PracticeFinal

Jim Liu

12/5/2020

```r
library(TSA)
```

```
##
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':
##
##     acf, arima

## The following object is masked from 'package:utils':
##
##     tar
```

```r
library(mgcv)
```

```
## Loading required package: nlme

## This is mgcv 1.8-33. For overview type 'help("mgcv-package")'.
```

```r
library(vars)
```

```
## Loading required package: MASS

## Loading required package: strucchange

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: urca

## Loading required package: lmtest
```

```r
library(tseries)
```

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

```r
library(fGarch)
```

```
## Loading required package: timeDate
```

```
##
## Attaching package: 'timeDate'
```

```
## The following objects are masked from 'package:TSA':
##
##     kurtosis, skewness
```

```
## Loading required package: timeSeries
```

```
##
## Attaching package: 'timeSeries'
```

```
## The following object is masked from 'package:zoo':
##
##     time<-
```

```
## Loading required package: fBasics
```

```r
library(rugarch)
```

```
## Loading required package: parallel
```

```
##
## Attaching package: 'rugarch'
```

```
## The following object is masked from 'package:stats':
##
##     sigma
```

```r
erie = read.csv("/Users/jim/Dropbox (GaTech)/Courses/ISyE6402/Final/erie.csv", header = TRUE)
```

Lake Erie Water Levels Analysis

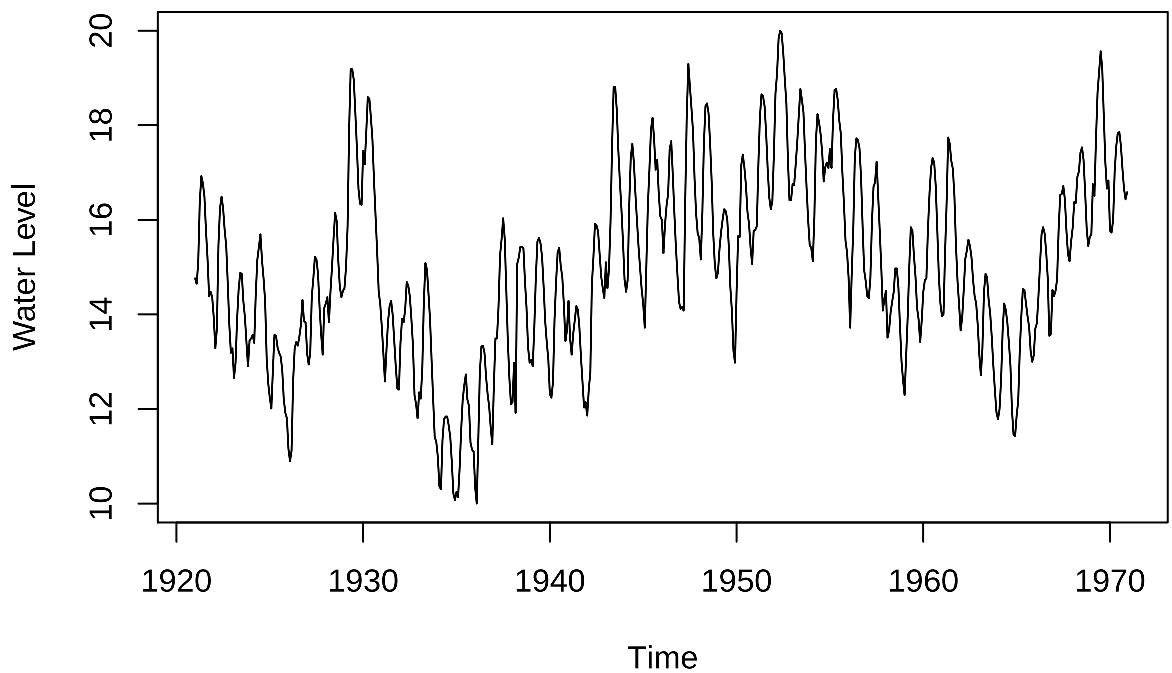
For this portion of the test, you will be examining the monthly average water levels of Lake Erie from 1921-1970.

For all questions in this R data analysis, make sure to provide the R code also. Show your work for full credit.

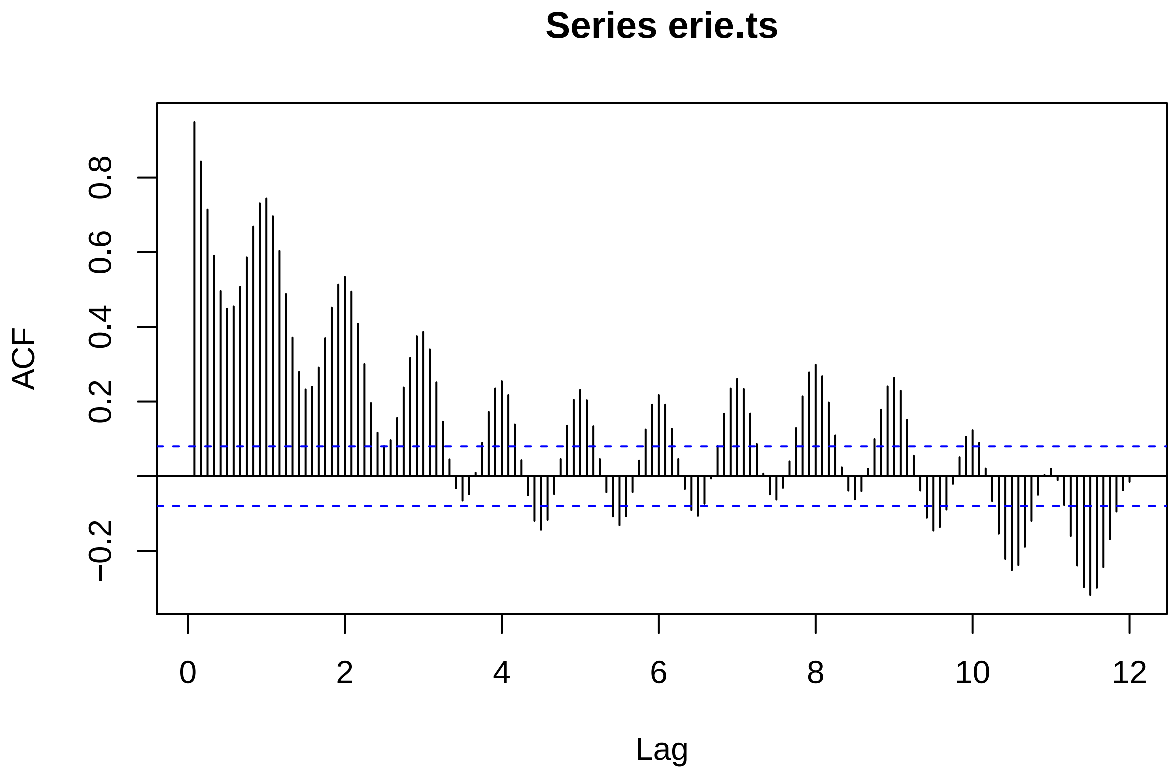
1. Exploratory Analysis.

