MATH1318 Time Series Analysis Assignment 2

Ashleigh Olney

09/05/2019

Contents

Import Data

```r
# read in data
eggs <- read_csv("eggs.csv")
```

```
## Parsed with column specification:
## cols(
##   year = col_integer(),
##   eggs = col_double()
## )
```

```r
# check dimensions
dim(eggs)
```

```
## [1] 16  2
```

```r
# remove extra column (years)
eggs <- eggs[,-(1)]

# convert to time series object and re-specify year labels
eggs.ts <- ts(as.vector(eggs), start = 1981, end = 1996)

# re-check dimensions
dim(eggs)
```

```
## [1] 16  1
```

Inspect Data

Plot

```r
plot(eggs.ts, type = 'o',
     ylab = "egg depositions (millions)",
     sub = "Figure 1. Lake Huron Bloaters Eggs",
     font.sub = "2",
     main = NULL)
```

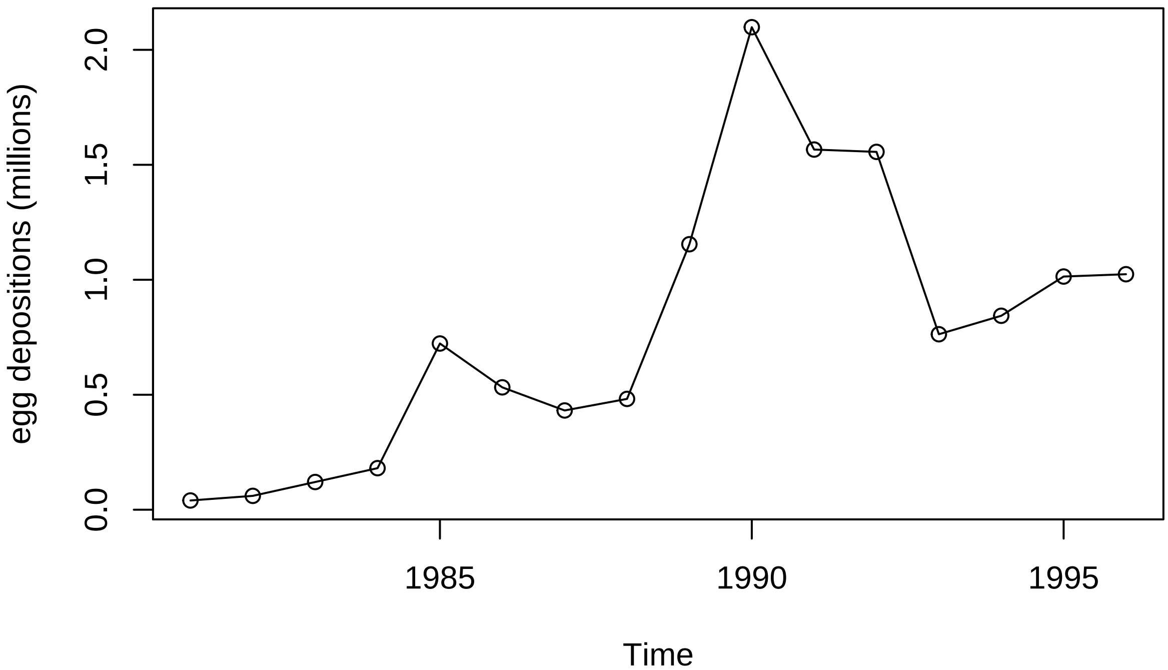


Figure 1. Lake Huron Bloaters Eggs

- There appears to be a slight upward trend
- There is no seasonality
- The distribution of the data appears to be a mix of autoregressive and moving average
- There may be an intonation point at 1989 & 1990 but the data regresses back to the slight upward trend

Test for Normality

```
qqnorm(eggs.ts,
       sub = "Figure 2. Lake Huron Bloaters Eggs Normal Q-Q Plot",
       font.sub = "2",
       main = NULL)

qqline(eggs.ts, col = 2)
```

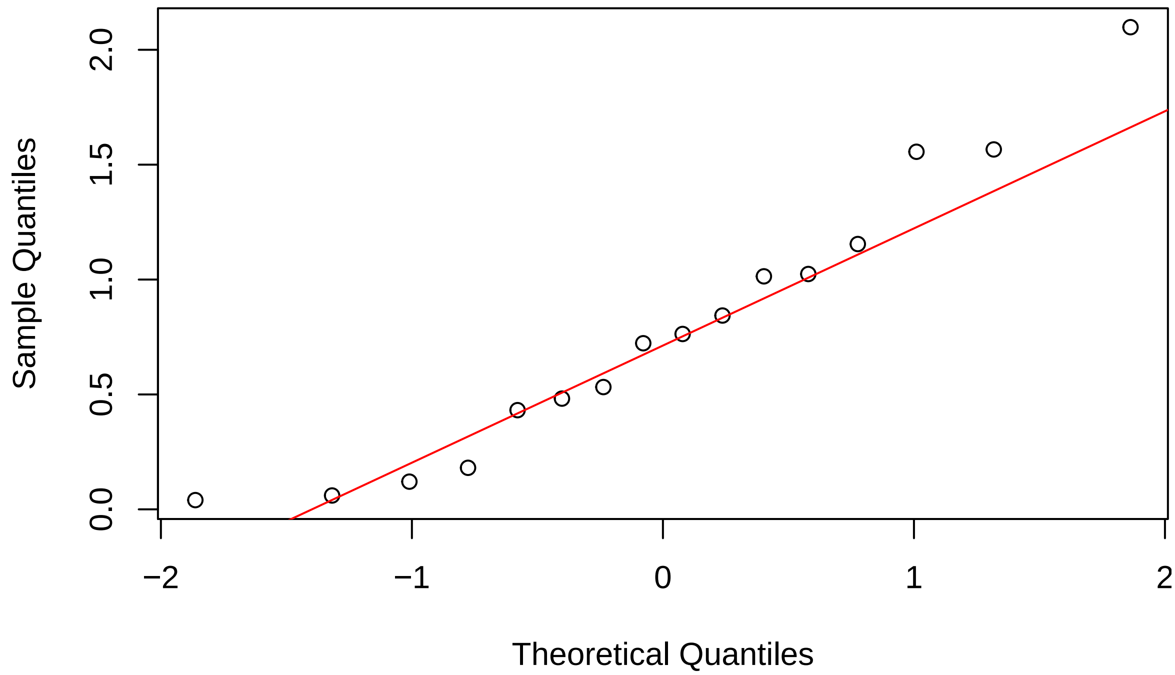


Figure 2. Lake Huron Bloaters Eggs Normal Q−Q Plot

```
shapiro.test(eggs.ts)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  eggs.ts
## W = 0.94201, p-value = 0.3744
```

Most of the data points sit quite close, if not along, the line in figure 2. The Shapiro-Wilk normality test returns a p-value of 0.3744, which means we fail to reject the null hypothesis that the series is normal.

ACF & PACF

```
acf(eggs.ts,
    sub = "Figure 3. Lake Huron Bloaters Eggs ACF",
    font.sub = "2",
    main = "")
```

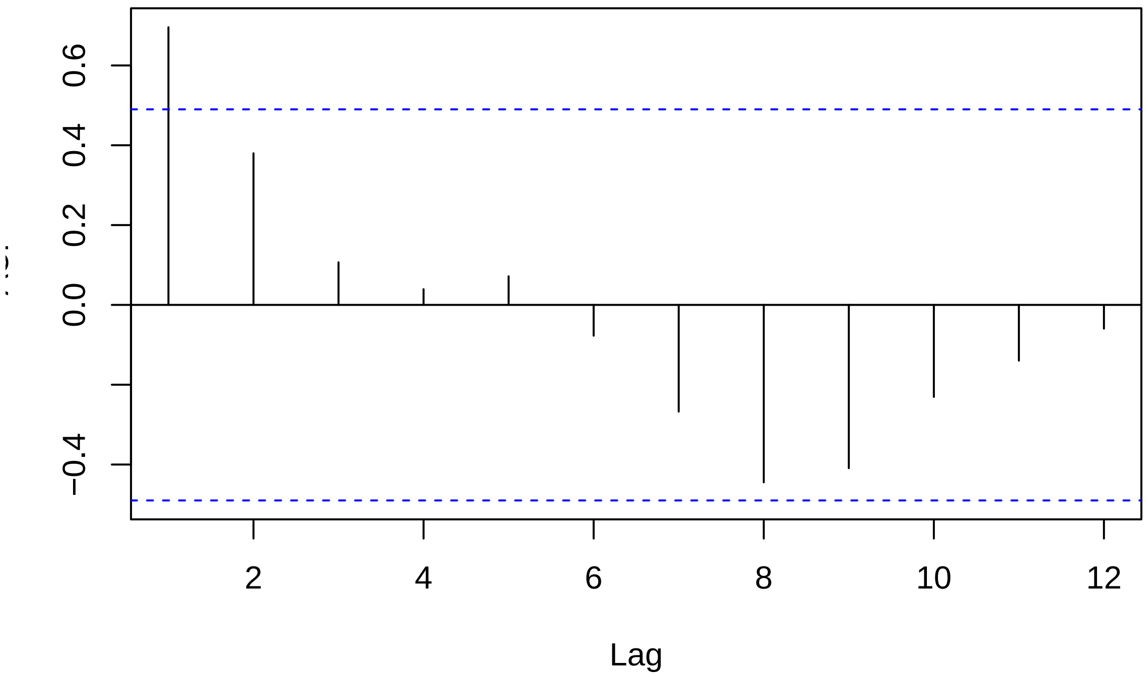


Figure 3. Lake Huron Bloaters Eggs ACF

Figure 3 shows one significant ACF lag (at lag 1). There is not a slowly decaying pattern. This graph suggests a MA(1) process.

```
pacf(eggs.ts,
     sub = "Figure 4. Lake Huron Bloaters Eggs PACF",
    font.sub = "2",
    main = "")
```

