Problem_Set_3

Kristopher C. Toll

April 3, 2018

Theory

Problem 1

$$Y_t = e_t = \sigma_t Z_t$$

where $Z_t \sim$ iid $N(0,1)$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2$$

with $\alpha_0 > 0, 0 < \alpha_1 < 1$ and Z_t and $Y_{t-j}, j\epsilon\mathbb{N}$, independent for all $t\epsilon\mathbb{Z}$

$$Y_t^2 = (\sigma_t Z_t)^2 = \sigma_t^2 Z_t^2 = (\alpha_0 + \alpha_1 Y_{t-1}^2) Z_t^2 = \alpha_0 Z_t^2 + \alpha_1 Z_t^2 Y_{t-1}^2$$

$$Y_{t-1}^2 = \sigma_{t-1}^2 Z_{t-1}^2$$

$$\sigma_{t-1}^2 = \alpha_0 + \alpha_1 Y_{t-2}^2$$

$$Y_{t-1}^2 = (\alpha_0 + \alpha_1 Y_{t-2}^2) Z_{t-1}^2$$

so,

$$Y_t^2 = \alpha_0 Z_t^2 + \alpha_1 Z_t^2 (\alpha_0 + \alpha_1 Y_{t-2}^2) Z_{t-1}^2 = \alpha_0 Z_t^2 + \alpha_0 \alpha_1 Z_{t-1}^2 + \alpha_1^2 Y_{t-2}^2 Z_t^2 Z_{t-1}^2$$

$$Y_{t-2}^2 = (\alpha_0 + \alpha_1 Y_{t-3}^2) Z_{t-2}^2$$

so,

$$Y_t^2 = \alpha_0 Z_t^2 + \alpha_0 \alpha_1 Z_{t-1}^2 + \alpha_1^2 (\alpha_0 + \alpha_1 Y_{t-3}^2) Z_{t-2}^2 Z_t^2 Z_{t-1}^2$$
$$= \alpha_0 Z_t^2 + \alpha_0 \alpha_1 Z_{t-1}^2 + \alpha_0 \alpha_1^2 Z_t^2 Z_{t-1}^2 Z_{t-2}^2 + \alpha_1^3 Y_{t-3}^2 Z_t^2 Z_{t-1}^2 Z_{t-2}^2$$

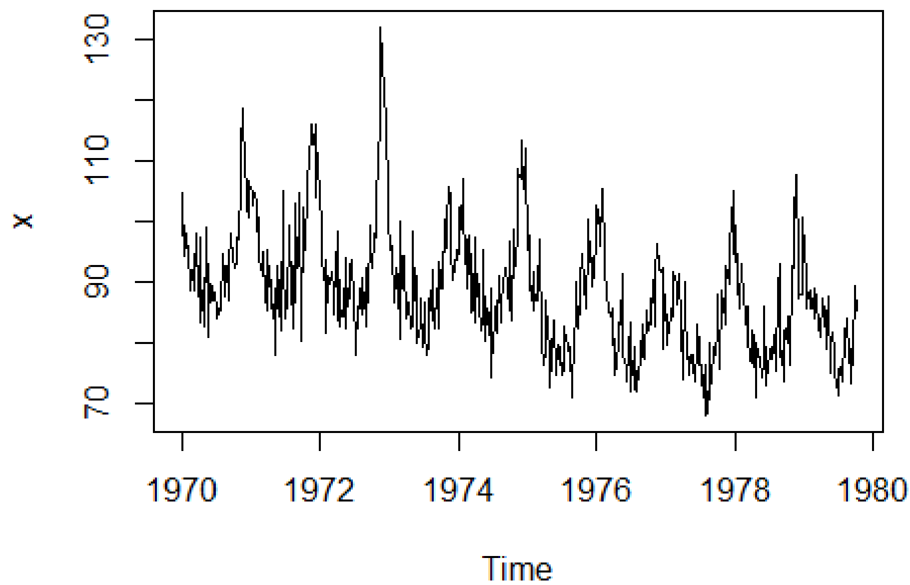
$$Y_t^2 = \alpha_0 \sum_{i=0}^{n} (\, \alpha_1^i Z_t^2 Z_{t-1}^2 Z_{t-2}^2 \cdots Z_{t-i}^2) + \alpha_1^{n+1} Y_{t-n-1}^2 Z_t^2 Z_{t-1}^2 Z_{t-2}^2 \cdots Z_{t-n}^2$$