(a) Plot the time series and identify its main features, including trend and seasonality. Which assumptions of stationarity seem to be violated?

```
erie.ts = ts(erie[, 2], start=c(1921,1), frequency = 12)
ts.plot(erie.ts, ylab="Water Level")
```




```
acf(erie.ts, lag.max = 12 * 12, cex=0.3, na.action = na.pass)
```



From the time series plot, we could see that this time series does not follow three assumptions: (1) constant mean: from the time series plot, we could see there is a trend. (2) constant variance: we could see the variance around 1930 is larger than other periods. So the variance is not constant. (3) covariance independent of time: based on the autocorelation plot, we could see there is a seasonality in this plot.

2. Estimating the trend.

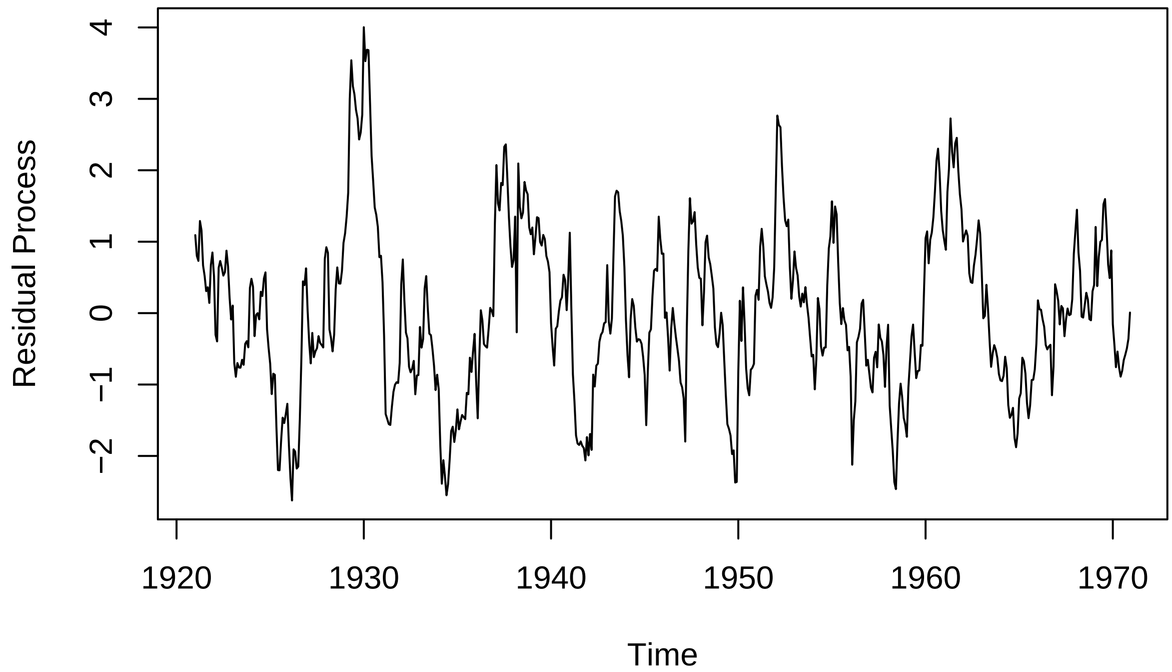
(a)

Fit a nonparametric regression model to estimate the trend and seasonality of the time series. Provide the R output. Are the parameters significant at the 10% significant level? The following R code may be useful in getting you started in fitting the model.

