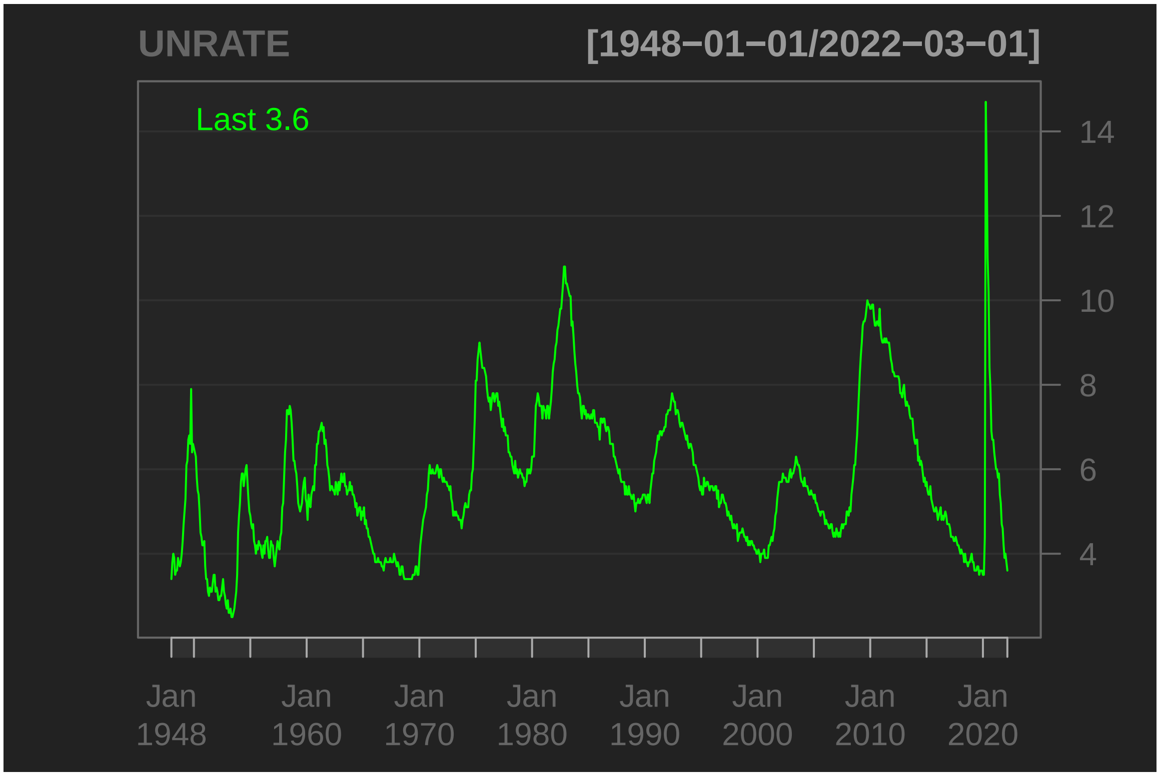
```
## [1] "UNRATE"
```

```r
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
```

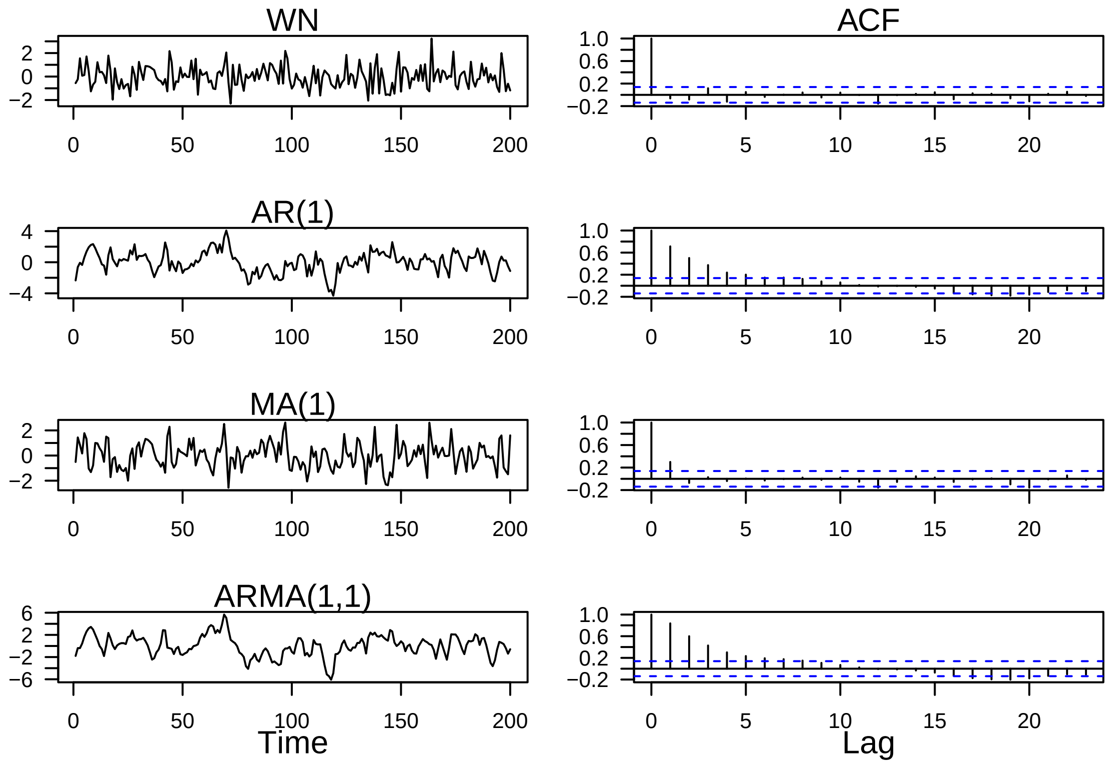
```
##            UNRATE
## 2021-10-01    4.6
## 2021-11-01    4.2
## 2021-12-01    3.9
## 2022-01-01    4.0
## 2022-02-01    3.8
## 2022-03-01    3.6
```

```r
chartSeries(UNRATE)
```

