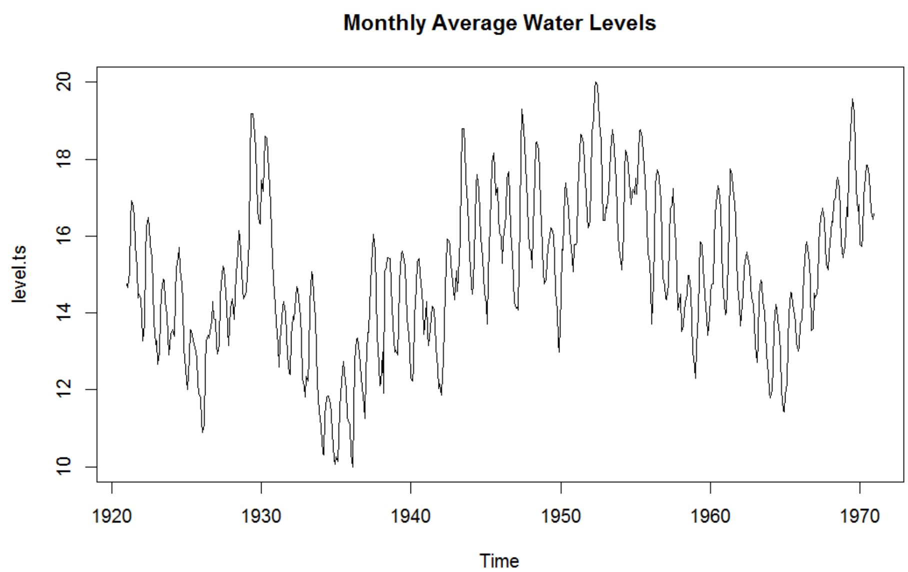
Time Series
Practice Final Solutions

I Lake Erie

1 Exploratory Analysis

(a) Based on the plot, water levels may overall be increasing over time, although more years of data are needed to say for sure. There is however clear seasonality. The data does not appear to be stationary due to the lack of constant mean and autocovariance.

```
data = read.csv('erie.csv')
colnames(data) = c('Month', 'Level')
level.ts = ts(data$Level, start=c(1921,1), frequency = 12)
plot.ts(level.ts, main="Monthly Average Water Levels")
```