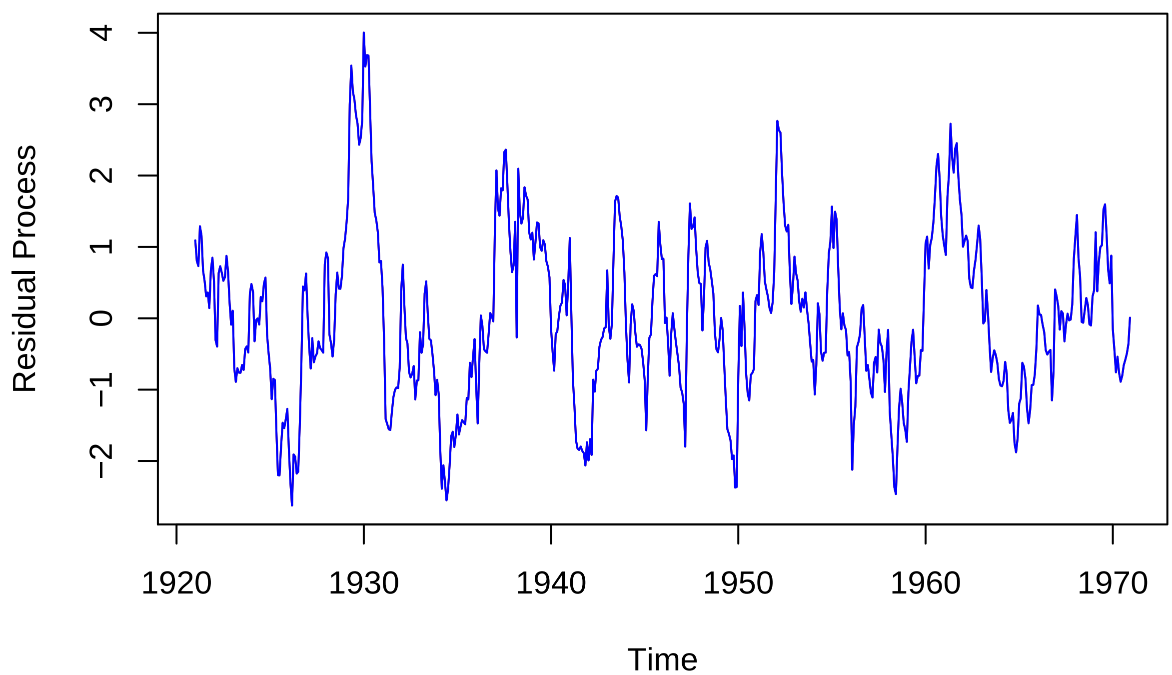
```
level = erie[,2]
level = level[1:600]
H = harmonic(ts(level, start = 1/12, end = 50, deltat = 1/12), 2)
```

```
time.pts = c(1:length(level))
time.pts = c(time.pts - min(time.pts))/max(time.pts)
```

```
gam.fit = gam(level~s(time.pts)+H)
dif.fit.gam = ts((level-fitted(gam.fit)),start=1921,frequency=12)
ts.plot(dif.fit.gam,ylab="Residual Process")
```




```r
## Compare approaches
ts.plot(dif.fit.gam,ylab="Residual Process",col="brown")
lines(dif.fit.gam,col="blue")
```




```r
summary(gam.fit)
```

```
##
## Family: gaussian
## Link function: identity
```

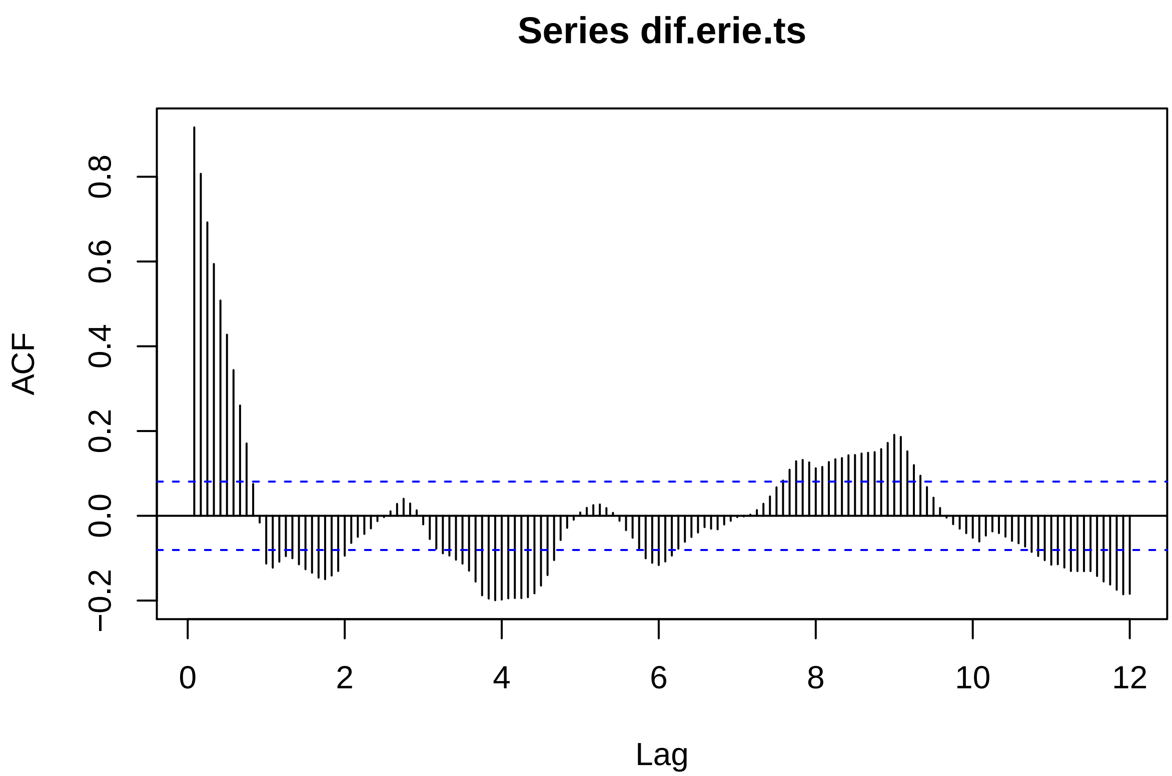
```
##
## Formula:
## level ~ s(time.pts) + H
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.99305    0.04776 313.951  < 2e-16 ***
## Hcos(2*pi*t) -1.23075    0.06754 -18.222  < 2e-16 ***
## Hcos(4*pi*t)  0.14467    0.06754   2.142   0.0326 *
## Hsin(2*pi*t) -0.29734    0.06759  -4.399 1.29e-05 ***
## Hsin(4*pi*t) -0.03063    0.06755  -0.453   0.6504
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##               edf Ref.df     F p-value
## s(time.pts) 8.917  8.998 92.05  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.662   Deviance explained = 66.9%
## GCV = 1.4009  Scale est. = 1.3684    n = 600
```

Most parameters except Hsin($4pit$) are statistically significant.