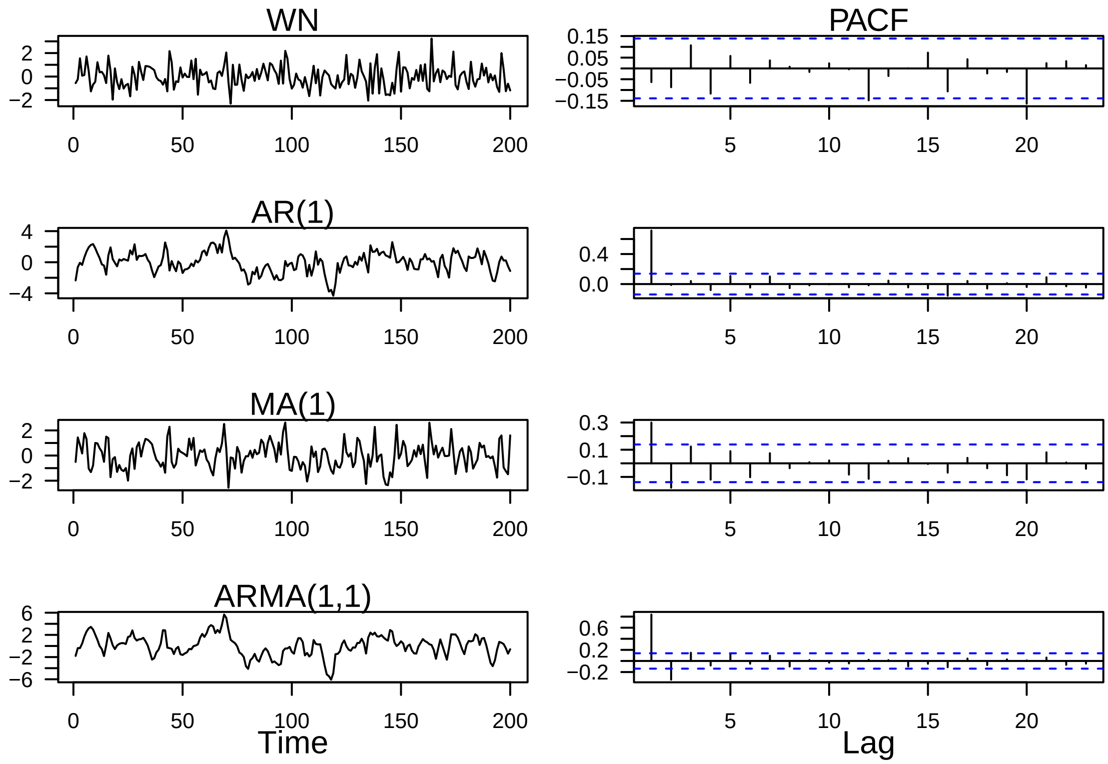


ARMA: ACF and PACF

```
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)
```



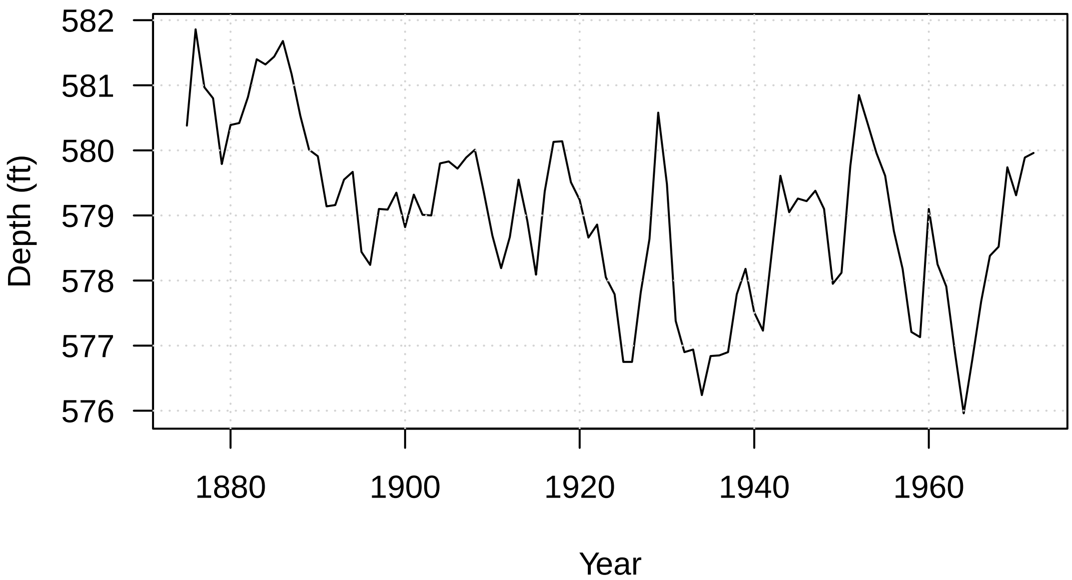

```r
par(mfrow = c(4, 2), mar = c(3.6, 3.6, 1.2, 0.6))
plot(1:n, WN, type = "l", las = 1, xlab = "", ylab = "")
mtext("WN")
pacf(WN, xlab = "", ylab = "", main = "", las = 1)
mtext("PACF")
# AR(1) phi = 0.8
set.seed(123)
AR <- arima.sim(n = n, model = list(ar = 0.8))
plot(1:n, AR, type = "l", las = 1, 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)
```