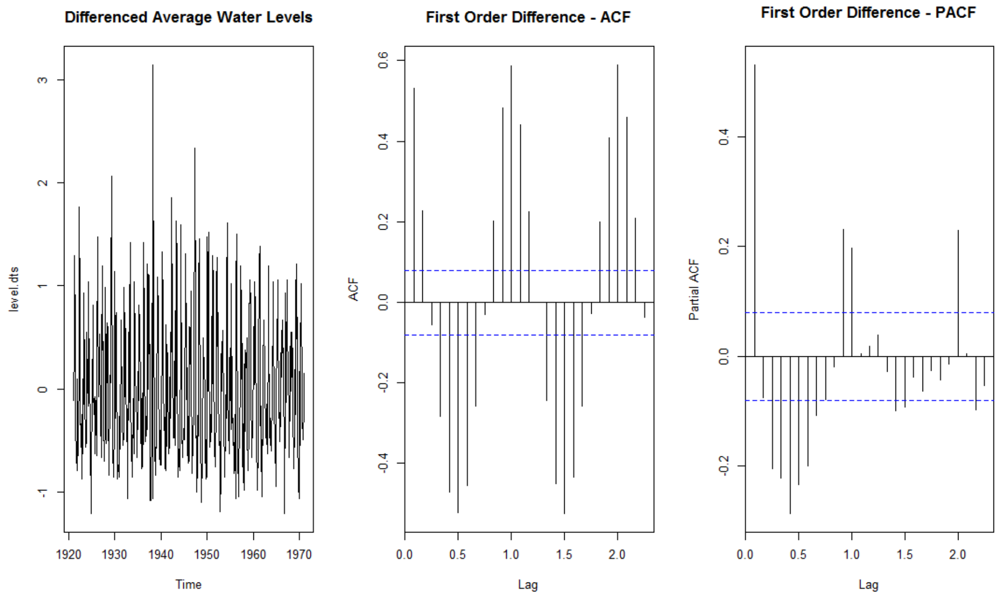
2 Estimating the Trend

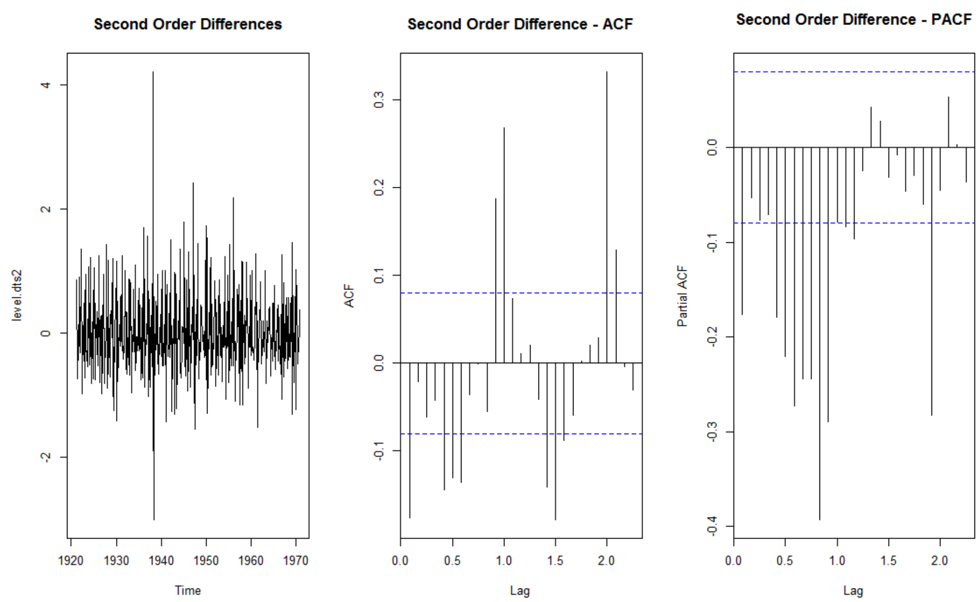
(a) Both trend and seasonality are significant, with all coefficients other than that of Hsin(4*pi*t) being significantly nonzero at the 10% level.

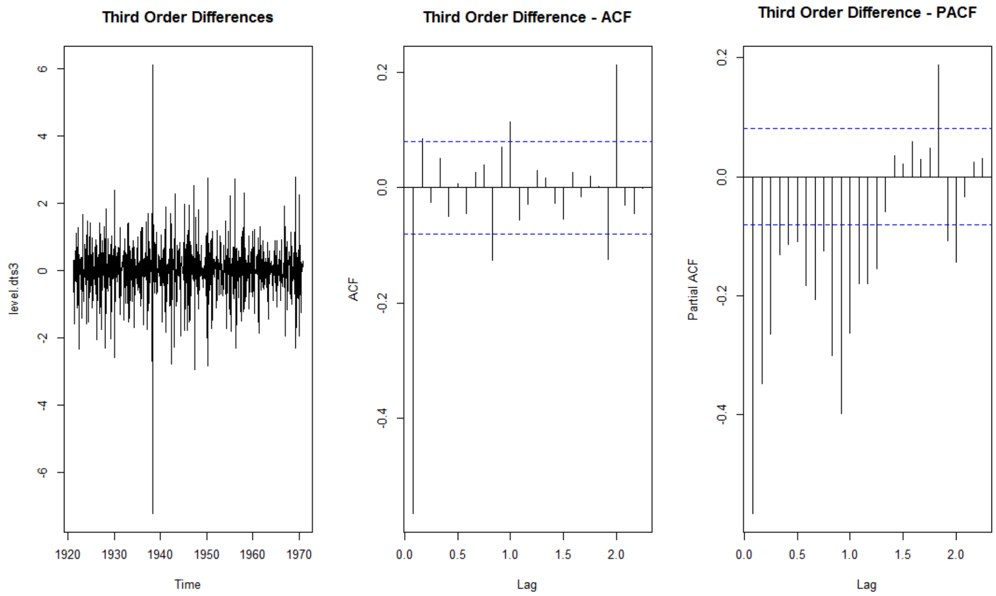
```
H = harmonic(ts(level, start = 1/12, end = 50, deltat = 1/12), 2)
time.pts = c(1:length(level))
time.pts = c(time.pts-min(time.pts))/max(time.pts)
erie.np = gam(level~s(time.pts)+H)
summary(erie.np)

Family: gaussian
Link function: identity
```

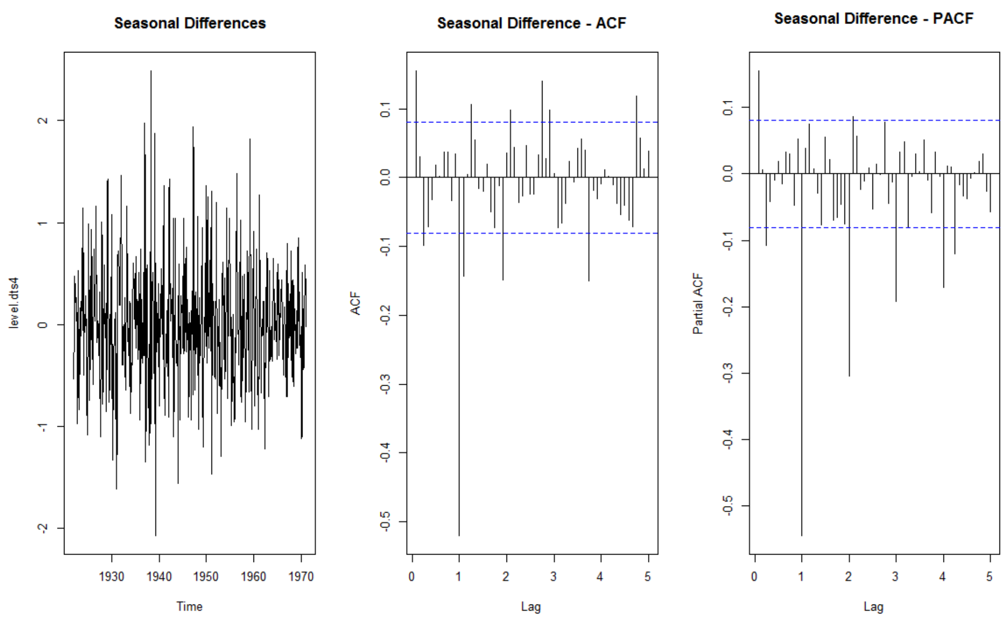
```
Formula:
level ~ s(time.pts) + H

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
```