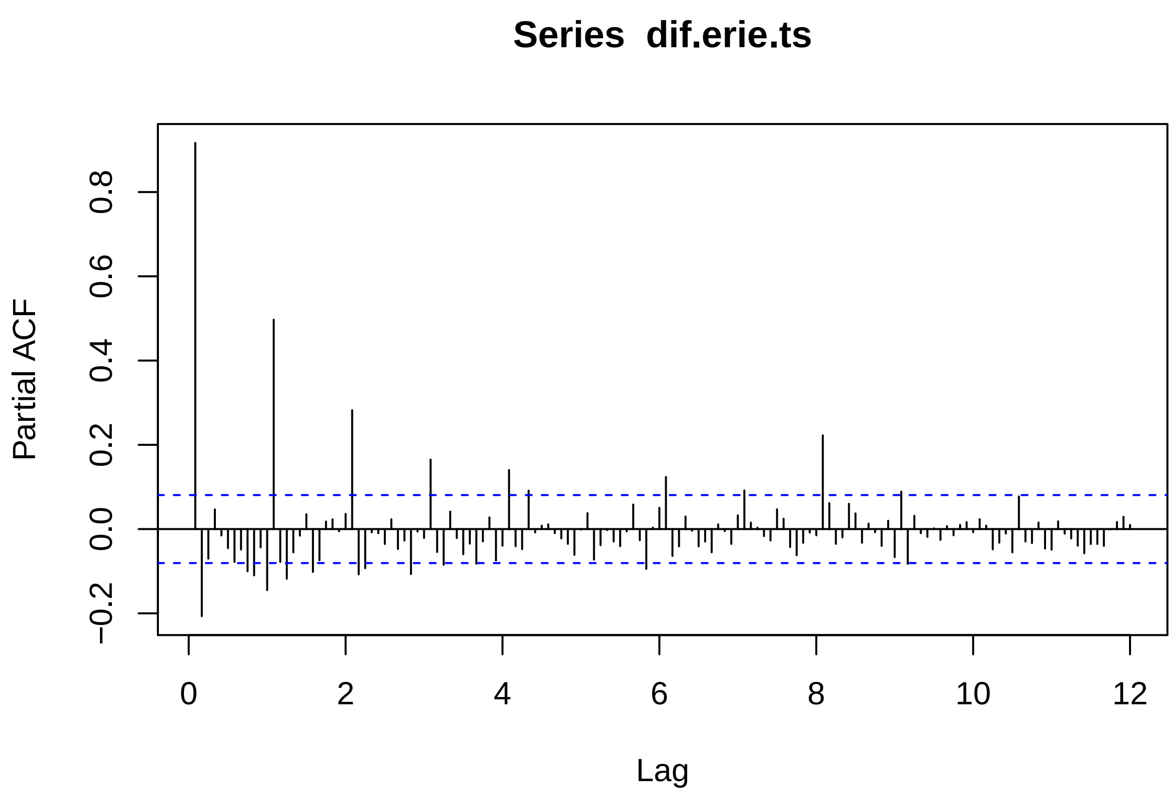
(b)

Use differencing to directly remove the trend and seasonality from the original time series. Report the order of integration. Plot the differenced time series, ACF and PACF and comment on the differenced time series.

```
dif.erie.ts = diff(erie.ts, lag=12)
acf(dif.erie.ts, na.action = na.pass, lag.max = 12 * 12)
```




```
pacf(dif.erie.ts, na.action = na.pass, lag.max = 12 * 12)
```



Since the water level is related to months, we take the order = 12 to get differenced data. However, based on ACF and PACF, we could still see seasonality from the ACF plot.

(c)

Based on the plots made in the previous step, suggest an appropriate ARIMA model to fit. Refer to specific aspects of the ACF and/or PACF plots that lead you to suggest the model.

ARIMA(0, 12, 8)

3. ARIMA Modeling.

(a) The are 81 ARIMA(p,1,q)(P,1,D) models. Select the best model based on AIC.

```
test_modelA <- function(p,d,q){
mod = arima(erie.ts, order=c(p,d,q), method="ML")
current.aic = AIC(mod)
df = data.frame(p,d,q,current.aic)
names(df) <- c("p","d","q","AIC")
print(paste(p,d,q,current.aic,sep=" "))
return(df)
}

orders = data.frame(Inf,Inf,Inf,Inf)
names(orders) <- c("p","d","q","AIC")


for (p in 0:2){
  for (d in 0:2){
    for (q in 0:2) {
      possibleError <- tryCatch(
        orders<-rbind(orders,test_modelA(p,d,q)),
        error=function(e) e
      )
      if(inherits(possibleError, "error")) next

    }
  }
}
```

```
## [1] "0 0 0 2544.73991906472"
## [1] "0 0 1 1877.39920969981"
## [1] "0 0 2 1438.90932136935"
## [1] "0 1 0 1171.44111148771"
## [1] "0 1 1 1015.48882202163"
## [1] "0 1 2 972.759441243084"
## [1] "0 2 0 1131.84794002762"
## [1] "0 2 1 1112.70838178796"
## [1] "0 2 2 1022.45458547027"
```

```
## Warning in stats::arima(x = x, order = order, seasonal = seasonal, xreg =
## xreg, : possible convergence problem: optim gave code = 1
```

```
## [1] "1 0 0 1171.14973199869"
## [1] "1 0 1 998.770350314576"
## [1] "1 0 2 980.460414464743"
```

```
## [1] "1 1 0 975.141093100406"
## [1] "1 1 1 975.132796728166"
## [1] "1 1 2 969.051261448883"
## [1] "1 2 0 1115.00733847981"
## [1] "1 2 1 981.397104726438"
## [1] "1 2 2 981.467550882602"
## [1] "2 0 0 929.310345503853"
## [1] "2 0 1 931.265104046663"
## [1] "2 0 2 927.495556777966"
## [1] "2 1 0 973.691167565993"
## [1] "2 1 1 884.944430573136"
## [1] "2 1 2 748.689052662587"
## [1] "2 2 0 1115.28782011429"
## [1] "2 2 1 980.096177071801"
## [1] "2 2 2 985.368266340088"
```

```
orders <- orders[order(-orders$AIC),]
tail(orders)
```

```
##    p d q      AIC
## 16 1 1 2 969.0513
## 21 2 0 1 931.2651
## 20 2 0 0 929.3103
## 22 2 0 2 927.4956
## 24 2 1 1 884.9444
## 25 2 1 2 748.6891
```

```
final_model = arima(erie.ts, order = c(2,1,2), method = "ML")
```