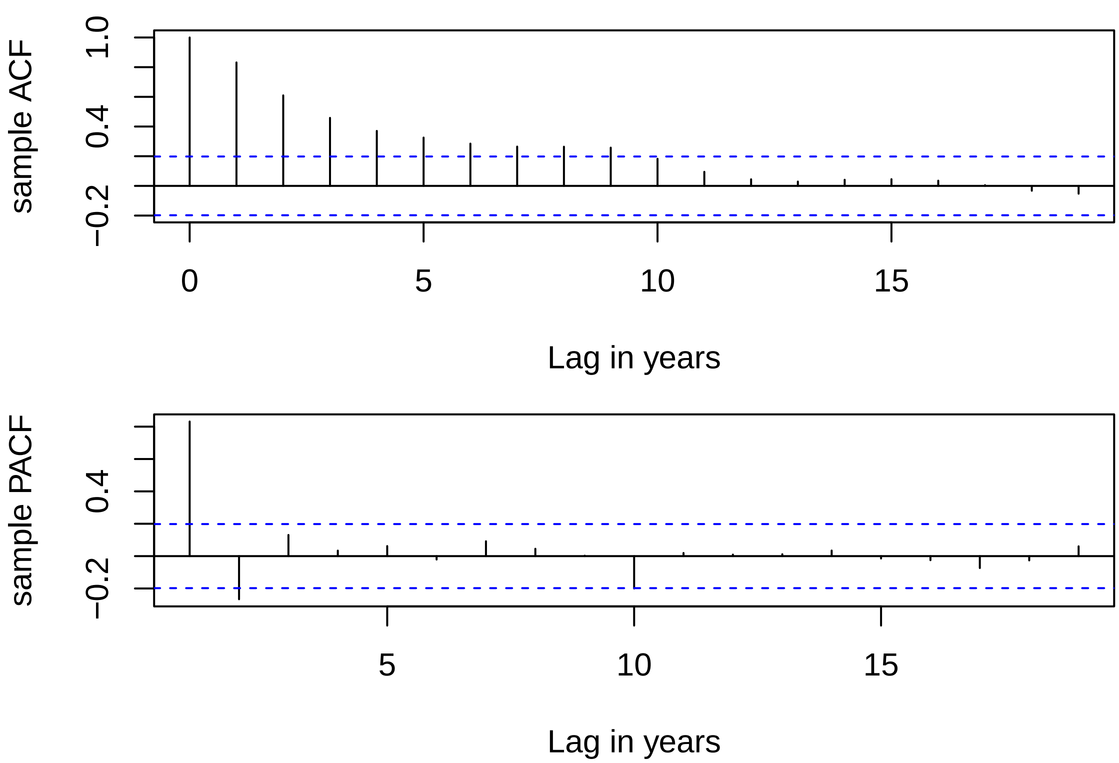
Lake Huron Case Study

Detrend

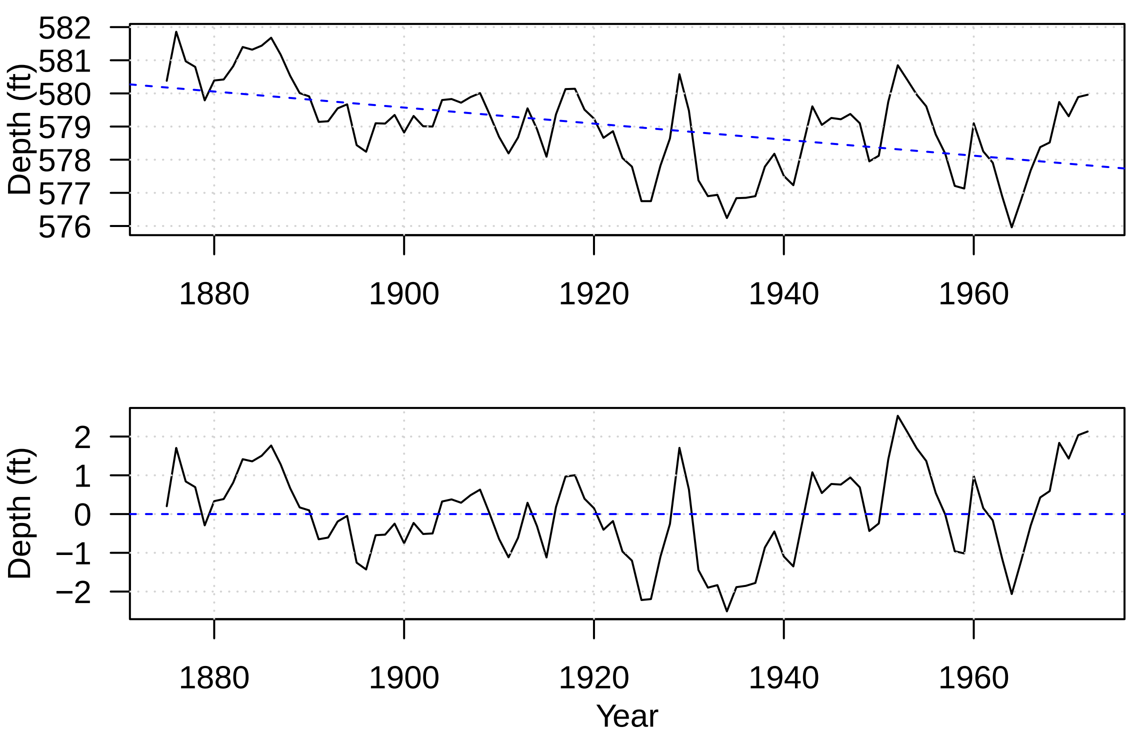
```
## Let us create a 'years' variable.
years <- time(LakeHuron)
## Plot time series
plot(LakeHuron, ylab = "Depth (ft)", xlab = "Year", las = 1)