(b) While differencing appears to remove any trend, there are still signs of seasonality. This is seen in the ACF plot displaying a pattern. At least 3rd order differncing is needed for the pattern to be mosly insignificant. Seasonality is reduced more quickly by taking 1st-order seasonal and regular differences.


Differenced Average Water Levels
First Order Difference - ACF
First Order Difference - PACF
level.dts
Time
ACF
Lag
Partial ACF
Lag


Second Order Differences
Second Order Difference - ACF
Second Order Difference - PACF
level.dts2
Time
ACF
Lag
Partial ACF
Lag





(c) It appears as though three orders of differencing are required to remove both trend and seasonality. Given the cutoff at lag 15 on the third order-differenced PACF plot, and the lack of significant lags after 1 on the 3rd-diffed ACF plot, this would suggest starting with an ARIMA(15,3,1) model – also quite complex! A seasonal ARIMA model would likely be preferred. For example, we may want to consider ARIMA(1,1,1)(0,1,1), where p and q were selected due to the large first lags in the seaosnal-differenced ACF and PACF, and P and Q were selected due to the ACF seasonal lags largely cuting off after the first and the PACF seasonal lags decaying more gradually.

3 ARIMA Modeling

(a) The model producing the lowest AIC was ARIMA(2,1,2)(0,1,1)

```
final.aic = Inf
final.arima.order = c(0,1,0)
final.seasonal.order = c(0,1,0)
for (p in 0:2){
  for (q in 0:2){
    for (P in 0:2) {
      for (Q in 0:2) {
        mod = arima(level.ts, order=c(p,1,q),
           seasonal=list(order=c(P,1,Q), period=12), method="ML")
        current.aic = AIC(mod)
  print(paste(p,q,P,Q,current.aic,sep="    "))
        if (current.aic < final.aic){
          final.aic = current.aic
          final.arima.order = c(p,1,q)
    final.seasonal.order = c(P,1,Q)
          final.arima = mod
        }
      }
    }
  }
}
```

(b) Coefficient estimates and confidence intervals are shown in the output below.

```
final.arima$coef

       ar1        ar2        ma1        ma2       sma1
 0.4672521 -0.7264566 -0.3171286  0.7588018 -0.9566066

confint(final.arima,level=.99)

          0.5 %     99.5 %
ar1   0.2583976  0.6761065
ar2  -1.0546556 -0.3982575
ma1  -0.5189220 -0.1153353
ma2   0.4231982  1.0944053
sma1 -1.0273707 -0.8858426
```

(c) The model from 3(a) has a lower AIC (641.371 versus 641.6963) and a lower BIC (667.6212 versus 667.9465) and is therefore preferred.

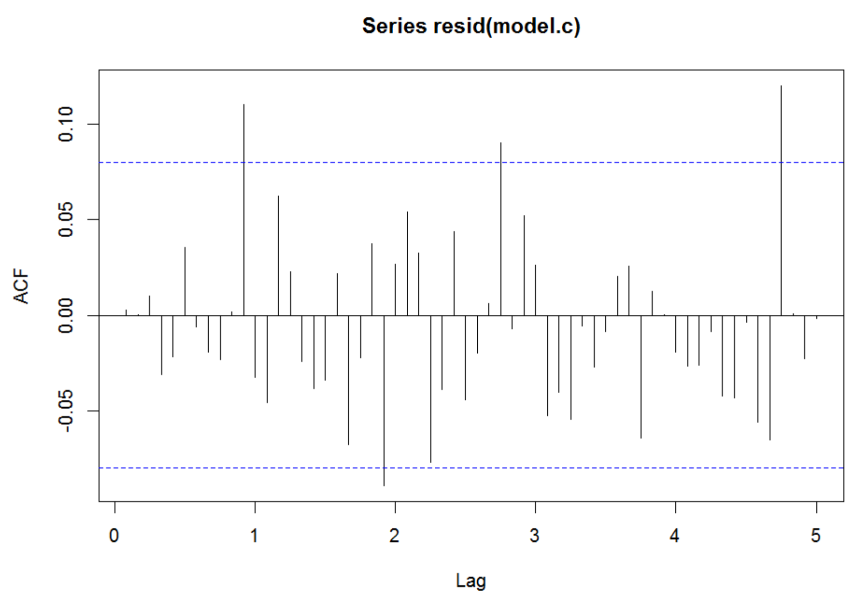
```
model.c = arima(level.ts, order=c(1,1,3),
```

```
                seasonal=list(order=c(0,1,1), period=12), method="ML")
AIC(model.c)
BIC(model.c)
AIC(final.arima)
BIC(final.arima)
```

4 Residual Analysis

(a) The ACF plot suggests no significant autocorrelation as spikes are all either insignificant or close to 0. The high p-value from the Box test confirms this.