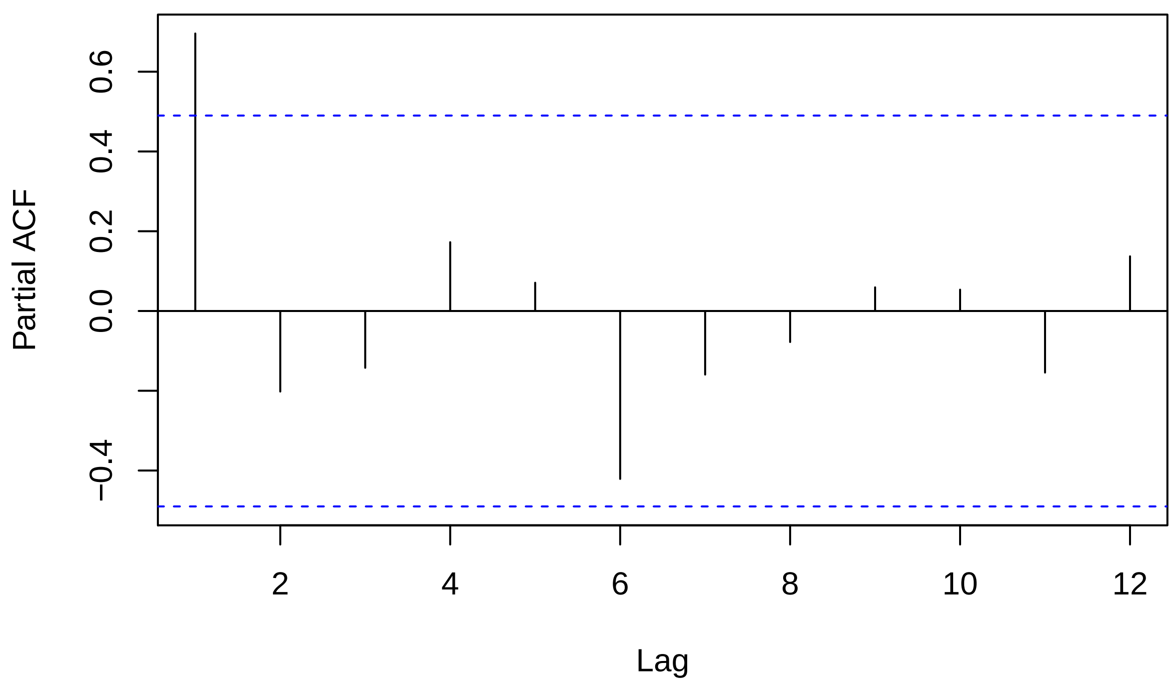


Figure 4. Lake Huron Bloaters Eggs PACF

Figure 4 shows there is one significant PACF lag (at lag 1). This graph suggests an AR(1) process.

Yt vs Yt-1

```r
plot(y = eggs.ts, x = zlag(eggs.ts), ylab = "Yt", xlab = "Yt-1",
     sub = "Figure 5. Correlation of First Difference",
     font.sub = "2", main = NULL)
```

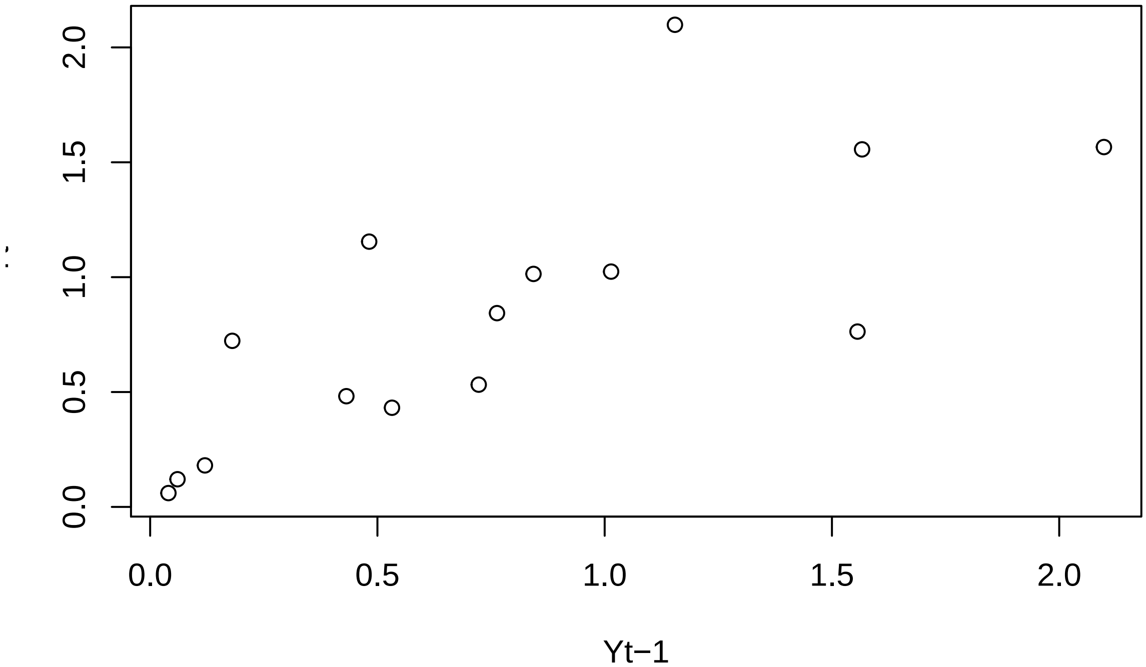


Figure 5. Correlation of First Difference

```r
x <- zlag(eggs.ts)
index <- 2:length(x)
cor(eggs.ts[index],x[index])
```

```
## [1] 0.7445657
```

There is a moderately strong positive correlation between Yt and Yt-1 of 0.74. This suggests that there is a relationship between one years data and the next.

ADF Test for Stationarity

```r
order <- ar(diff(eggs.ts))$order
adfTest(eggs.ts)
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 1
##   STATISTIC:
##     Dickey-Fuller: -0.5115
##   P VALUE:
##     0.4455
##
```

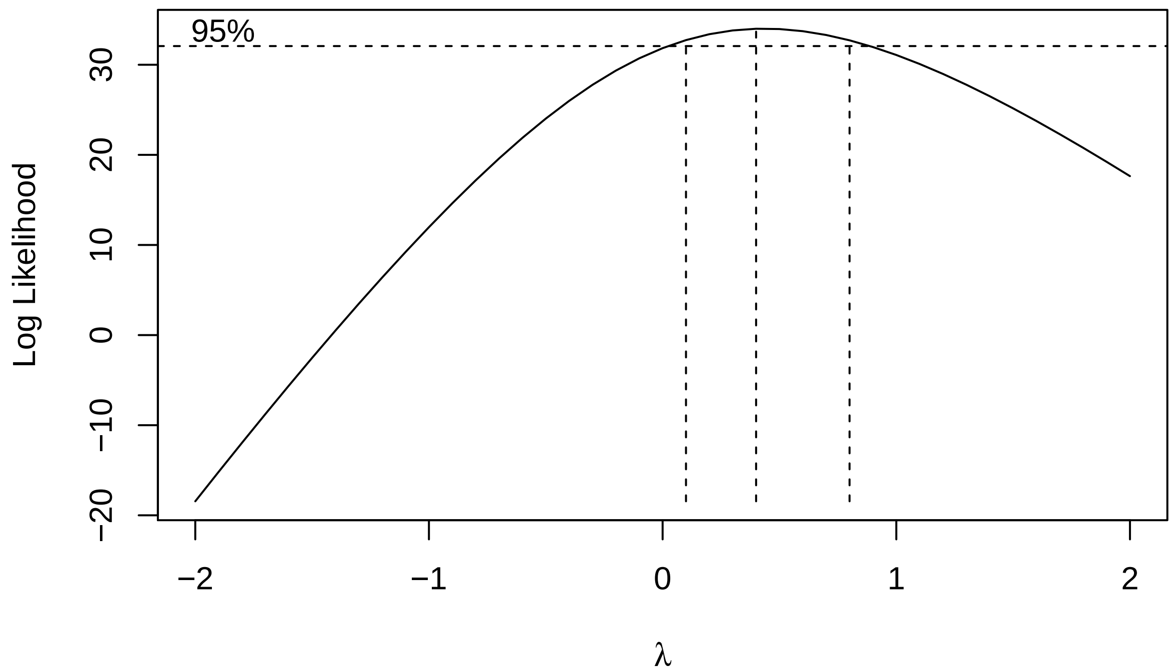
```
## Description:
##  Sun May 12 19:23:21 2019 by user:
```

The ADF test confirms that the time series is not stationary (p value > 0.05). In order to determine an ARIMA model for this egg data, the series must first be transformed into a stationary series through transformation and/or differencing.