grid()
```



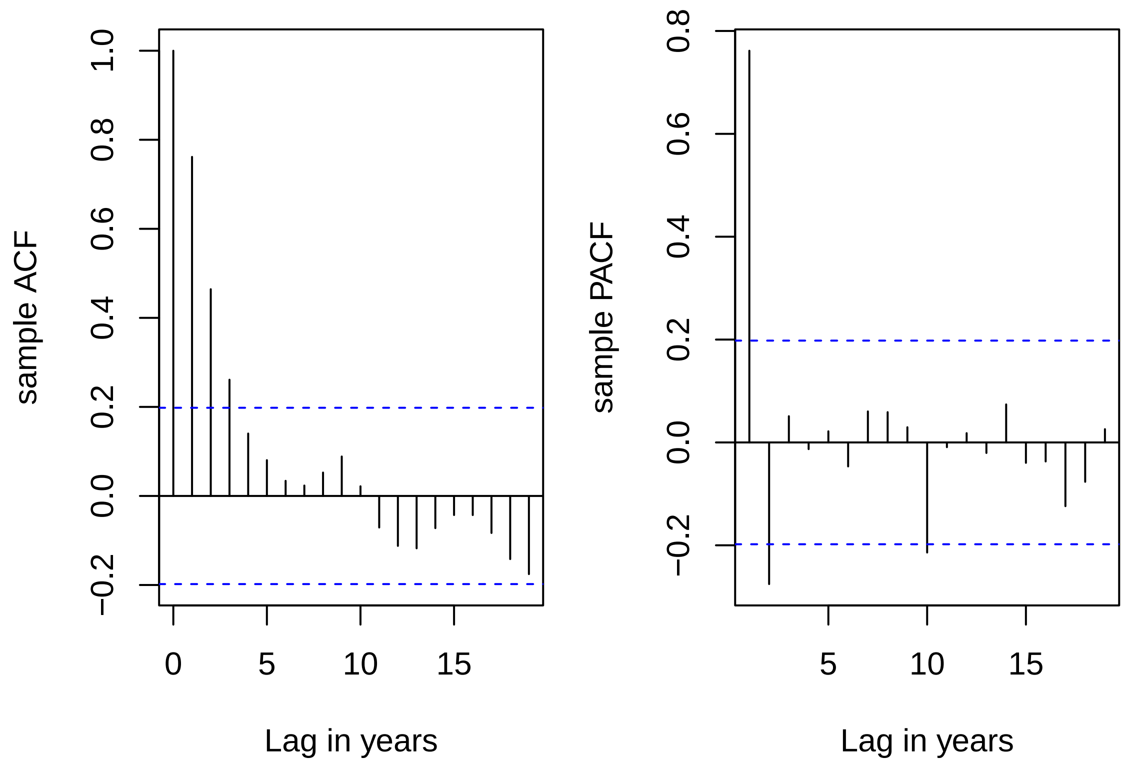

```r
## ACF and PACF
par(mfrow = c(2, 1), mar = c(4, 4, 1, 1))
acf(LakeHuron, xlab="Lag in years", ylab = "sample ACF", main = "")
pacf(LakeHuron, xlab="Lag in years", ylab = "sample PACF", main = "")
```