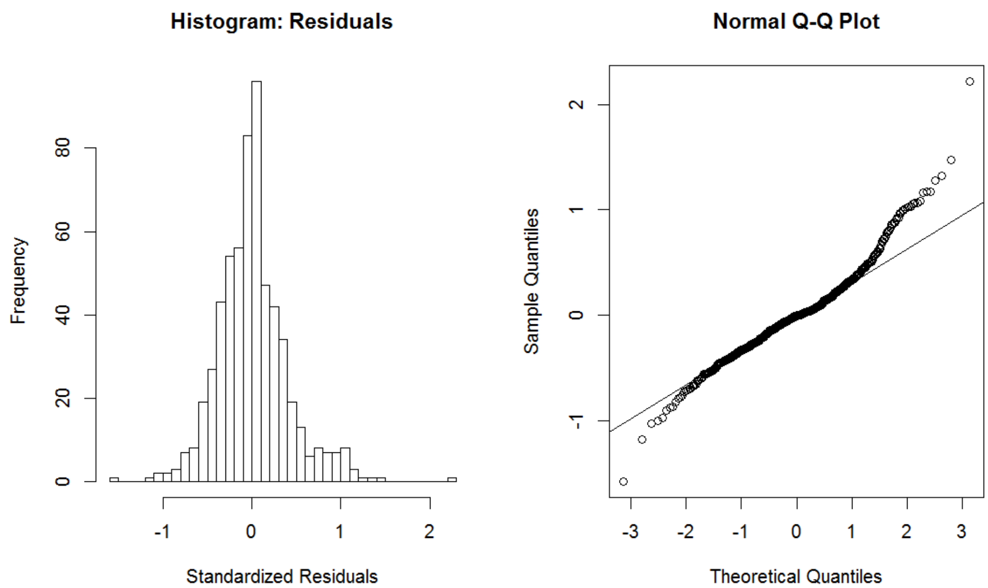
```
Box.test(resid(model.c),lag=24,type="Ljung",fitdf=5)

        Box-Ljung test

data:  resid(model.c)
X-squared = 26.019, df = 19, p-value = 0.1297

acf(resid(model.c),lag=60)
```



(b) The histogram of residuals seems slightly skewed right. The QQ plot also shows signs of non-normality. This is further confirmed by the low p-value from the Jarque Bera test.

```
jarque.bera.test(resid(model.c))

        Jarque Bera Test

data:  resid(model.c)
```

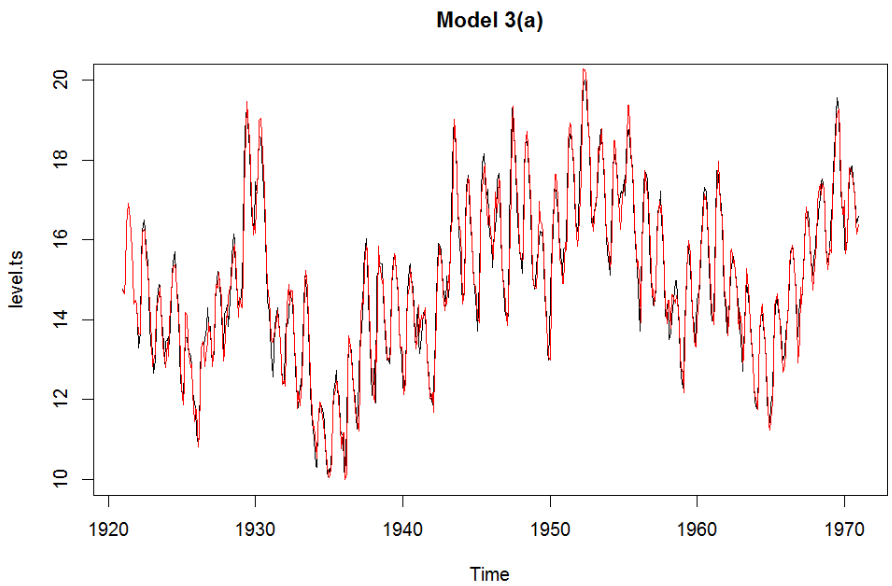
```
X-squared = 214.25, df = 2, p-value < 2.2e-16

par(mfrow=c(1,2))
hist(resid(model.c),breaks="FD",xlab='Standardized Residuals',mai
qqnorm(resid(model.c))
qqline(resid(model.c))
```