Transform to Stationary

Box Cox Transformation

```
# find lambda
eggs.transform <- BoxCox.ar(eggs.ts, method = "yule-walker",
                           sub = "Figure 6. Lambda Log-Liklihood")
```




```
eggs.transform$ci
```

```
## [1] 0.1 0.8
```

```
lambda <- 0.5

# transform data
BC.eggs <- (eggs.ts^lambda-1)/lambda

# plot transformed data
plot(BC.eggs, type = 'o',
     ylab = "egg depositions (millions)",
     sub = "Figure 7. Box Cox Transformed",
     font.sub = "2", main = NULL)
```

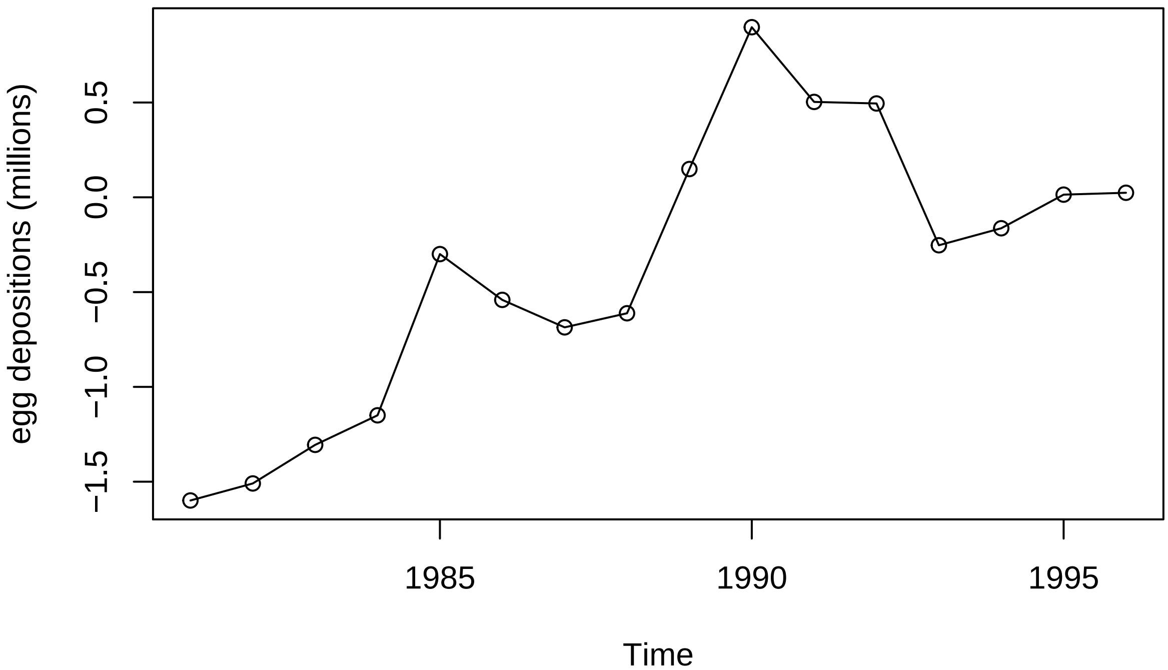


Figure 7. Box Cox Transformed

Based on a visual assessment of the transformed data in figure 7, the series does not appear to be stationary as the distribution is largely similar to the original data.

```
# test for stationarity
order <- ar(diff(BC.eggs))$order
adfTest(BC.eggs, lags = order)
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 0
##   STATISTIC:
##     Dickey-Fuller: -2.1242
##   P VALUE:
##     0.03595
##
## Description:
##  Sun May 12 19:23:23 2019 by user:
```

The ADF test p-value is 0.035, which is below alpha (0.05) which would cause the null hypothesis of non-stationary to be rejected. However, based on a visual assessment of figure 7, the series does not appear to be stationary so we will proceed with differencing.

```
# test for normality
qqnorm(BC.eggs, sub = "Figure 8. Box-Cox Transformed Normal Q-Q Plot",
       font.sub = "2", main = NULL)
qqline(BC.eggs, col = 2)
```

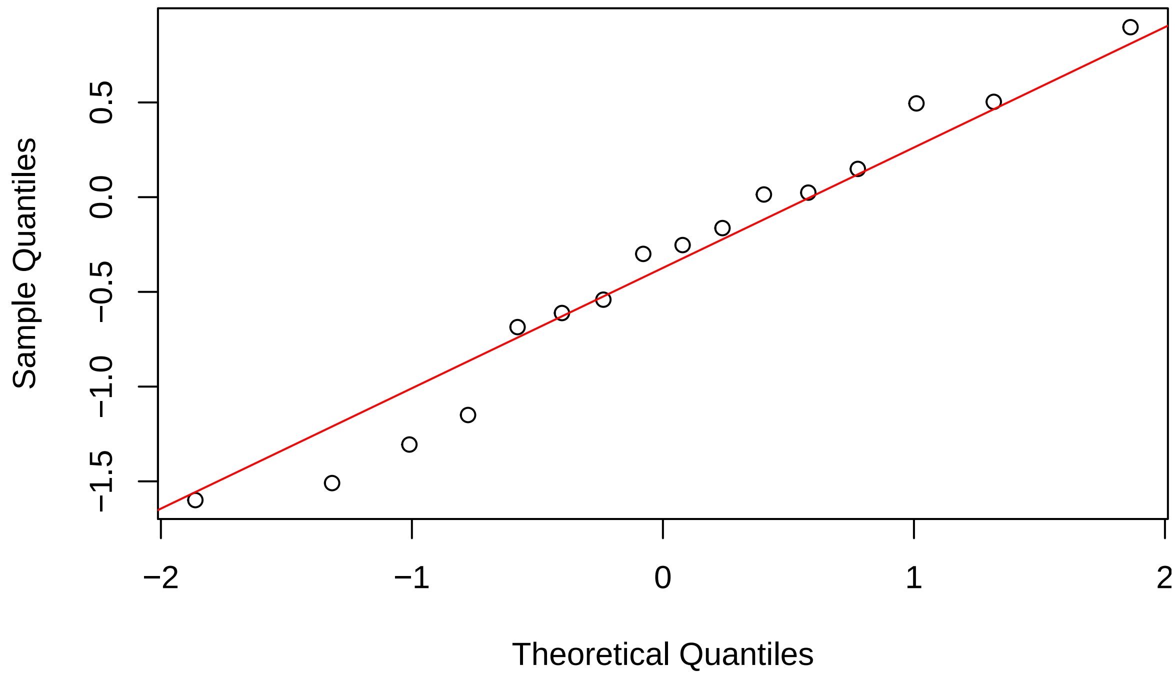


Figure 8. Box−Cox Transformed Normal Q−Q Plot

```
shapiro.test(BC.eggs)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  BC.eggs
## W = 0.96562, p-value = 0.7636
```

The QQ plot and Shapiro-Wilk test show that the data continues to be normal.

First Difference

```
eggsDiff1 <- diff(BC.eggs)
plot(eggsDiff1, type = 'o',
      ylab = "egg depositions (millions)",
     sub = "Figure 9. First Difference of Box Cox Transformed Eggs",
     font.sub = "2", main = NULL)
```

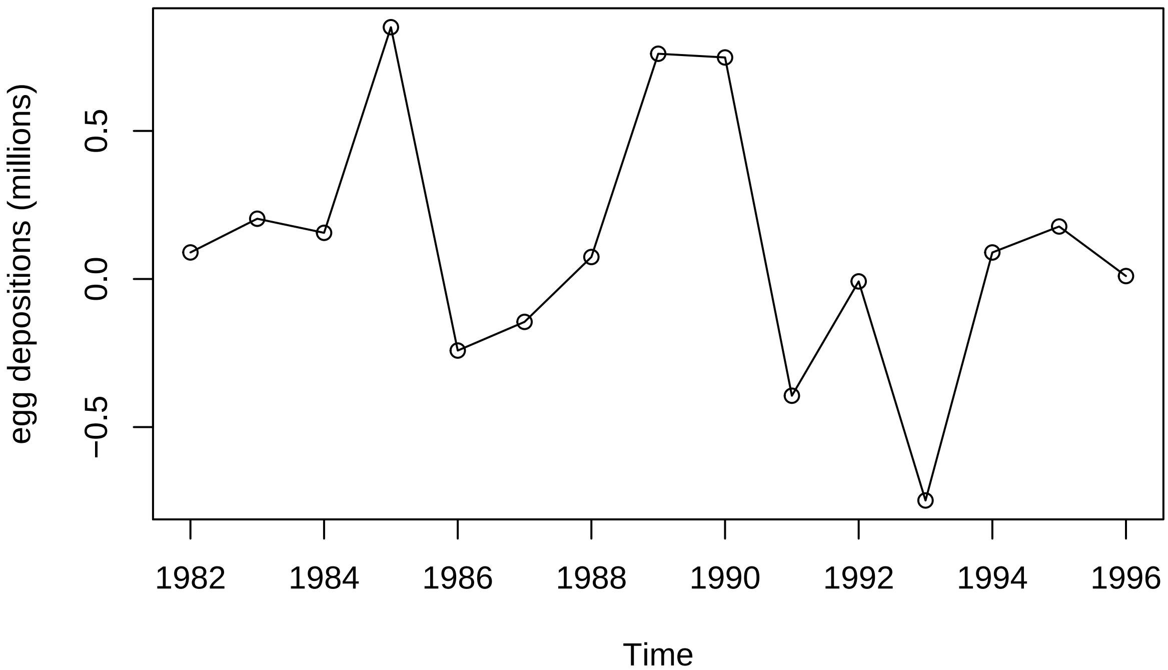


Figure 9. First Difference of Box Cox Transformed Eggs

Taking the first difference has led to the data appearing closer to a moving average distribution.

```
order = ar(diff(eggsDiff1))$order
adfTest(eggsDiff1, lags = order)
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 4
##   STATISTIC:
##     Dickey-Fuller: -0.8002
##   P VALUE:
##     0.3539
##
## Description:
##  Sun May 12 19:23:24 2019 by user:
```

```
qqnorm(eggsDiff1, sub = "Figure 10. First Difference of Eggs Normal Q-Q Plot",
       font.sub = "2", main = NULL)
qqline(eggsDiff1, col = 2)
```