Application

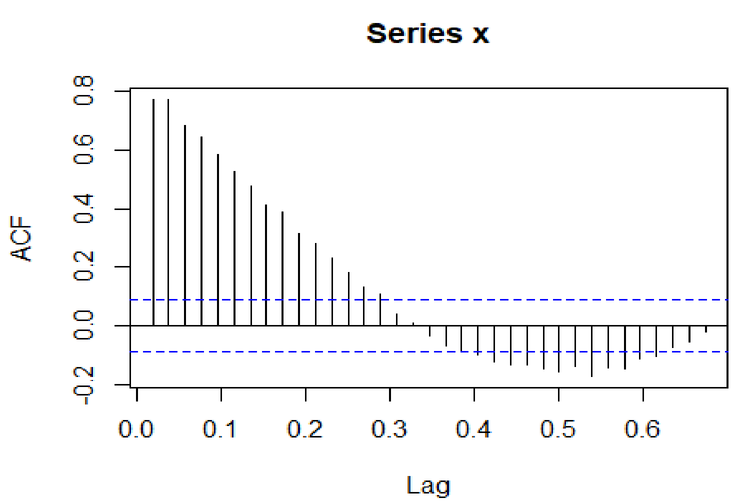
Problem 1

Part A

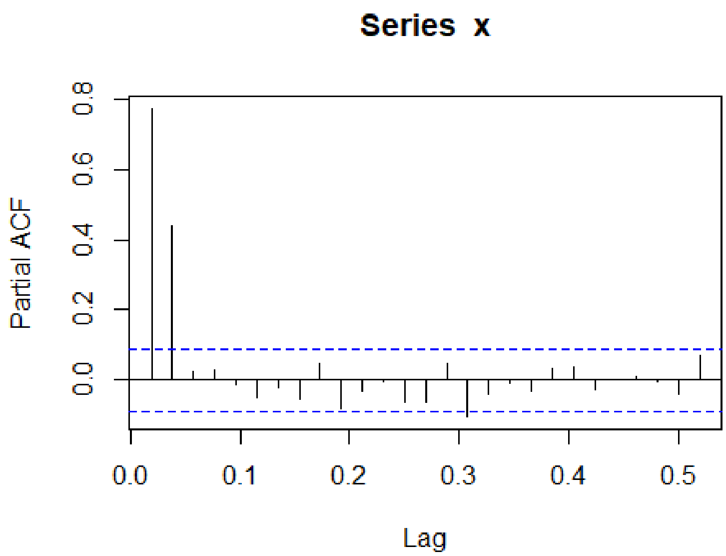
```
x <- cmort
plot(x)
```




```
acf(x, lag.max = 35)
```



```r
pacf(x)
```




```r
adf.test(x, alternative = "stationary",k = 0)

## Warning in adf.test(x, alternative = "stationary", k = 0): p-value smaller
## than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  x
## Dickey-Fuller = -9.2037, Lag order = 0, p-value = 0.01
## alternative hypothesis: stationary

adf.test(x, alternative = "stationary")

## Warning in adf.test(x, alternative = "stationary"): p-value smaller than
## printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  x
## Dickey-Fuller = -5.4125, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

By examining the plot of the data we can see a seasonal pattern. In the ACF plot, we see that there are significant spikes that are declining at a constant rate toward zero and may even develop a sine wave. From the PACF we can see that there are two significant spikes. This suggests that this data follows an AR(2) process. By running the Dickey-Fuller Test we can see that this is indeed a stationary dataset.

Part b

```r
lag2reg_ols <- ar.ols(x, order = 2, demean = FALSE, intercept = TRUE)
lag2reg_mle <- arima(x, order = c(2,0,0))
print(lag2reg_mle)
```