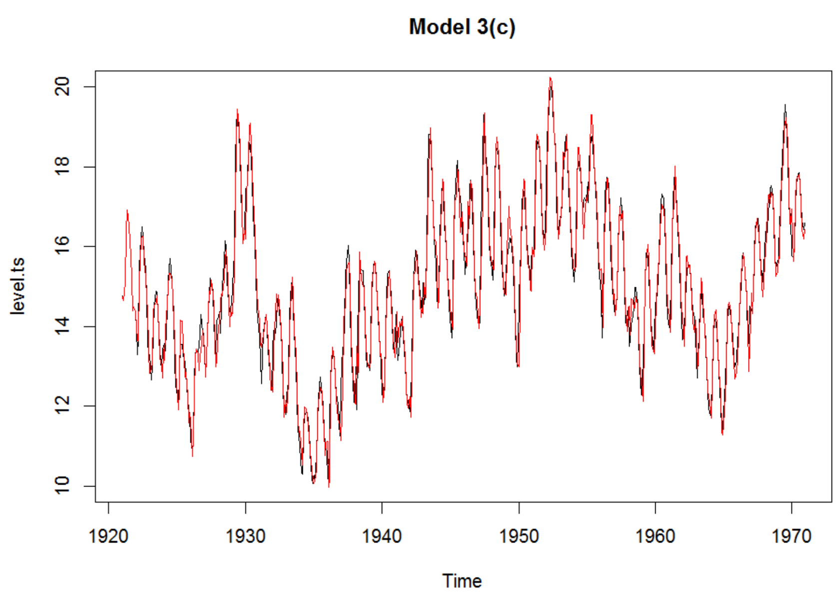


5 Model Fit

(a) The model from question 3c fits slightly better. Its MAE and PM are 0.2861072 and 0.03935116 respectively while those of the model from 3a are 0.2906788 and 0.03936394.

```
#3a model
fit=level.ts-final.arima$residuals
mae=mean(abs(fit-level.ts))
pm=sum((fit-level.ts)^2)/sum((level.ts-mean(level.ts))^2)
plot(level.ts, main = "Model 3(a)")
lines(fit, col="red")

#3c model
fit=level.ts-model.c$residuals
mae=mean(abs(fit-level.ts))
pm=sum((fit-level.ts)^2)/sum((level.ts-mean(level.ts))^2)
plot(level.ts, main = "Model 3(c)")
lines(fit, col="red")
```





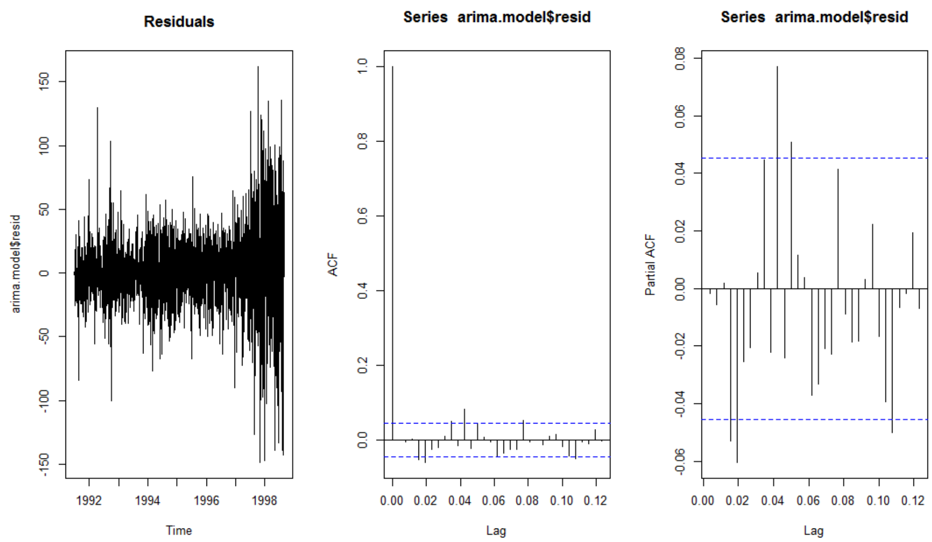
II Stock Index

1 Residual Analysis

(a) The residual plot shows signs of heteroskedasticity, with data after 1997 having larger variability than earlier data.

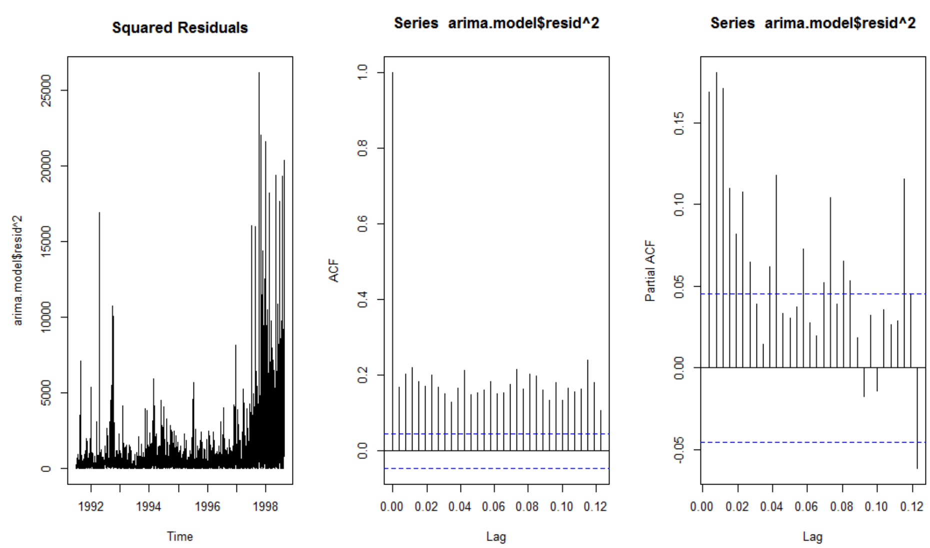
```
par(mfrow=c(1,3))
plot(arima.model$resid, main="Residuals")
acf(arima.model$resid)
```

```
pacf(arima.model$resid)
```