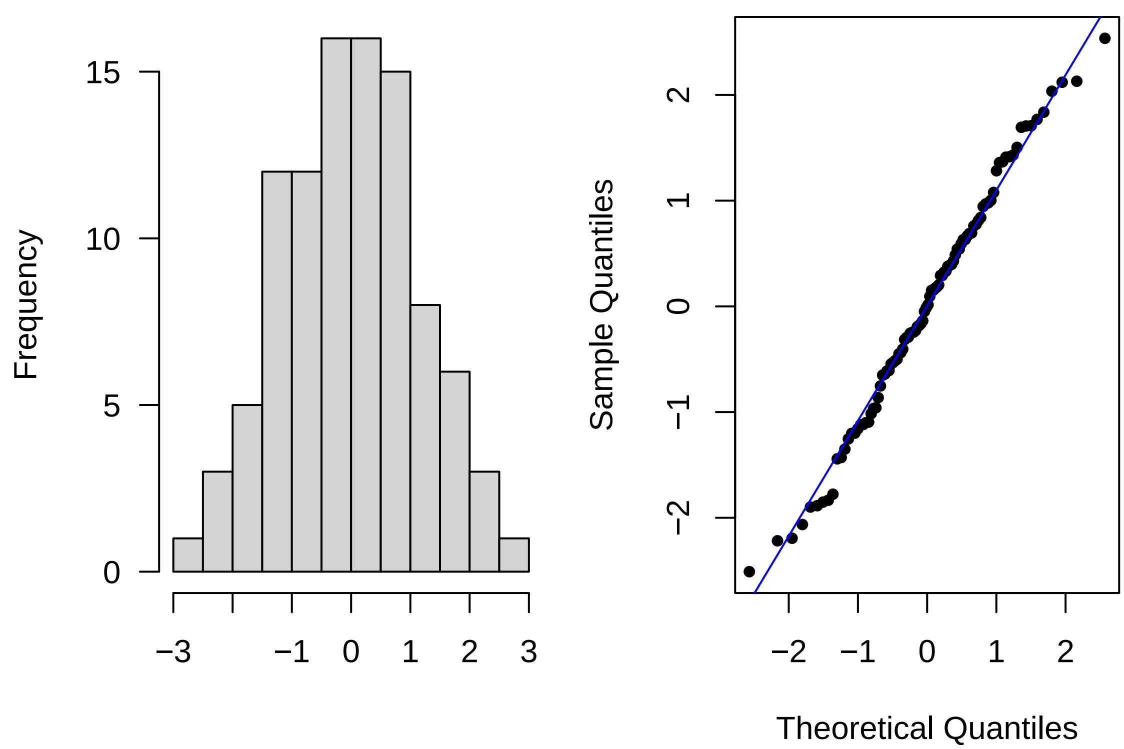
```r
# Estimate the linear trend
lm <- lm(LakeHuron ~ years)
par(mfrow = c(2, 1), mar = c(3.5, 3.5, 1, 0.6))
plot(LakeHuron, ylab = "", xlab = "", las = 1); grid()
abline(lm, col = "blue", lty = 2)
mtext("Depth (ft)", 2, line = 2.4)
deTrend <- resid(lm)
plot(1875:1972, deTrend, type = "l", ylab = "", xlab = "", las = 1); grid()
abline(h = 0, col = "blue", lty = 2)
mtext("Year", 1, line = 2)
mtext("Depth (ft)", 2, line = 2.4)
```




```r
## ACF and PACF
par(mfrow = c(1, 2), mar = c(4, 4, 1, 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, main = "", xlab = "", las = 1)
qqnorm(deTrend, main = "", pch = 16, cex = 0.8); qqline(deTrend, col = "blue")
```