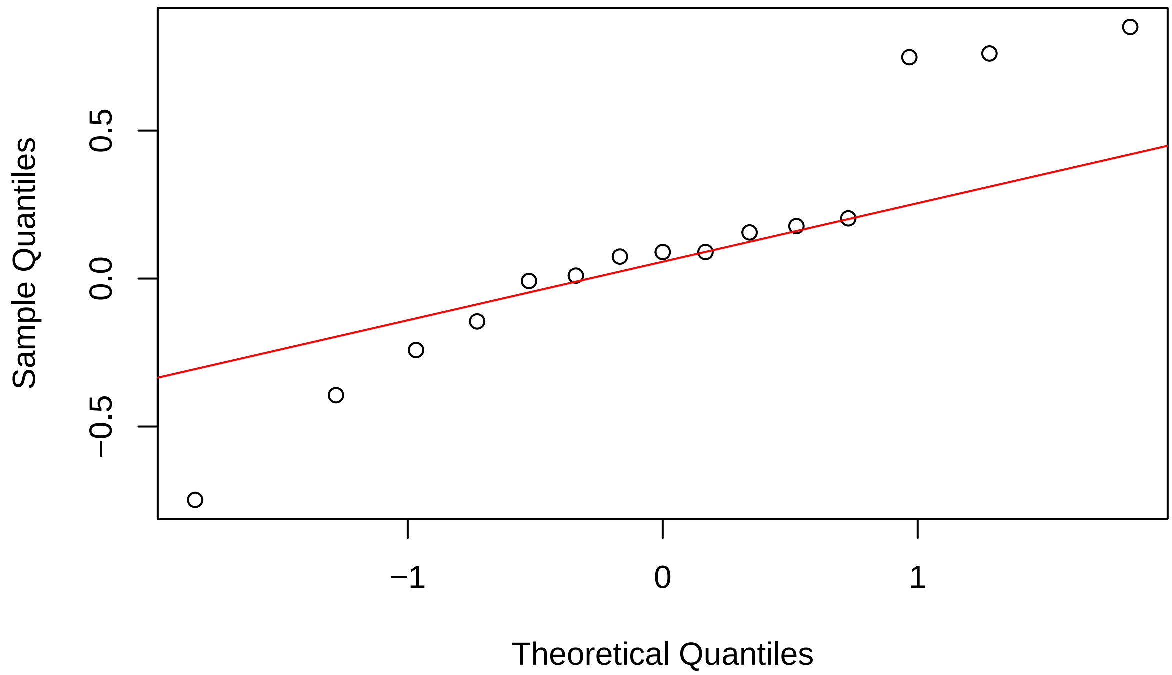


Figure 10. First Difference of Eggs Normal Q−Q Plot

```
shapiro.test(eggsDiff1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  eggsDiff1
## W = 0.93127, p-value = 0.2851
```

The ADF test confirms that the series is not stationary and remains normal based on the Shapiro-Wilk test.

Second Difference

```
eggsDiff2 <- diff(BC.eggs, difference = 2)
plot(eggsDiff2, type = 'o',
      ylab = "egg depositions (millions)",
     sub = "Figure 11: Second Difference of Box-Cox Transformed Eggs",
     font.sub = "2", main = NULL)
```

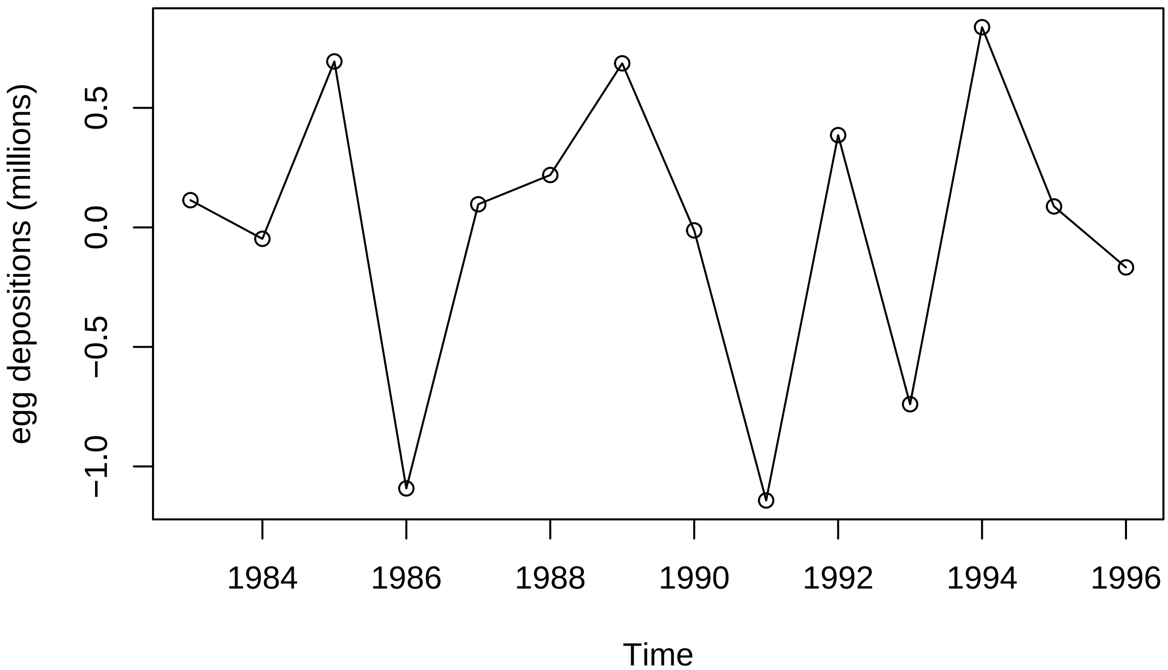


Figure 11: Second Difference of Box−Cox Transformed Eggs

Taking the second difference has led to the data appearing closer to a moving average distribution.

```
order <- ar(diff(eggsDiff2))$order
adfTest(eggsDiff2, lags = order)
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 4
##   STATISTIC:
##     Dickey-Fuller: -1.5305
##   P VALUE:
##     0.1221
##
## Description:
##  Sun May 12 19:23:25 2019 by user:
```

```
qqnorm(eggsDiff2, sub = "Figure 12: Second Difference of Eggs Normal Q-Q Plot",
       font.sub = "2", main = NULL)
qqline(eggsDiff2, col = 2)
```

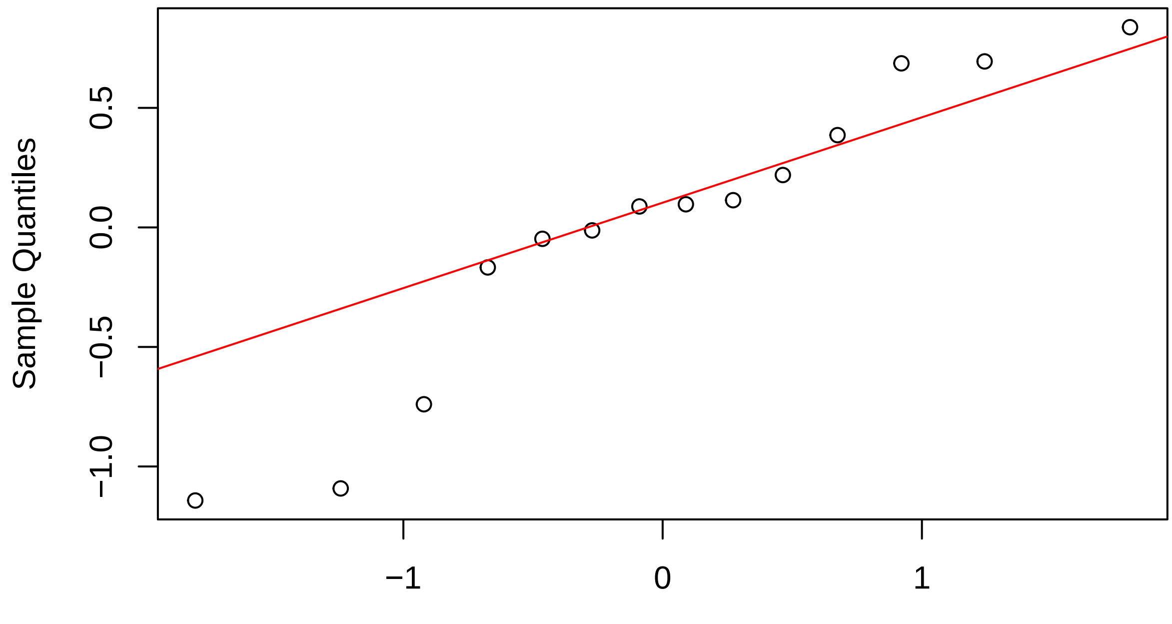


Figure 12: Second Difference of Eggs Normal Q−Q Plot

```
shapiro.test(eggsDiff2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  eggsDiff2
## W = 0.91116, p-value = 0.1638
```

The ADF test confirms that the series is not stationary and remains normal based on the Shapiro-Wilk test.

Third Difference

```
eggsDiff3 <- diff(eggs.ts, difference = 3)
plot(eggsDiff3, type = 'o',
      ylab = "egg depositions (millions)",
     sub = "Figure 13. Third Difference of Box-Cox Transformed Eggs",
     font.sub = "2", main = NULL)
```

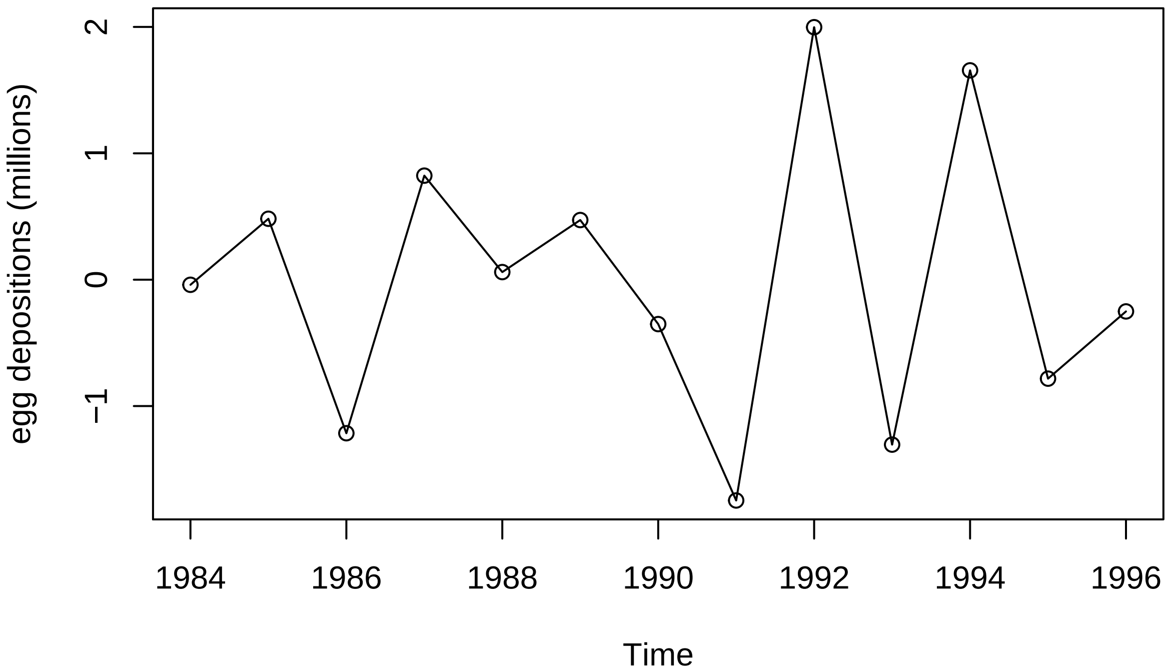


Figure 13. Third Difference of Box−Cox Transformed Eggs

Taking the third difference has led to the data appearing closer to a moving average distribution.

```
order <- ar(diff(eggsDiff3))$order
adfTest(eggsDiff3, lags = order)
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 4
##   STATISTIC:
##     Dickey-Fuller: -0.7284
##   P VALUE:
##     0.3767
##
## Description:
##  Sun May 12 19:23:26 2019 by user:
```

```
qqnorm(eggsDiff3, sub = "Figure 14. Third Difference of Eggs Normal Q-Q Plot",
       font.sub = "2", main = NULL)
qqline(eggsDiff3, col = 2)
```