(b) The squared residuals show more signs of heteroskedasticity, with the correlations in the ACF plot all being significant.

```
par(mfrow=c(1,3))
plot(arima.model$resid^2, main="Squared Residuals")
acf(arima.model$resid^2)
pacf(arima.model$resid^2)
```



2 GARCH Modeling

(a) The selected model is ARIMA(0,2,2)-GARCH(1,1).

```
FTSE.diff2 = diff(diff(FTSE))
final.bic = Inf
```

```
final.order.garch = c(0,0)
for (m in 0:3) for (n in 0:3) {
  spec = ugarchspec(variance.model=list(garchOrder=c(m,n)),
                    mean.model=list(armaOrder=c(0,2),
                    include.mean=T),distribution.model="std")
  fit = ugarchfit(spec, FTSE.diff2, solver = "hybrid")
  current.bic = infocriteria(fit)[2]
  if (current.bic < final.bic) {
    final.bic = current.bic
    final.order.garch = c(m, n)
  }
}
```

(b) The chosen GARCH model does have slightly lower AIC and BIC compared to the ARIMA(0,2,2)-GARCH(0,0) model, indicating a better fit. However, its fitted values are comparable to those from the ARIMA-only model, with slightly higher mean absolute and mean squared errors. The more simple ARIMA-only model may be preferred.

```
infocriteria(fit)

Akaike       9.370454
Bayes        9.391278
Shibata      9.370426
Hannan-Quinn 9.378129

mean(residuals(fit)^2)
[1] 937.9557
mean(abs(residuals(fit)))
[1] 21.86196

spec0 = ugarchspec(variance.model=list(garchOrder=c(0,0)),
                     mean.model=list(armaOrder=c(0,2),
                     include.mean=T),distribution.model="std")
fit0 = ugarchfit(spec0, FTSE.diff2, solver = "hybrid")
infocriteria(fit0)

Akaike       9.545082
Bayes        9.559956
Shibata      9.545067
Hannan-Quinn 9.550564

mean(riduals(fit0)^2)
[1] 936.2503
```

```
mean(abs(residuals(fit0)))
[1] 21.85999
```

3 Refine Order Selection

(a) If we're concerned only with minimizing BIC, the final model we select is ARIMA(2,2,2)-GARCH(1,1). However, the ARIMA(0,2,2)-GARCH(1,1) model performs similarly and is a more simple model.

```
final.bic = Inf
final.order.arma2 = c(0,0)
for (p in 0:3) for (q in 0:3) {
  spec2 = ugarchspec(variance.model=list(garchOrder=c(final.order
                  mean.model=list(armaOrder=c(p, q),
                  include.mean=T),distribution.model="std")
  mod2 = ugarchfit(spec2, FTSE.diff2, solver = "hybrid")
  current.bic = infocriteria(mod2)[2]
  if (current.bic < final.bic) {
    final.bic = current.bic
    final.order.arma2 = c(p, q)
    fit2 = mod2
  }
}
final.order.arma2
[1] 2 2

final.bic = Inf
final.order.garch2 = c(0,0)
for (m in 0:3) for (n in 0:3){
  spec3 = ugarchspec(variance.model=list(garchOrder=c(m,n)),
                  mean.model=list(armaOrder=c(final.order.arma2
                  include.mean=T),distribution.model="std")
  mod3 = ugarchfit(spec3, FTSE.diff2, solver = "hybrid")
  current.bic = infocriteria(mod3)[2]
  if (current.bic < final.bic){
    final.bic = current.bic
    final.order.garch2 = c(m,n)
    fit3 = mod3
  }
}
final.order.garch2
[1] 1 1

infocriteria(fit3)
```

```
Akaike       9.350262
Bayes        9.377036
Shibata      9.350216
Hannan-Quinn 9.360130
```