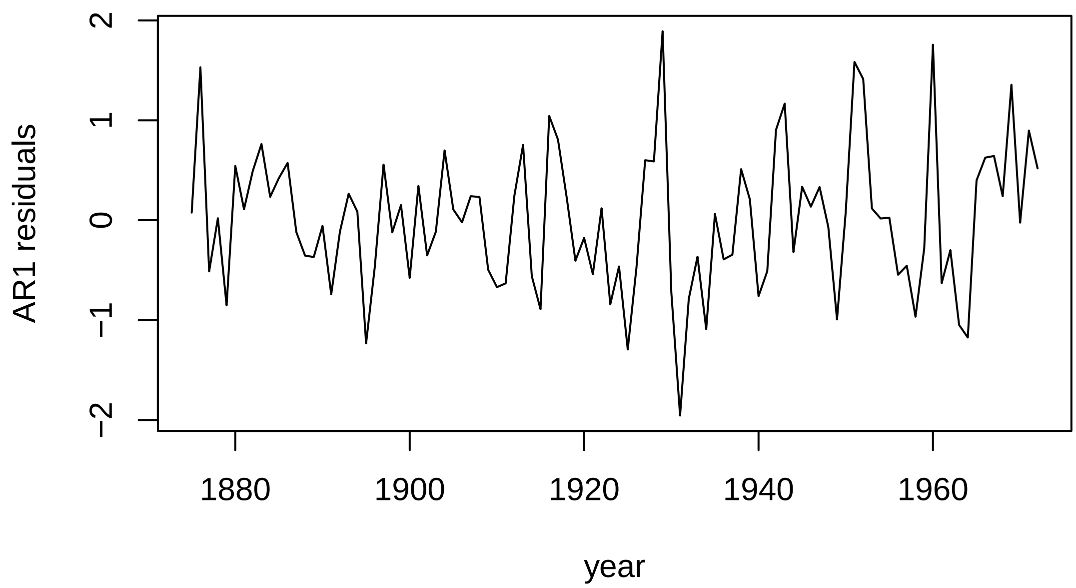


Model Selection/Fitting

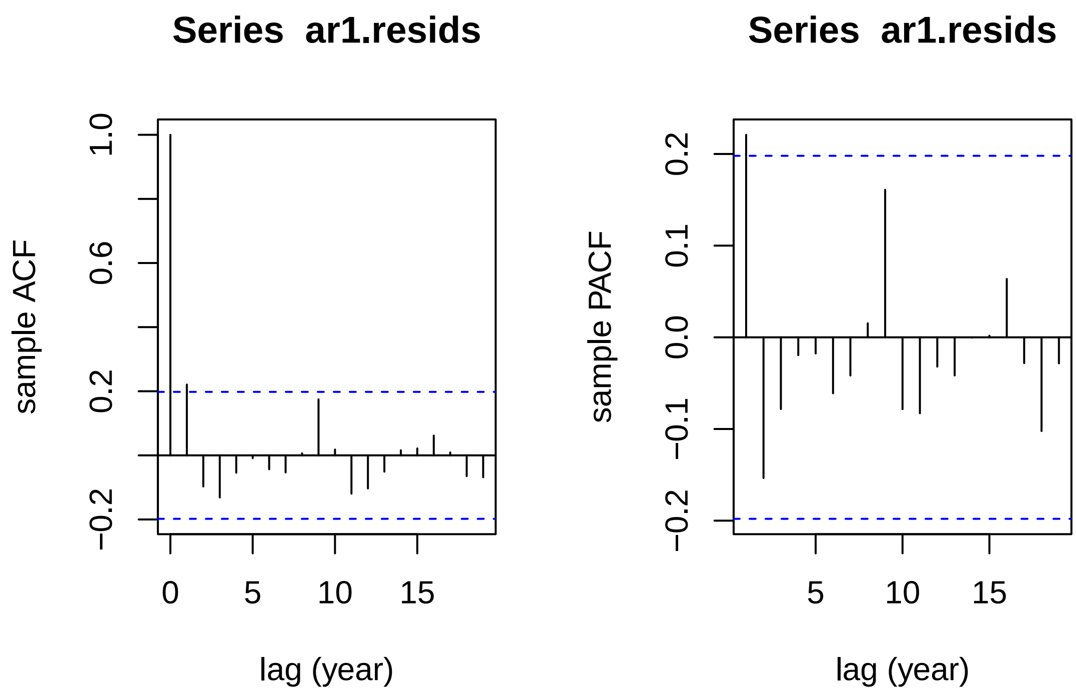
```r
## AR(1)
(ar1.model <- arima(deTrend, order = c(1, 0, 0)))
```

```
##
## Call:
## arima(x = deTrend, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.7829     0.0797
## s.e.  0.0634     0.3178
##
## sigma^2 estimated as 0.4972:  log likelihood = -105.29,  aic = 216.58
```

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




```r
## 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)")
```




```r
## 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, type = "Ljung-Box")
```

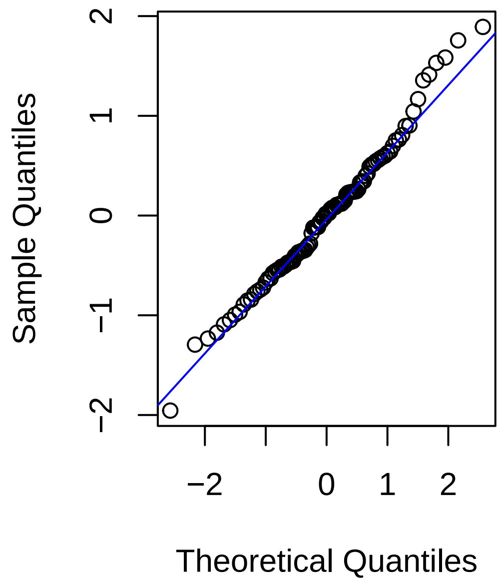
```
##
##  Box-Ljung test
##
## data:  ar1.resids
## X-squared = 4.93, df = 1, p-value = 0.02639
```

```r
## AR(2)
(ar2.model <- arima(deTrend, order = c(2, 0, 0)))
```