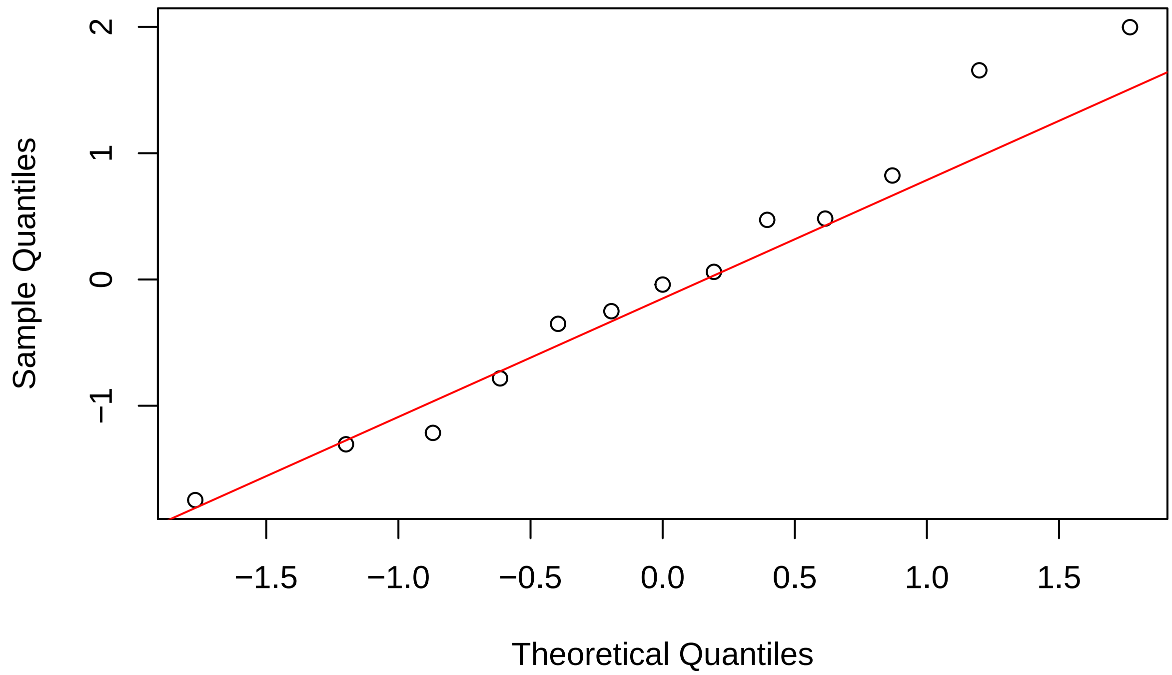


Figure 14. Third Difference of Eggs Normal Q−Q Plot

```
shapiro.test(eggsDiff3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  eggsDiff3
## W = 0.97224, p-value = 0.9195
```

The ADF test confirms that the series is not stationary and remains normal based on the Shapiro-Wilk test.

Fourth Difference

```
eggsDiff4 <- diff(eggs.ts, difference = 4)
plot(eggsDiff4, type = 'o',
      ylab = "egg depositions (millions)",
     sub = "Figure 15. Fourth Difference of Box Cox Transformed Eggs",
     font.sub = "2", main = NULL)
```

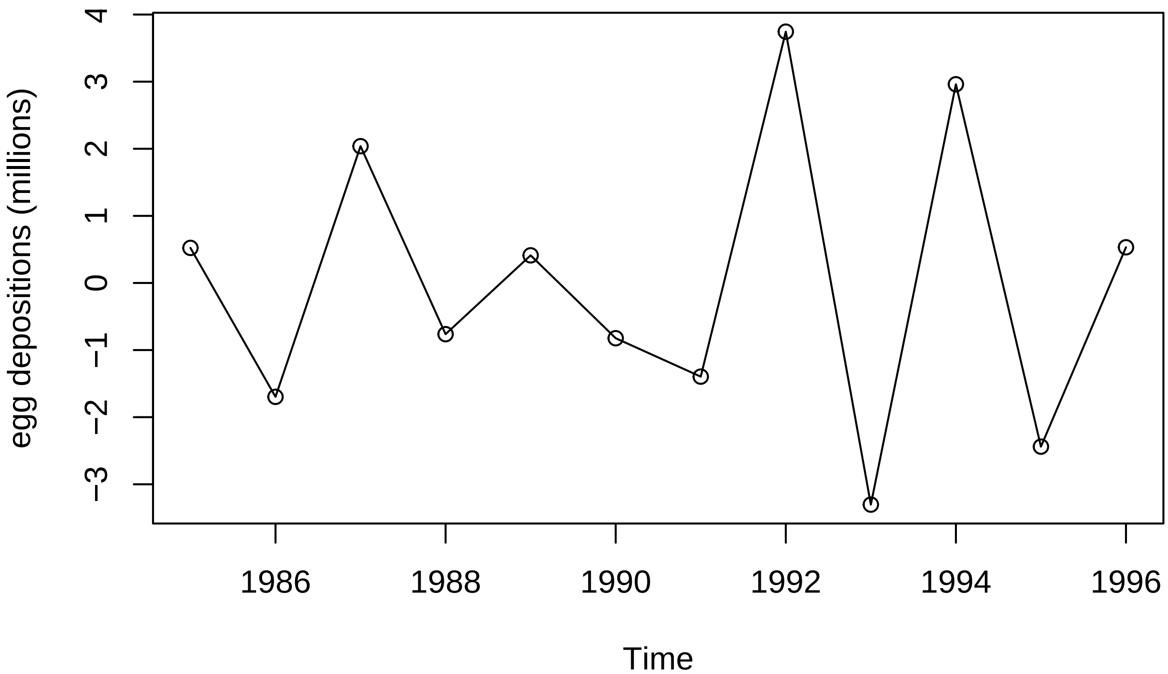


Figure 15. Fourth Difference of Box Cox Transformed Eggs

The fourth difference of the data series looks stationary based on the distribution of the data points in figure 15.

```
order <- ar(diff(eggsDiff4))$order
adfTest(eggsDiff4, lags = order)
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 2
##   STATISTIC:
##     Dickey-Fuller: -2.1524
##   P VALUE:
##     0.03368
##
## Description:
##  Sun May 12 19:23:27 2019 by user:
```

```
qqnorm(eggsDiff4, sub = "Figure 16. Fourth Difference of Eggs Normal Q-Q Plot",
       font.sub = "2", main = NULL)
qqline(eggsDiff4, col = 2)
```

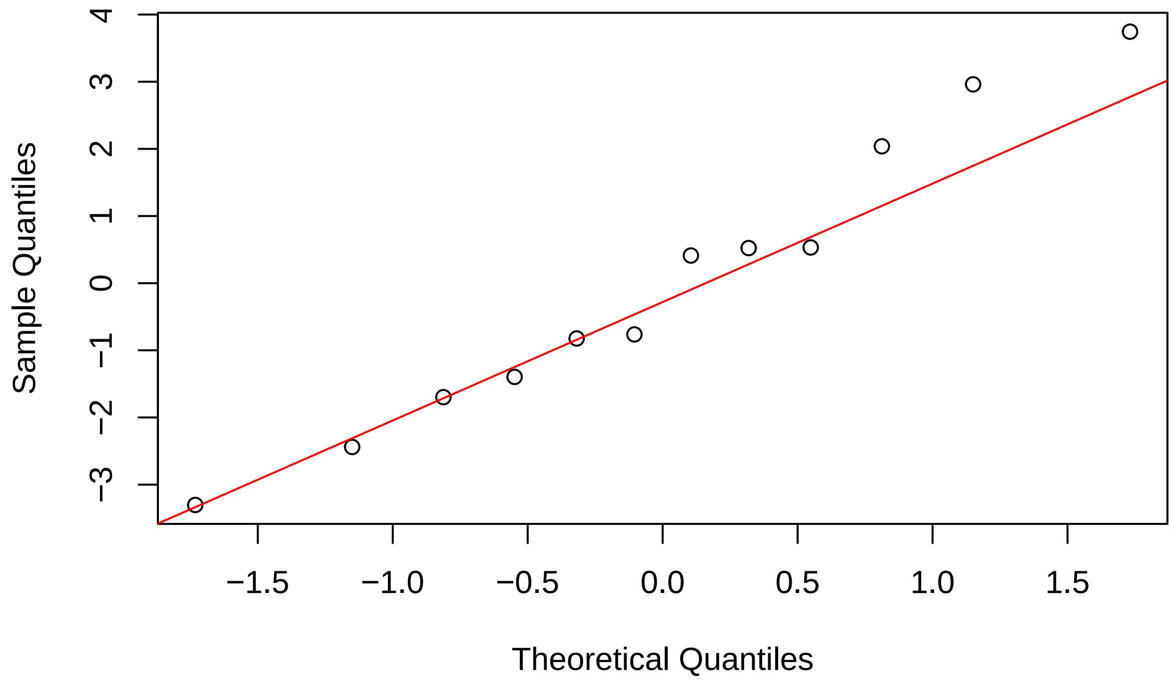


Figure 16. Fourth Difference of Eggs Normal Q−Q Plot

```
shapiro.test(eggsDiff4)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  eggsDiff4
## W = 0.96712, p-value = 0.8785
```

The p-value in the ADF test (0.033) is less than alpha (0.05) meaning that the null hypothesis that the series in non-stationary is rejected and the specification of ARIMA models can proceed on the fourth difference of the box-cox transformed data.