(b) Report the estimated model parameters and their 99% confidence intervals for the model selected in the previous question.

```
library(lmtest)
coeftest(final_model)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error  z value  Pr(>|z|)
## ar1  1.7309844  0.0014735 1174.743 < 2.2e-16 ***
## ar2 -0.9991695  0.0013365 -747.605 < 2.2e-16 ***
## ma1 -1.7016190  0.0231228  -73.591 < 2.2e-16 ***
## ma2  0.9610347  0.0274502   35.010 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c) Fit an ARIMA(1,1,3)(0,1,1) model to the time series. Compare this model to the model you selected in the previous question 3(a) using AIC and BIC. Which is the preferred model?

```
arim1113 = arima(erie.ts, order = c(1,1,3), method = "ML")
arima011 = arima(erie.ts, order = c(0,1,1), method = "ML")
```

```r
AIC(arim1113)
```

```
## [1] 969.4627
```

```r
AIC(arima011)
```

```
## [1] 1015.489
```

```r
BIC(arim1113)
```

```
## [1] 991.439
```

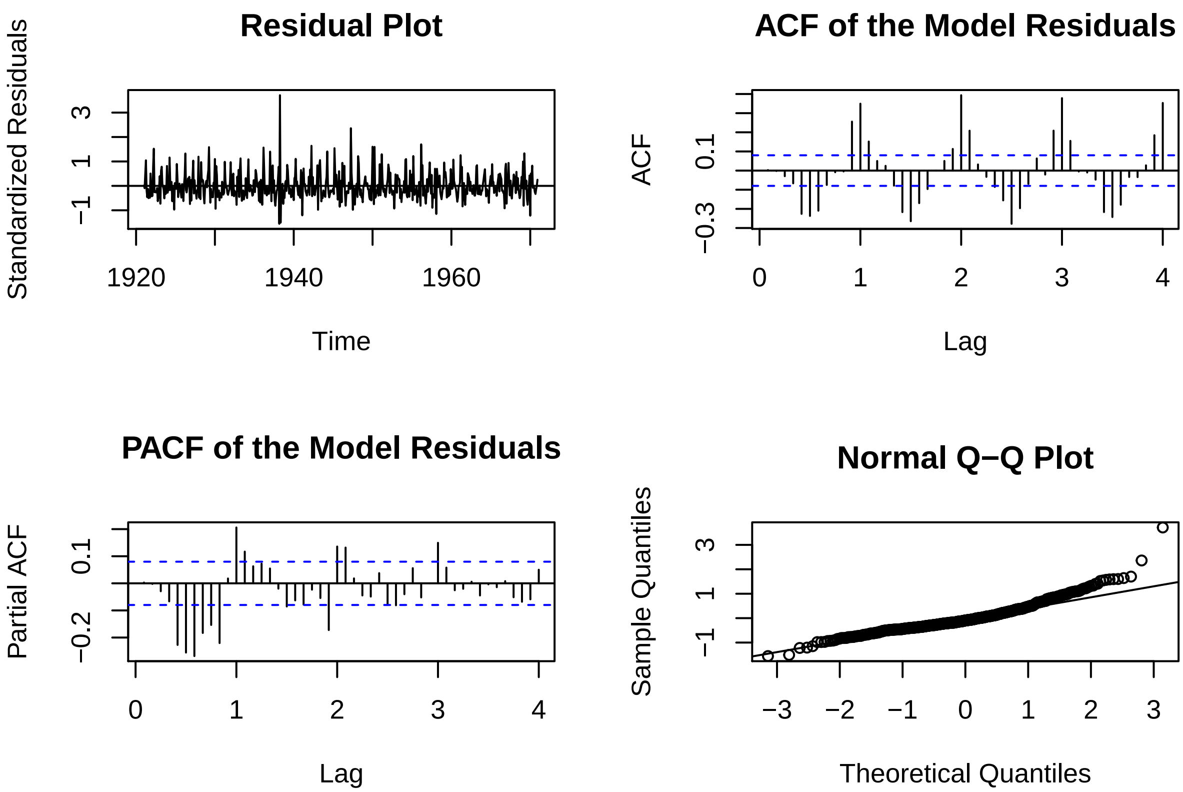
```r
BIC(arima011)
```

```
## [1] 1024.279
```

4. Residual Analysis.

(a) Do the residuals of the ARIMA(1,1,3)(0,1,1) model display autocorrelation? Use appropriate plots and/or hypothesis tests to support your answer.