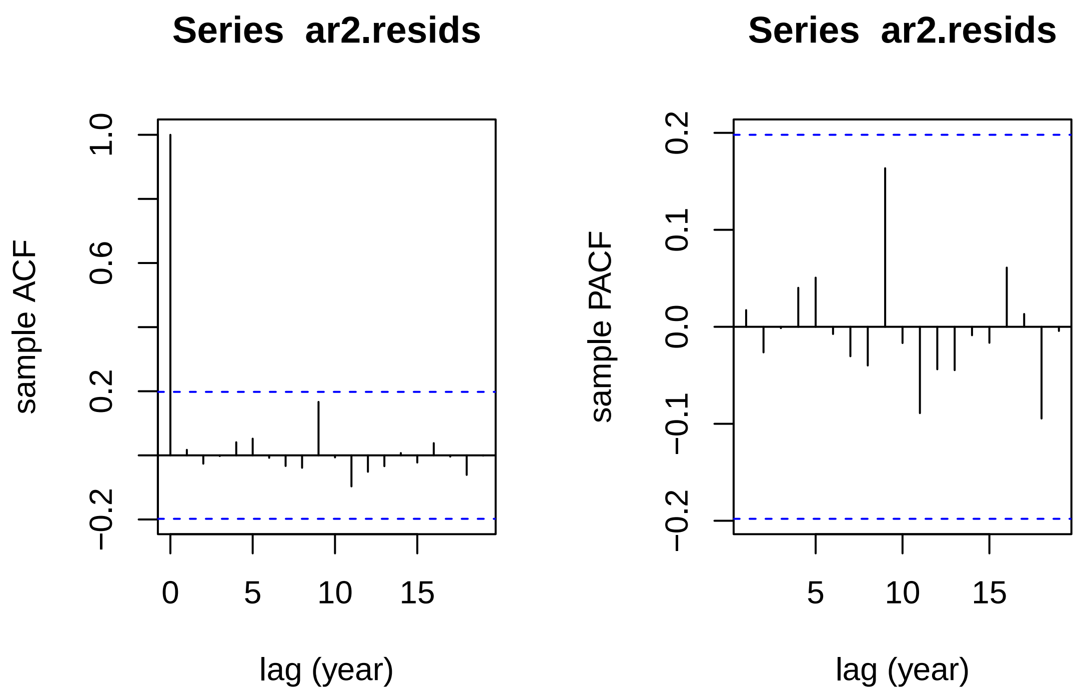
```
##
## Call:
## arima(x = deTrend, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       1.0047  -0.2919     0.0196
## s.e.  0.0977   0.1004     0.2351
##
## sigma^2 estimated as 0.4571:  log likelihood = -101.25,  aic = 210.5
```

```r
## calculate the residuals
ar2.resids <- resid(ar2.model)
```

```r
## Sample ACF and PACF of the residuals
par(mfrow = c(1, 2))
```




```r
acf(ar2.resids,  ylab = "sample ACF", xlab = "lag (year)")
pacf(ar2.resids,  ylab = "sample PACF", xlab = "lag (year)")
```



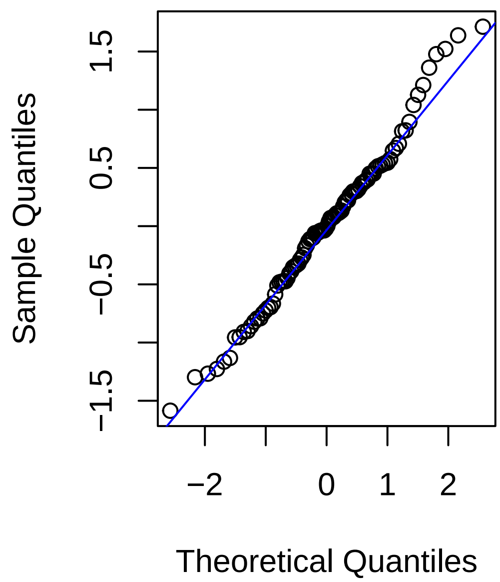

```r
## Test for time dependence for the residuals
Box.test(ar2.resids, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  ar2.resids
## X-squared = 0.029966, df = 1, p-value = 0.8626
```

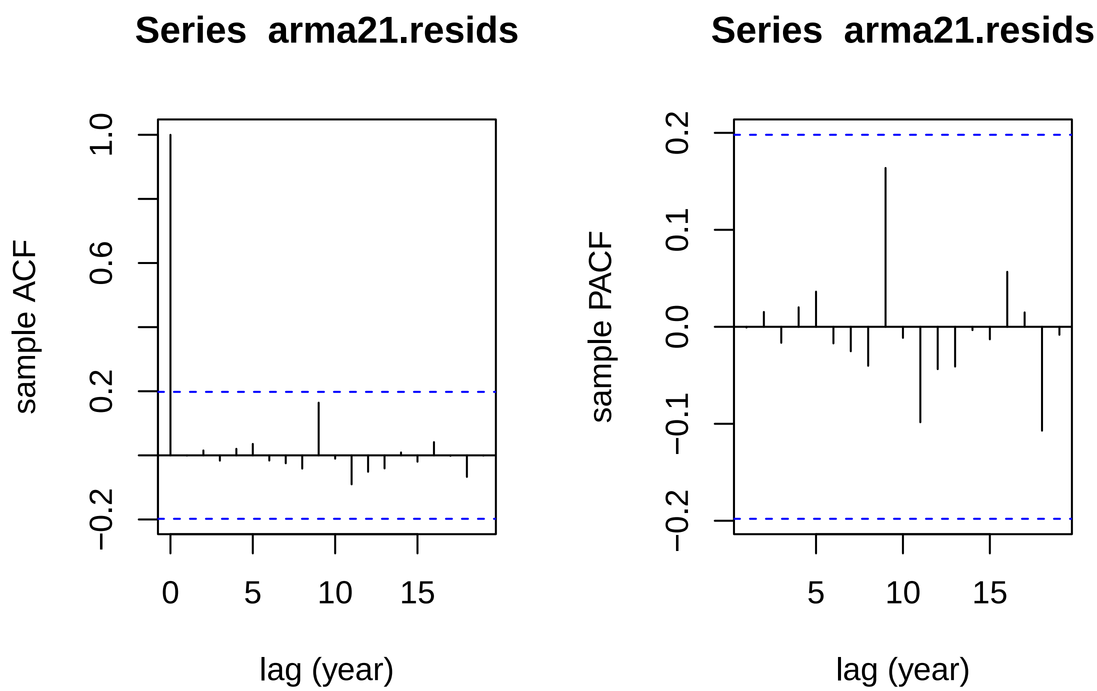
```r
## Normal Q-Q plot for the residuals
qqnorm(ar2.resids, main = ""); qqline(ar2.resids, col = "blue")

## Fit the ARMA(2, 1) model
(arma21.model <- arima(deTrend, order = c(2, 0, 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)")
```