Model Specification

```
acf(eggsDiff4, sub = "Figure 17. Fourth Difference of Eggs ACF",
    font.sub = "2", main = "")
```

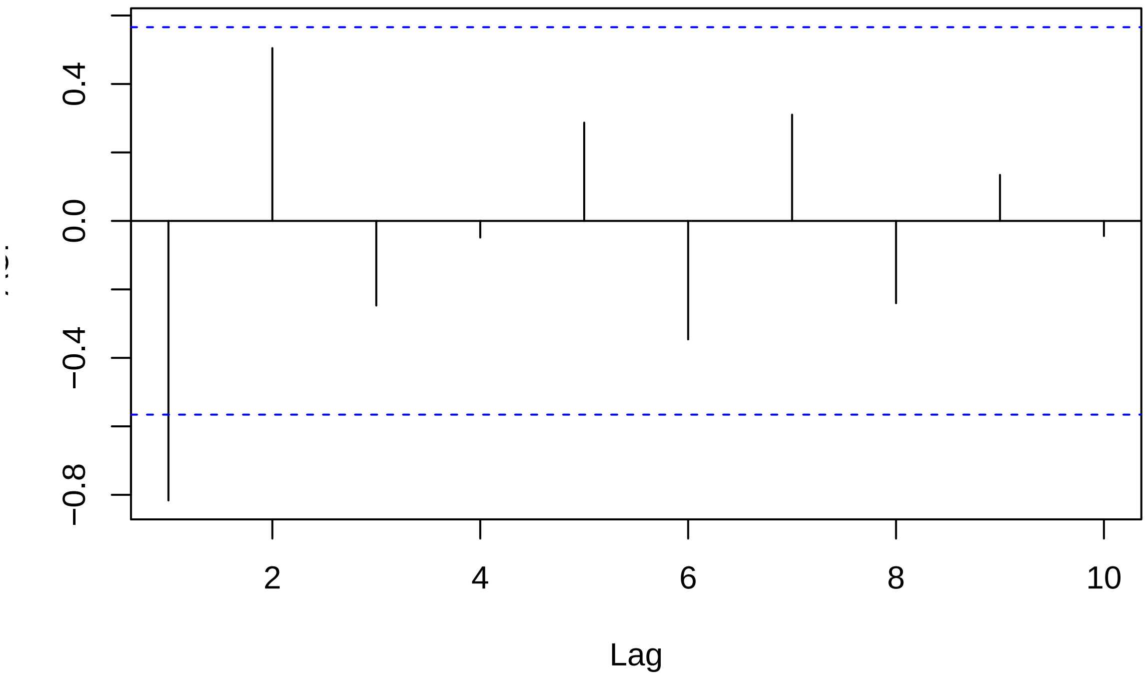


Figure 17. Fourth Difference of Eggs ACF

- Candidate models: ARIMA(0,4,1)

```r
pacf(eggsDiff4, sub = "Figure 18. Fourth Difference of Eggs PACF",
     font.sub = "2", main = "")
```

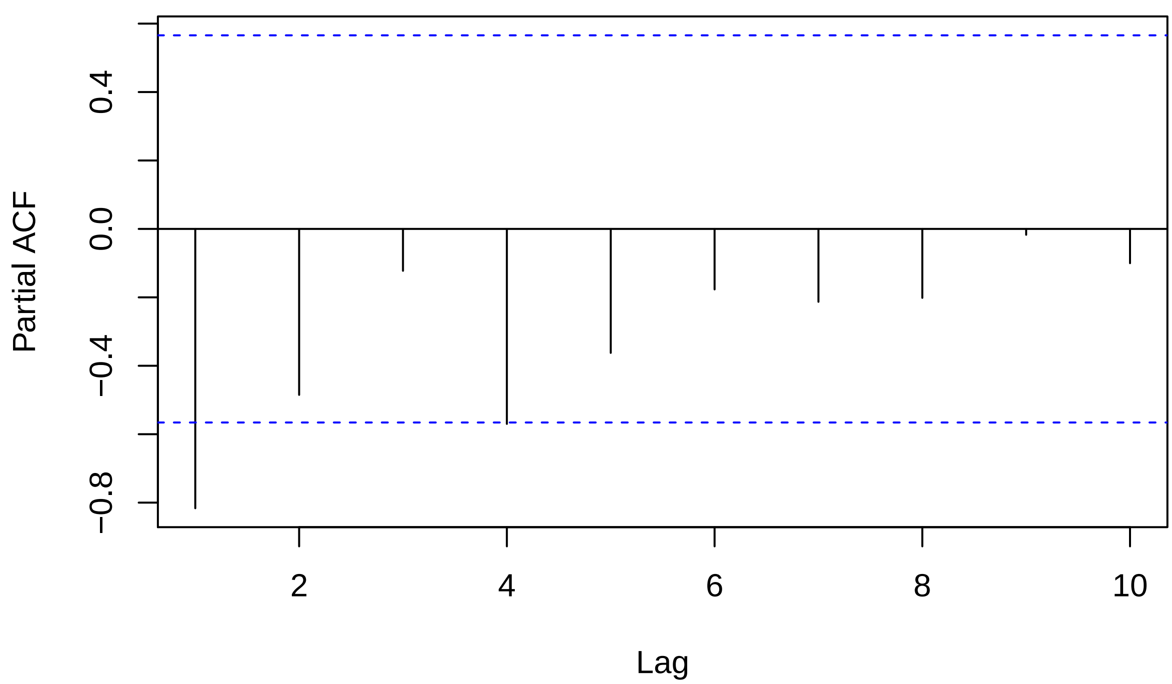


Figure 18. Fourth Difference of Eggs PACF

- Candidate models: ARIMA(1,4,0)

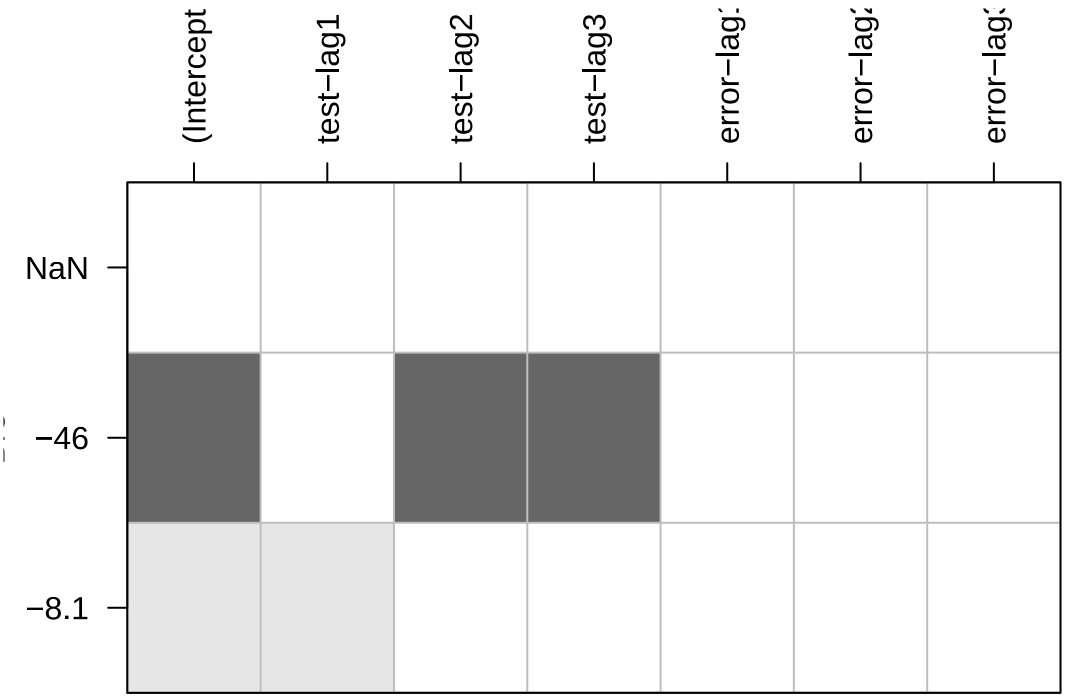
```r
eacf(eggsDiff4,ar.max = 2, ma.max =2)
```

```
## AR/MA
```

```
##   0 1 2
## 0 x o o
## 1 o o o
## 2 o o o
```

- Candidate models: ARIMA(0,4,2), ARIMA(1,4,2), ARIMA(1,4,1)

```
res = armasubsets(y = eggsDiff4, nar=3, nma=3, y.name='test', ar.method='ols')
plot(res)
```



- Candidate models: ARIMA(2,4,0), ARIMA(3,4,0)

Parameter Estimation

```
# ARIMA(0,4,1)
model_041_ml = arima(BC.eggs ,order=c(0,4,1),method='ML')
coeftest(model_041_ml)
```

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ma1 -0.97868    0.20857 -4.6923 2.702e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Significant

```
# ARIMA(1,4,0)
model_140_ml = arima(BC.eggs ,order=c(1,4,0),method='ML')
coeftest(model_140_ml)
```

```
##
```

```
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ar1 -0.75274    0.16056 -4.6883 2.754e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Significant

```r
# ARIMA(1,4,1)
model_141_ml = arima(BC.eggs ,order=c(1,4,1),method='ML')
coeftest(model_141_ml)
```

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ar1 -0.59356    0.20768 -2.8580  0.004264 **
## ma1 -0.97229    0.23583 -4.1229 3.741e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Both significant

```r
# ARIMA(0,4,2)
model_042_ml = arima(BC.eggs ,order=c(0,4,2), method='ML')
coeftest(model_042_ml)
```

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ma1 -1.86561    0.32347 -5.7675 8.044e-09 ***
## ma2  0.94485    0.31948  2.9574  0.003102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Both significant

```r
# ARIMA(1,4,2)
model_142_ml = arima(BC.eggs ,order=c(1,4,2), method='ML')
coeftest(model_142_ml)
```

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ar1 -0.26516    0.28000 -0.9470   0.34364
## ma1 -1.84442    0.39749 -4.6402 3.481e-06 ***
## ma2  0.90747    0.40222  2.2561   0.02406 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- AR1 not significant, ARIMA(1,4,2) removed from candidate models.