```r
resids.113 = resid(arim1113)
## Residual Analysis
par (mfrow=c(2,2))
plot(resids.113, ylab='Standardized Residuals', main="Residual Plot")
abline(h=0)
acf(resids.113,main= 'ACF of the Model Residuals', na.action = na.pass, lag.max = 12*4)
pacf(resids.113,main='PACF of the Model Residuals', na.action = na.pass, lag.max = 12*4)
qqnorm(resids.113)
qqline(resids.113)
```



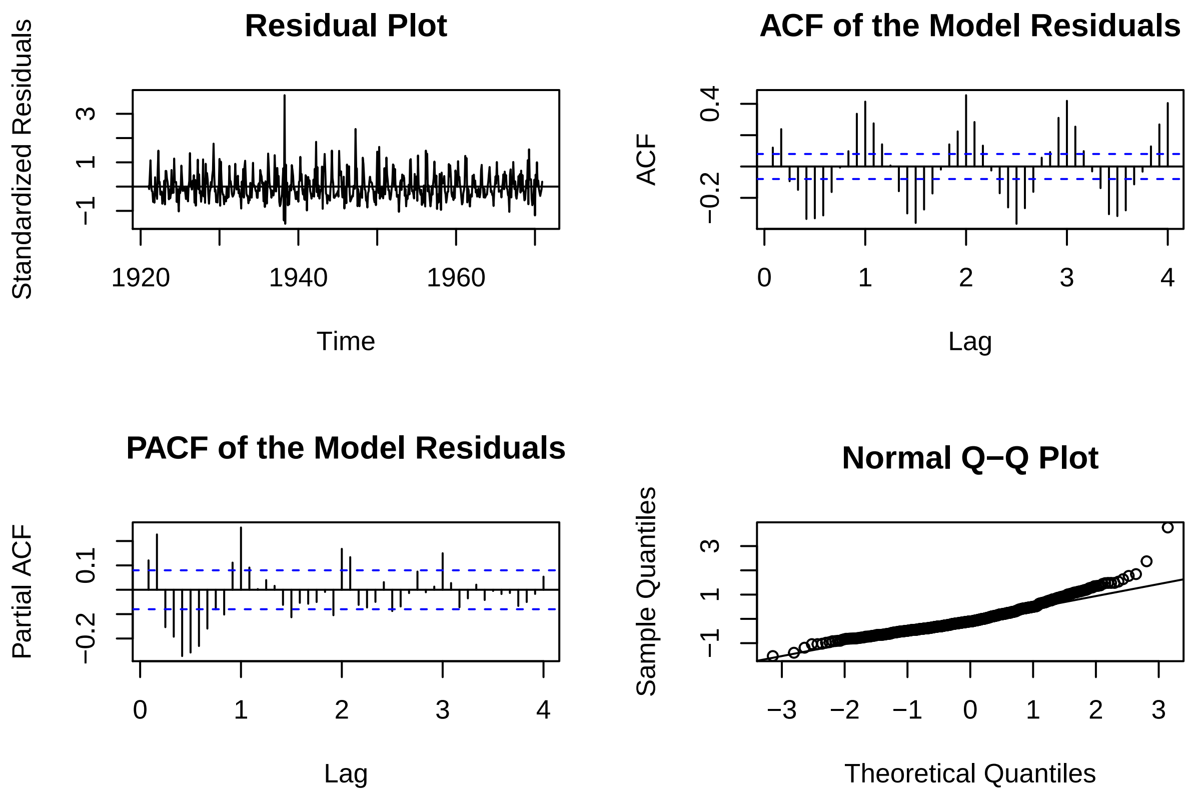

```
Box.test(resids.113, lag = (1+1+3), type = "Box-Pierce", fitdf = (1+3))
```

```
##
##  Box-Pierce test
##
## data:  resids.113
## X-squared = 33.869, df = 1, p-value = 5.894e-09
```

```
Box.test(resids.113, lag = (1+1+3), type = "Ljung-Box", fitdf = (1+3))
```

```
##
##  Box-Ljung test
##
## data:  resids.113
## X-squared = 34.262, df = 1, p-value = 4.818e-09
```

```
resids.011 = resid(arima011)
## Residual Analysis
par (mfrow=c(2,2))
plot(resids.011, ylab='Standardized Residuals', main="Residual Plot")
abline(h=0)
acf(resids.011,main= 'ACF of the Model Residuals', na.action = na.pass, lag.max = 12*4)
pacf(resids.011,main='PACF of the Model Residuals', na.action = na.pass, lag.max = 12*4)
qqnorm(resids.011)
qqline(resids.011)
```




```r
Box.test(resids.011, lag = (0+1+1), type = "Box-Pierce", fitdf = (0+1))
```

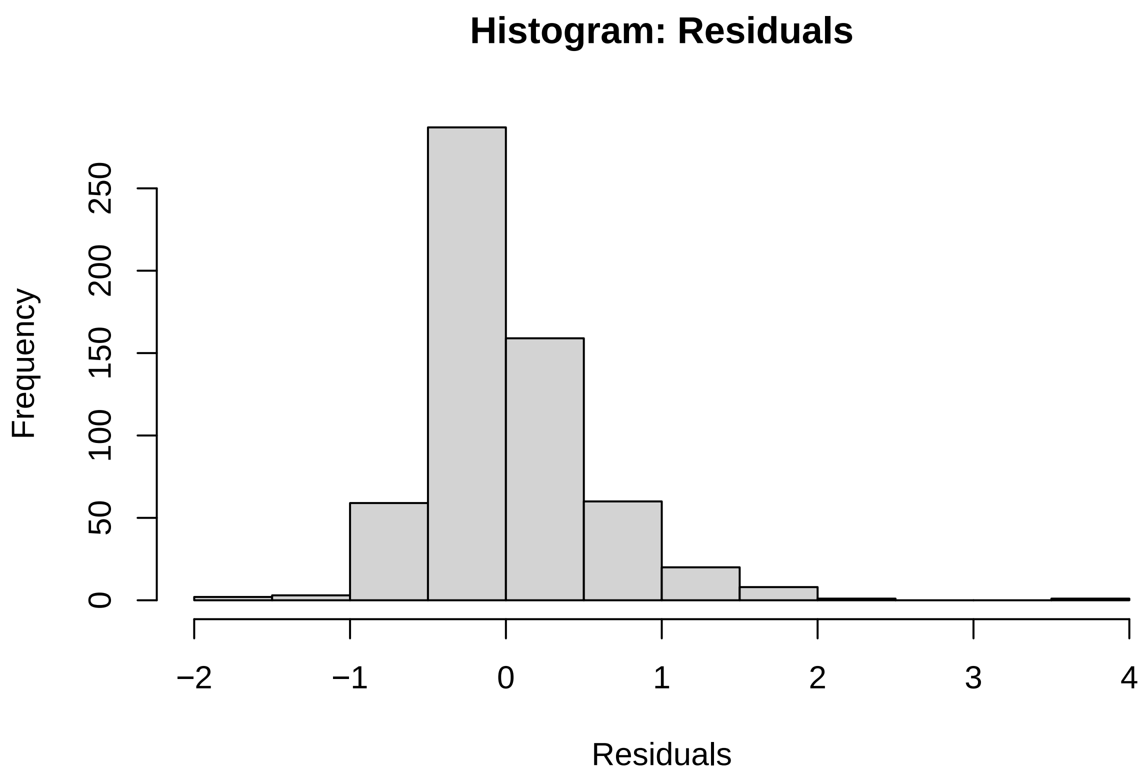
```
##
##  Box-Pierce test
##
## data:  resids.011
## X-squared = 42.83, df = 1, p-value = 5.97e-11
```