(b) If Y_t represents the 2nd-order differenced FTSE data, the ARIMA(2,2,2)-GARCH(1,1) model is

$$Y_t = -0.0005 - 0.773Y_{t-1} + 0.069Y_{t-2} + Z_t - 0.139Z_{t-1} - 0.852Z_{t-2}$$

where

$$Z_t = \sigma_t R_t$$

$$\sigma_t^2 = 2.968 + 0.038Z_{t-1}^2 + 0.960\sigma_{t-1}^2$$

$$R_t \sim N(0,1)$$

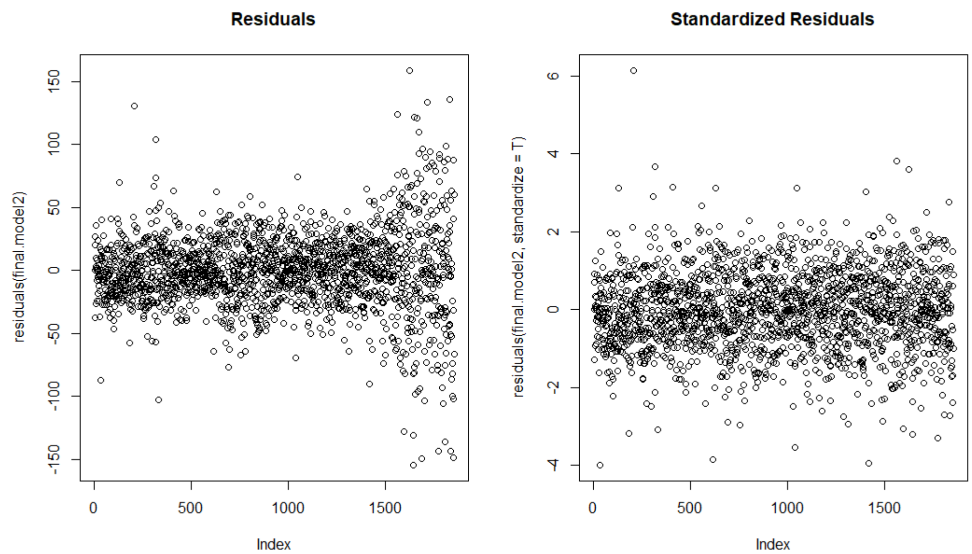
```
final.model2 = garchFit(~arma(2,2)+garch(1,1),data=FTSE.diff2,tra
summary(final.model2)

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(2, 2) + garch(1, 1), data = FTSE.diff2,
    trace = F)

Mean and Variance Equation:
 data ~ arma(2, 2) + garch(1, 1)
<environment: 0x000000001ce636b8>
 [data = FTSE.diff2]

Conditional Distribution:
 norm

Coefficient(s):
        mu          ar1          ar2          ma1          ma2
-0.00053077  -0.77328568   0.06866243  -0.13908118  -0.85151288
     alpha1        beta1
 0.03833442   0.95959284

Std. Errors:
 based on Hessian

Error Analysis:
```

```
         Estimate  Std. Error  t value Pr(>|t|)
mu     -0.0005308   0.0060386   -0.088  0.92996
ar1    -0.7732857   0.0833851   -9.274  < 2e-16 ***
ar2     0.0686624   0.0242586    2.830  0.00465 **
ma1    -0.1390812   0.0790472   -1.759  0.07850 .
ma2    -0.8515129   0.0804391  -10.586  < 2e-16 ***
omega   2.9680999   1.5564764    1.907  0.05653 .
alpha1  0.0383344   0.0070249    5.457 4.84e-08 ***
beta1   0.9595928   0.0075479  127.134  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 -8712.033    normalized:  -4.688931

Description:
 Mon Nov 19 21:15:33 2018 by user: Man1


Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  139.4056  0
 Shapiro-Wilk Test  R    W      0.9912023 3.753583e-09
 Ljung-Box Test     R    Q(10)  7.942167  0.634486
 Ljung-Box Test     R    Q(15)  17.7113   0.2781475
 Ljung-Box Test     R    Q(20)  23.4306   0.2681491
 Ljung-Box Test     R^2  Q(10)  4.268061  0.9344451
 Ljung-Box Test     R^2  Q(15)  9.69575   0.8384291
 Ljung-Box Test     R^2  Q(20)  13.5149   0.8542159
 LM Arch Test       R    TR^2   7.449436  0.8265416

Information Criterion Statistics:
     AIC      BIC      SIC     HQIC
9.386473 9.410272 9.386436 9.395244
```

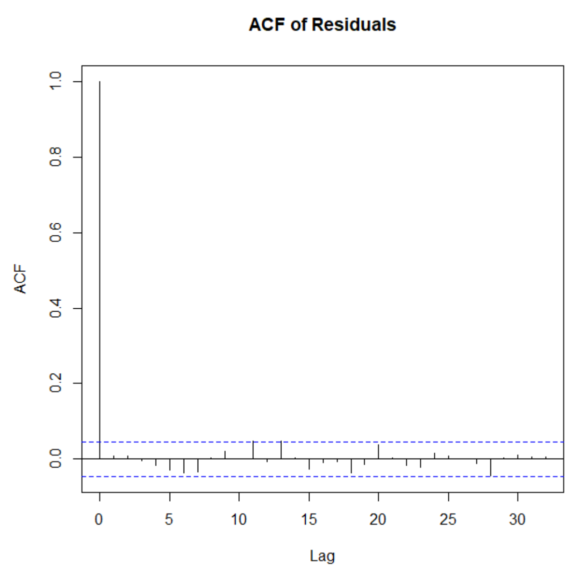
4 Residual Analysis, Revisited

(a) While the residual plot still shows signs of heteroskedasticity, it is reduced when looking at the standardized residuals.