```r
# ARIMA(2,4,0)
model_240_ml = arima(BC.eggs ,order=c(2,4,0),method='ML')
coeftest(model_240_ml)
```

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ar1 -1.13784    0.26283 -4.3291 1.497e-05 ***
## ar2 -0.49759    0.28586 -1.7407   0.08174 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Significant

```
# ARIMA(3,4,0)
model_340_ml = arima(BC.eggs ,order=c(3,4,0), method='ML')
coeftest(model_340_ml)
```

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ar1 -1.19081    0.28962 -4.1116 3.929e-05 ***
## ar2 -0.64468    0.42955 -1.5008    0.1334
## ar3 -0.15129    0.33267 -0.4548    0.6493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Only AR1 significant. ARIMA(3,4,0) removed from candidate models.

```
sort.score <- function(x, score = c("bic", "aic")){
  if (score == "aic"){
    x[with(x, order(AIC)),]
  } else if (score == "bic") {
    x[with(x, order(BIC)),]
  } else {
    warning('score = "x" only accepts valid arguments ("aic","bic")')
  }
}

sort.score(AIC(model_041_ml, model_140_ml,
               model_141_ml, model_042_ml,
               model_240_ml),
           score = "aic")
```

```
##              df      AIC
## model_042_ml  3 42.43888
## model_141_ml  3 44.72759
## model_041_ml  2 48.23230
## model_240_ml  3 48.69193
## model_140_ml  2 49.08322
```

```
sort.score(BIC(model_041_ml, model_140_ml,
               model_141_ml, model_042_ml,
               model_240_ml),
           score = "bic")
```

```
##              df      BIC
## model_042_ml  3 43.89360
## model_141_ml  3 46.18231
```

```
## model_041_ml  2 49.20211
## model_140_ml  2 50.05304
## model_240_ml  3 50.14665
```

ARIMA(0,4,2) has the highest AIC and BIC score therefore has been chosen as the model for diagnostic checking and forecasting.

Diagnostic Checking

```
# assign model
model = model_042_ml

# assign residuals
residuals = rstandard(model)

plot(residuals, type='p',
       ylab = 'Standardised residuals',
       sub = "Figure 19. Time series plot of standardised residuals",
     font.sub = "2",
     main = "")
  abline(h=0)
```

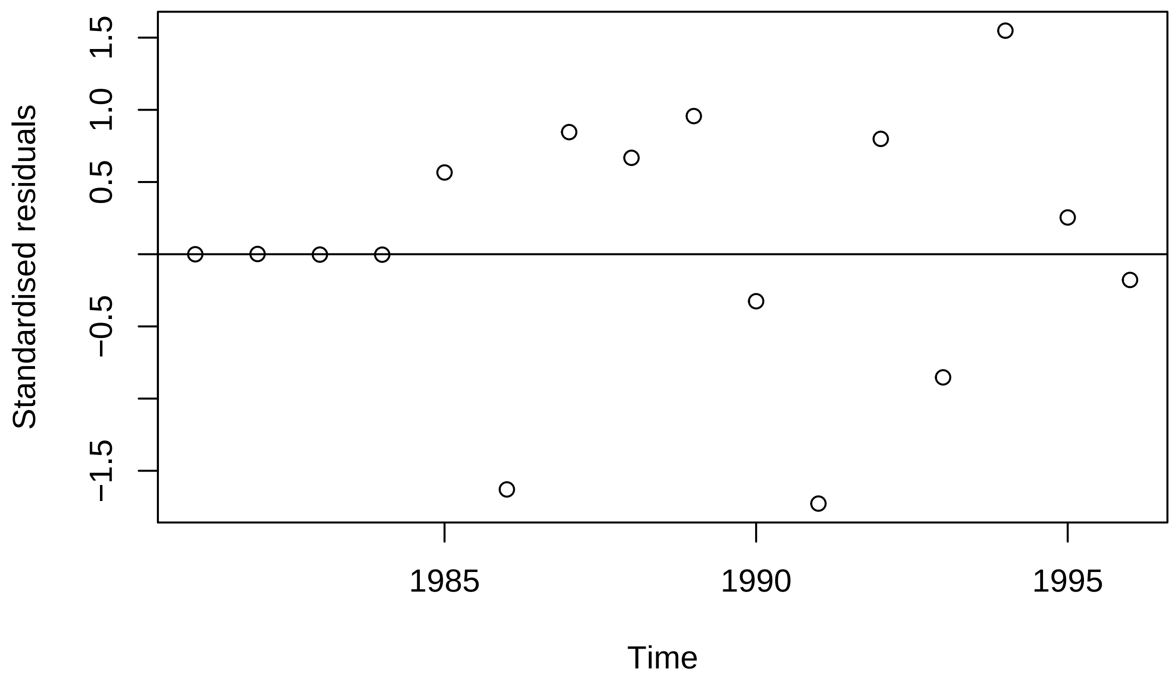


Figure 19. Time series plot of standardised residuals

There is no trend in the distribution of the residuals which suggests that the model captures the trend.

```
hist(residuals, sub = "Figure 20. Histogram of standardised residuals",
     font.sub = "2", main = "")
```

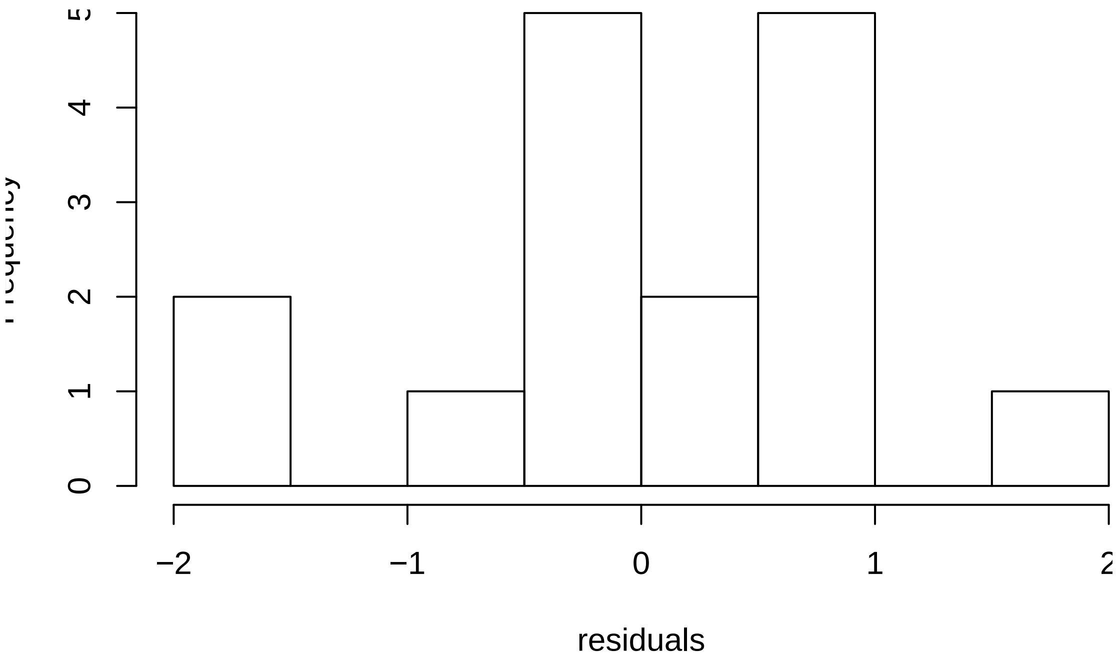


Figure 20. Histogram of standardised residuals

The residuals appear close to normally distributed.

```
qqnorm(residuals, sub = "Figure 21. QQ plot of standardised residuals",
       font.sub = "2", main = "")
qqline(residuals, col = 2)
```

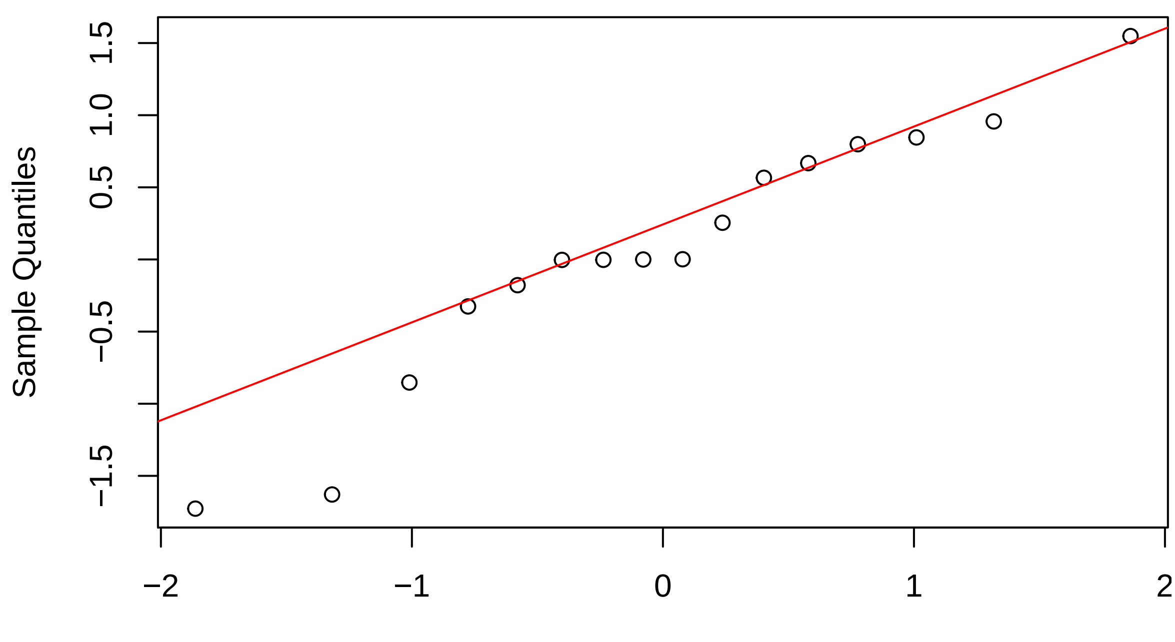


Figure 21. QQ plot of standardised residuals

```
print(shapiro.test(residuals))
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  residuals
## W = 0.93839, p-value = 0.3298
```

The residuals appear normally distributed based on the Shapiro-Wilk test.