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

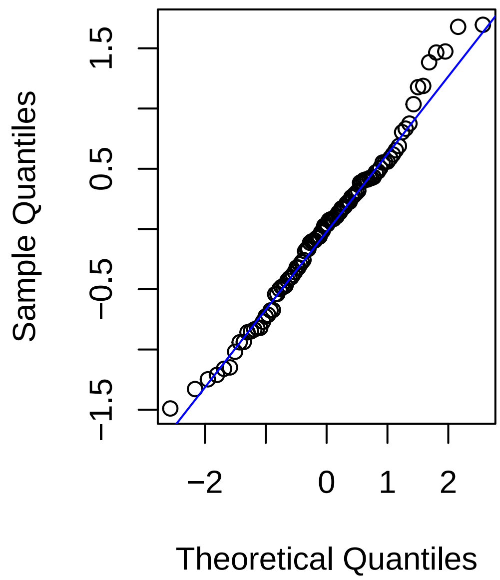
```
##
##  Box-Ljung test
##
## data:  arma21.resids
## X-squared = 5.5105e-05, df = 1, p-value = 0.9941
```

```
# Model selection using AIC
AIC(ar1.model); AIC(ar2.model); AIC(arma21.model)
```

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

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

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

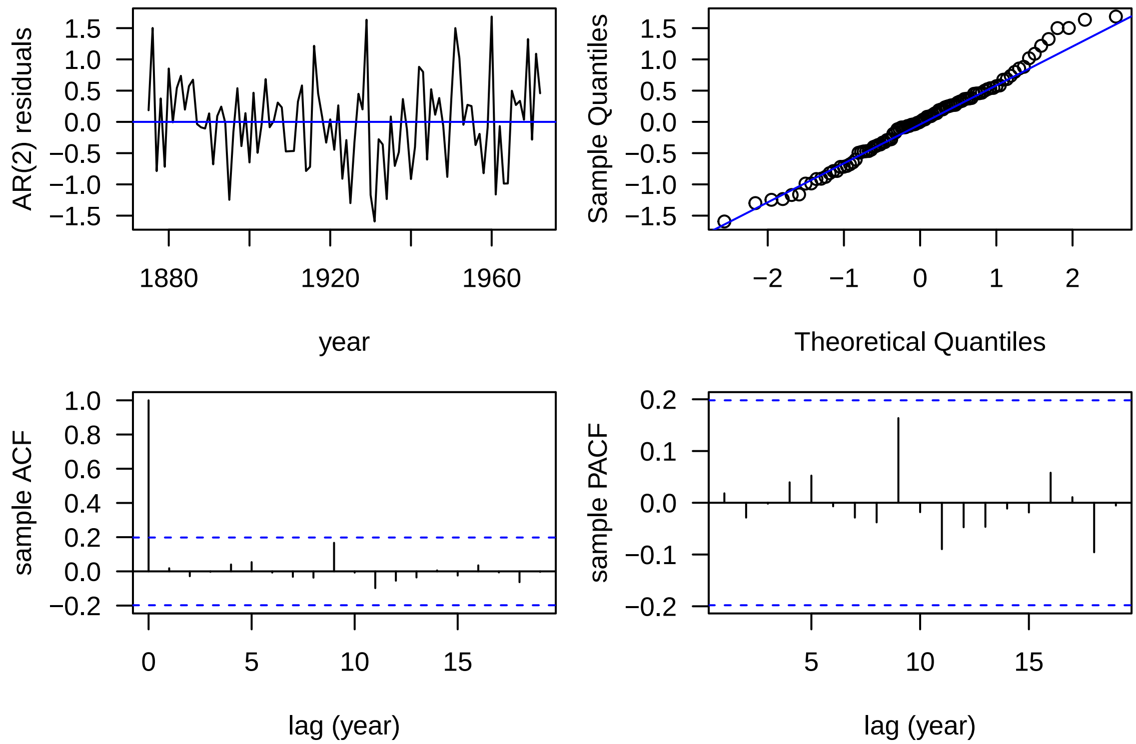


AR(2) Fitting and Forecasting

```
library(forecast)
(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)")
```




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