```r
Box.test(resids.011, lag = (0+1+1), type = "Ljung-Box", fitdf = (0+1))
```

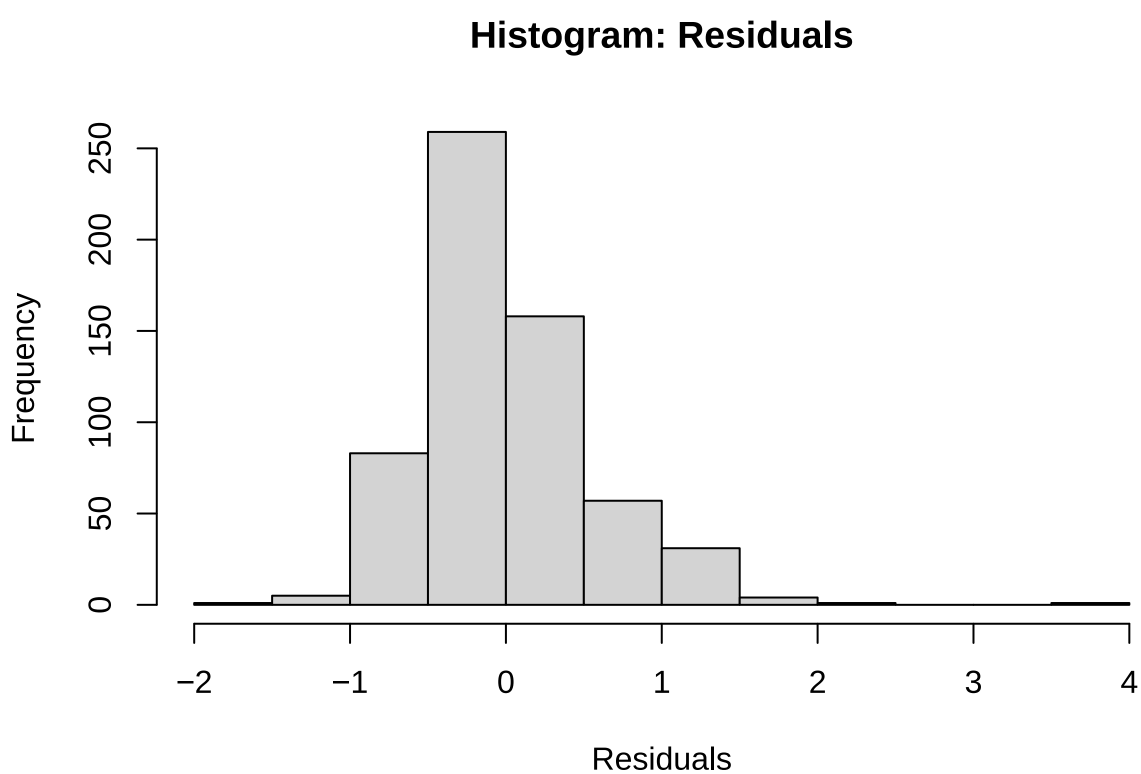
```
##
##  Box-Ljung test
##
## data:  resids.011
## X-squared = 43.102, df = 1, p-value = 5.195e-11
```

(b) Do the residuals of the ARIMA(1,1,3)(0,1,1) model follow a normal distribution? Use appropriate plots and/or hypothesis tests to support your answer.

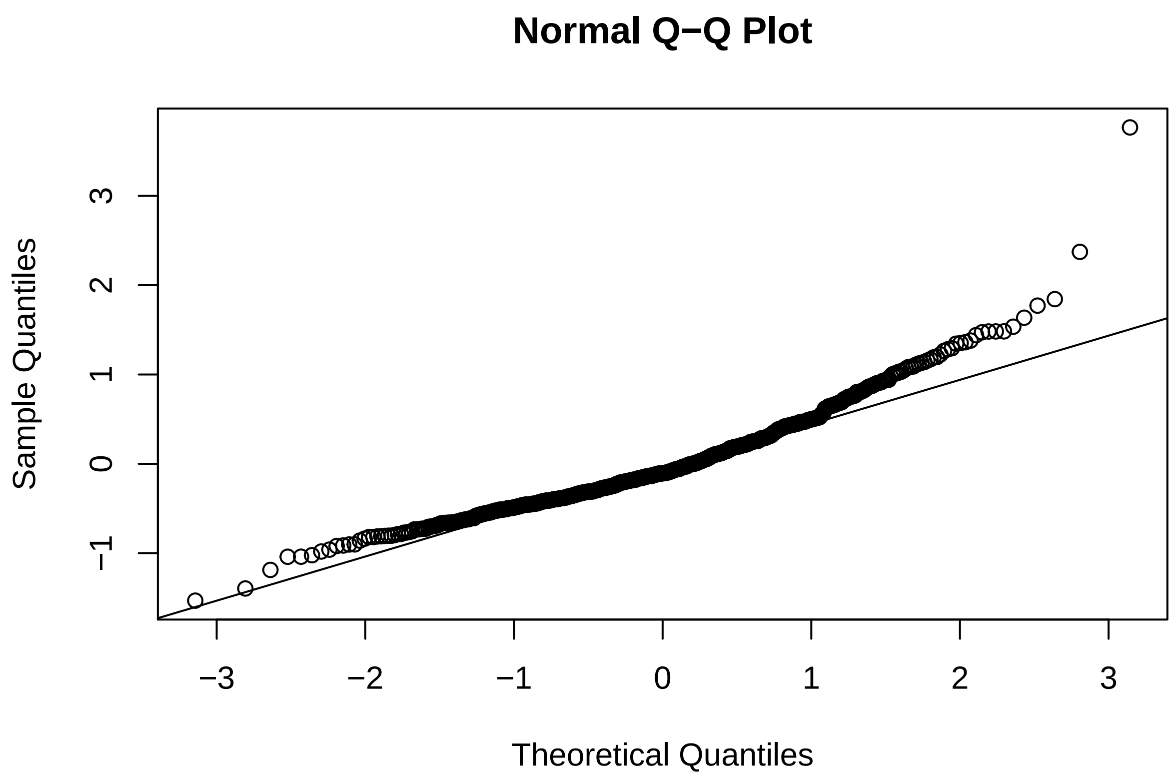
```r
hist(resids.113,xlab='Residuals',main='Histogram: Residuals')
```



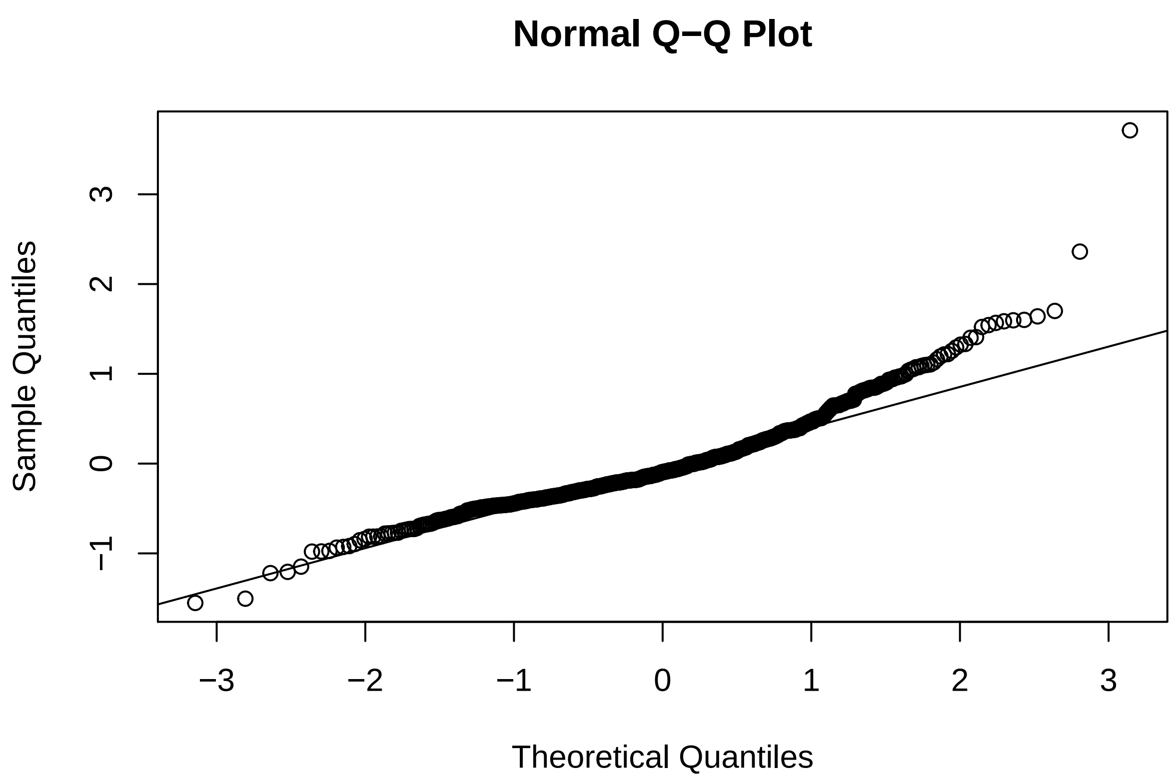

```r
hist(resids.011,xlab='Residuals',main='Histogram: Residuals')
```




```r
qqnorm(resids.011)
qqline(resids.011)
```




```r
qqnorm(resids.113)
qqline(resids.113)
```



5. Model Fit.

(a) Plot the fitted values of the model you selected in question 3(a) alongside the original values. Do the same for the ARIMA(1,1,3)(0,1,1) model. Which model fits better? Use the mean absolute error and the precision measure to evaluate the fit.

```
## MAE
mean(abs(fitted(final_model)[1:600]-level))
```

```
## [1] 0.328108
```

```
## PM
sum((fitted(final_model)[1:600]-level)^2)/sum((level-mean(level))^2)
```

```
## [1] 0.04921367
```

```
## MAE
mean(abs(fitted(arim1113)[1:600]-level))
```

```
## [1] 0.3988436
```

```
## PM
sum((fitted(arim1113)[1:600]-level)^2)/sum((level-mean(level))^2)
```

```
## [1] 0.07171313
```

```
## MAE
mean(abs(fitted(arima011)[1:600]-level))
```

```
## [1] 0.4280662
```

```
## PM
sum((fitted(arima011)[1:600]-level)^2)/sum((level-mean(level))^2)
```

```
## [1] 0.07824054
```