```r
acf(residuals, sub = "Figure 22. ACF of standardised residuals",
    font.sub = "2", main = "")
```

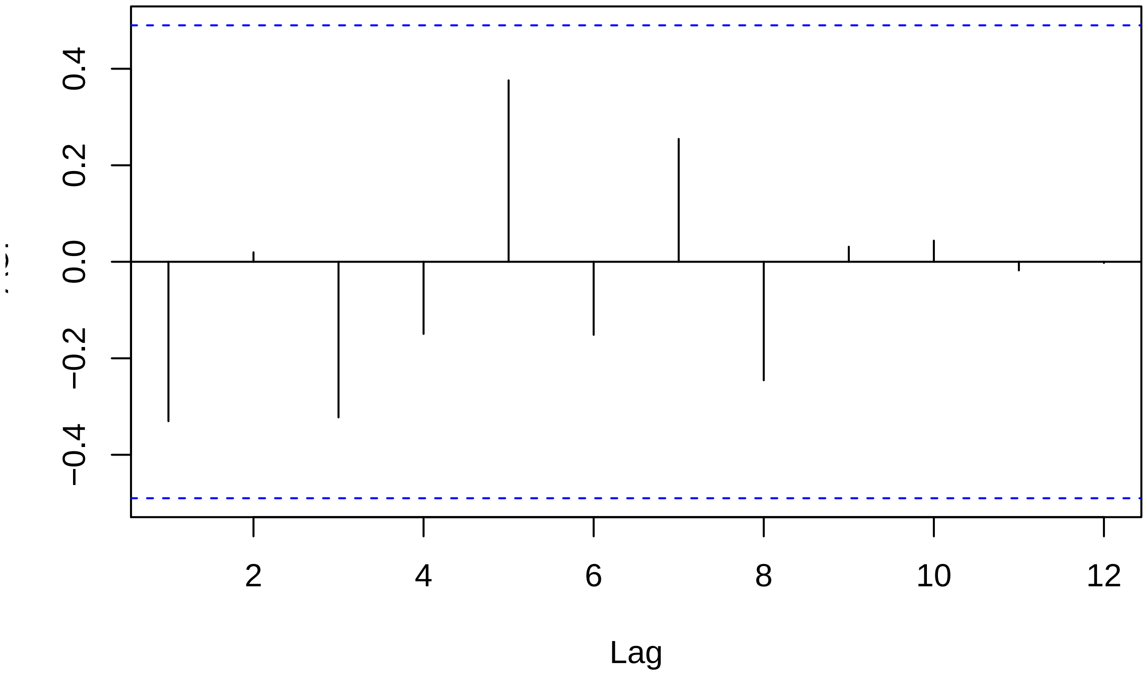
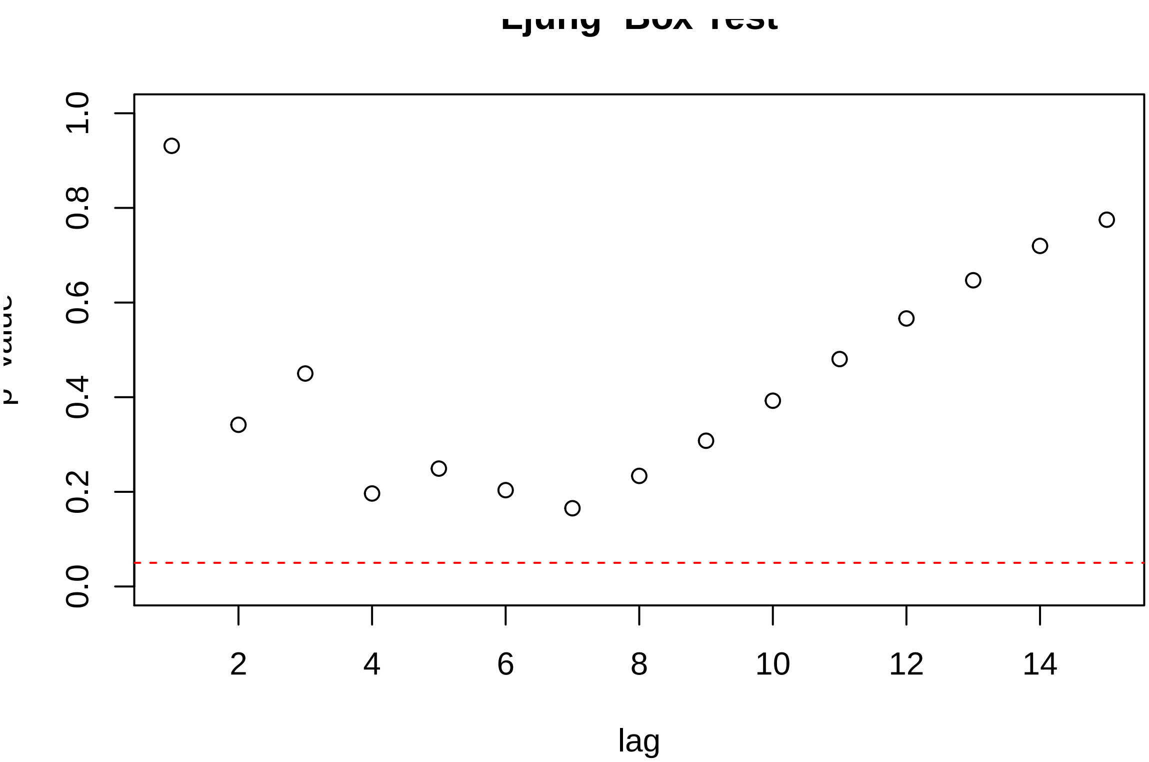


Figure 22. ACF of standardised residuals

There are no significant ACF lags which suggests that the residuals are a white noise series.

```r
k=0
LBQPlot(residuals, lag.max = length(residuals)-1,
        StartLag = k + 1,
        k = 0, SquaredQ = FALSE)
```



The lags are all outside of the alpha confidence level.

Forecast

```r
fit = Arima(eggs.ts,c(0,4,2), lambda = 0.5)
forecast(fit,h=5)
```

```
##      Point Forecast        Lo 80     Hi 80        Lo 95     Hi 95
## 1997      0.9981420   0.24606822  2.256270   0.05279473  3.127133
## 1998      0.9575859  -0.05829463  4.833706  -0.78725686  8.090619
## 1999      0.9229774  -1.61147784 10.181678  -6.00256775 19.109558
## 2000      0.9135410  -6.92876475 20.646496 -20.53587543 41.515334
## 2001      0.9487792 -19.25254585 40.143353 -52.21726065 84.166994
```

```r
plot(forecast(fit,h=5), sub = "Figure 24. Forecasts from ARIMA(0,4,2",
     font.sub = "2", main = "", type = 'o')
```

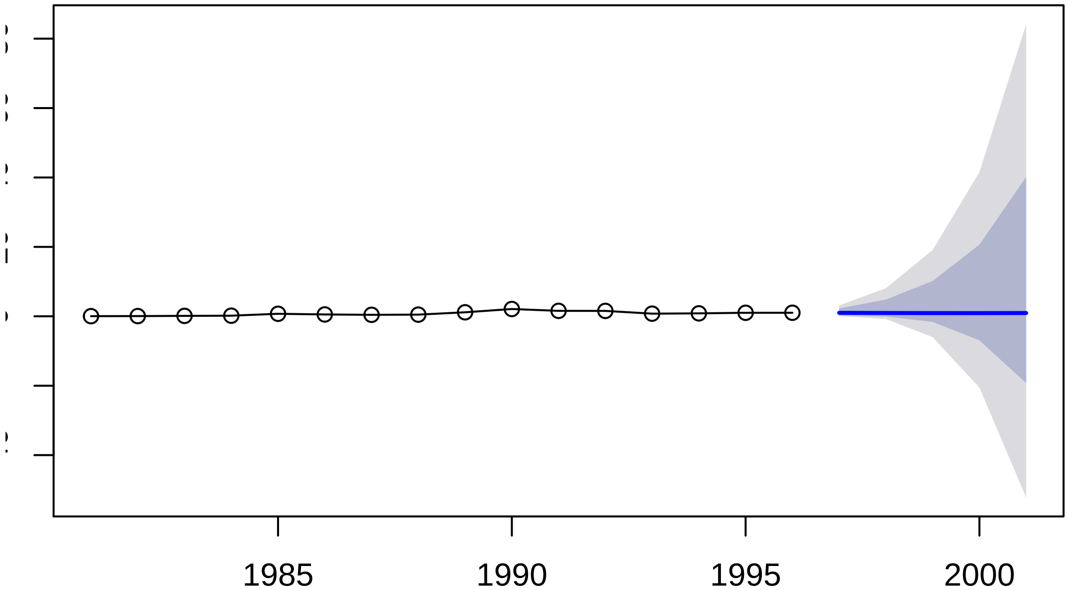


Figure 24. Forecasts from ARIMA(0,4,2

The plot shows the huge variation in estimates moving forward over the five years. It doesn't make logical sense with the data for the model to predict values with confidence intervals including negative values (i.e. fish can't lay less than zero eggs).