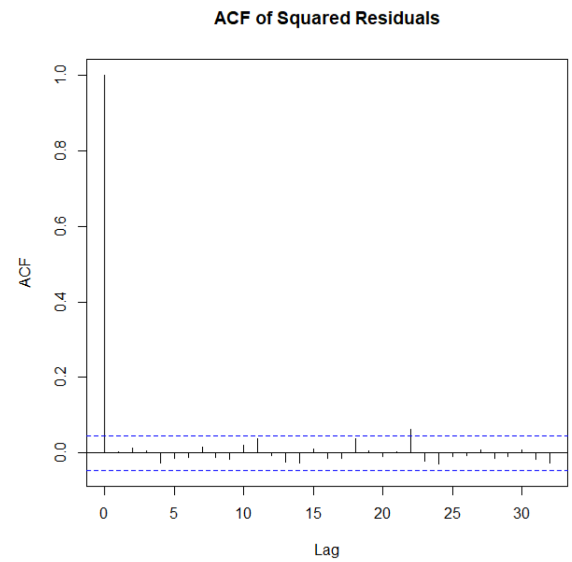
```
par(mfrow=c(1,2))
plot(residuals(final.model2),main="Residuals")
plot(residuals(final.model2,standardize=T),main="Standardized Res
```



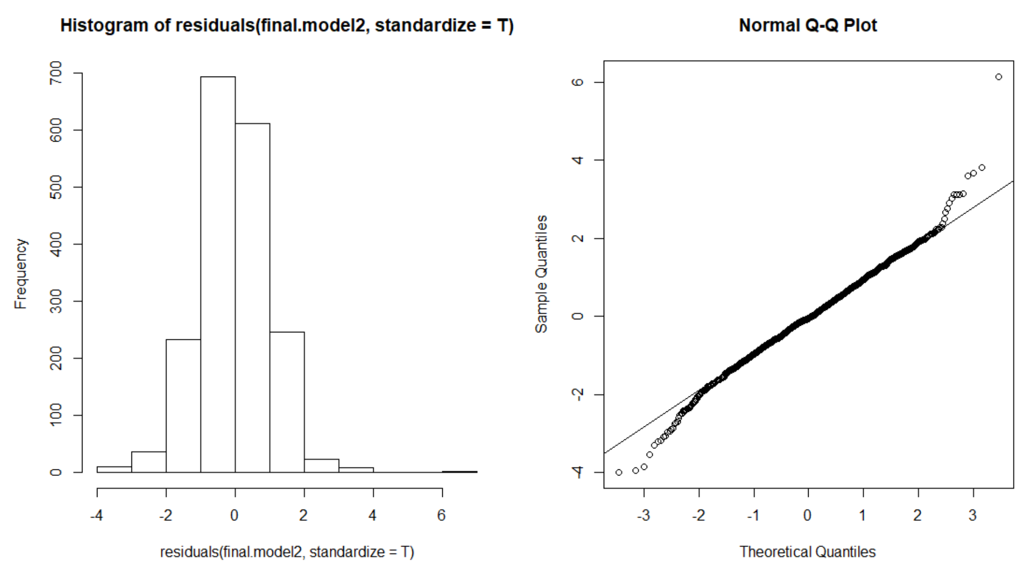
(b) They do not display significant autocorrelation. The ACF plot does not show signficant autocorrelations, and the Ljung-Box Test results in high p-values.



(c) They do not display significant autocorrelation. The ACF plot does not show signficant autocorrelations, and the Ljung-Box Test results in high p-values.

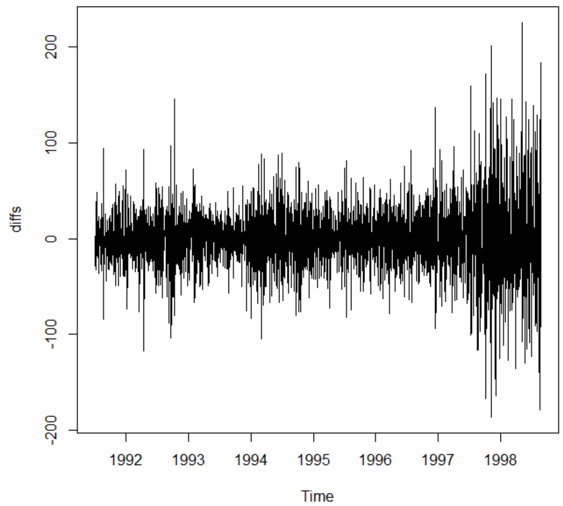


(d) The standardized residuals don't seem to follow a normal distribution. The Jarque-Bera Test has a p-value close to 0, and the QQ plot shows signs of non-normality near the tails



5 Model Fit

(a) Due to the heteroskedasticity, we would expect the ARIMA(2,2,2)-GARCH(1,1) model to fit better.



(b) The ARIMA model fits better based on MAE.

```
model1.fitted = fitted(model1)
model2.fitted = model2@fitted
mean(abs(model1.fitted - diffs))
[1] 21.65011
mean(abs(model2.fitted - diffs))
[1] 21.7453
```