```
##
## Call:
## arima(x = x, order = c(2, 0, 0))
##
## Coefficients:
##          ar1     ar2  intercept
##       0.4301  0.4424    88.8538
## s.e.  0.0397  0.0398     1.9407
##
## sigma^2 estimated as 32.37:  log likelihood = -1604.71,  aic = 3215.43
```

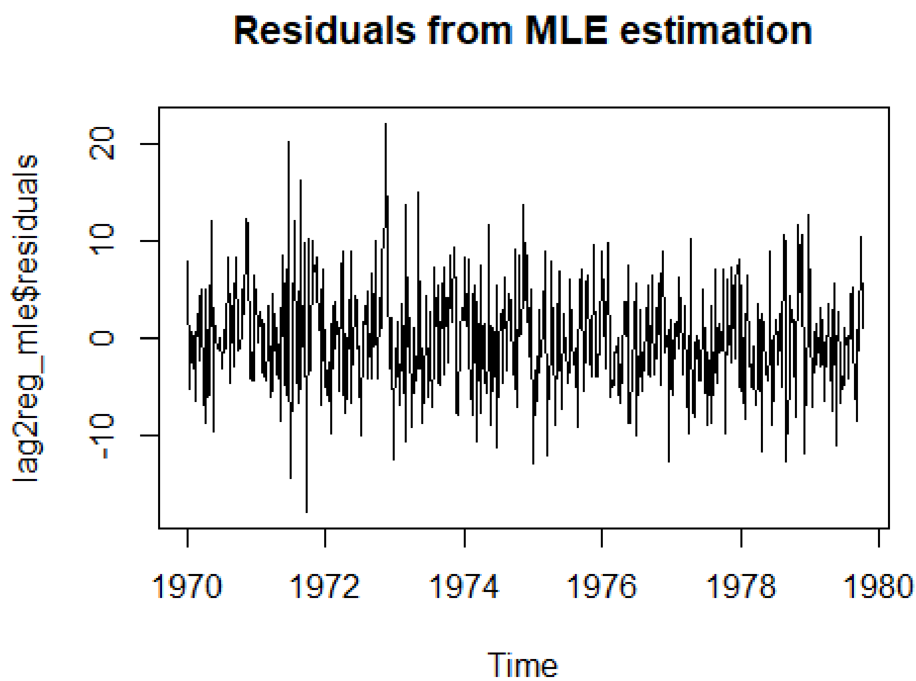
```r
print(lag2reg_ols)
```

```
##
## Call:
## ar.ols(x = x, order.max = 2, demean = FALSE, intercept = TRUE)
##
## Coefficients:
##      1       2
## 0.4286  0.4418
##
## Intercept: 11.45 (2.394)
##
## Order selected 2  sigma^2 estimated as  32.32
```

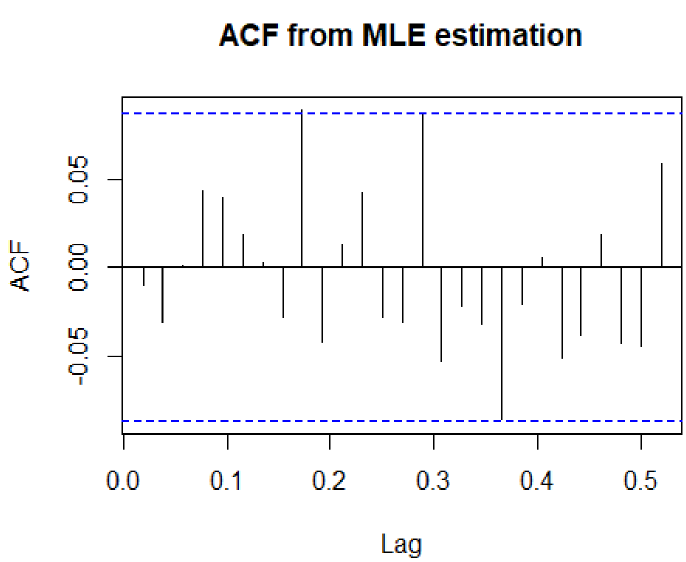
We can see that the OLS estimate for the intercept is much larger than the MLE estimate. Other than that, the coefficient estimates are almost the same. Even the MSE for the MLE and OLS models are very similar in value.

c

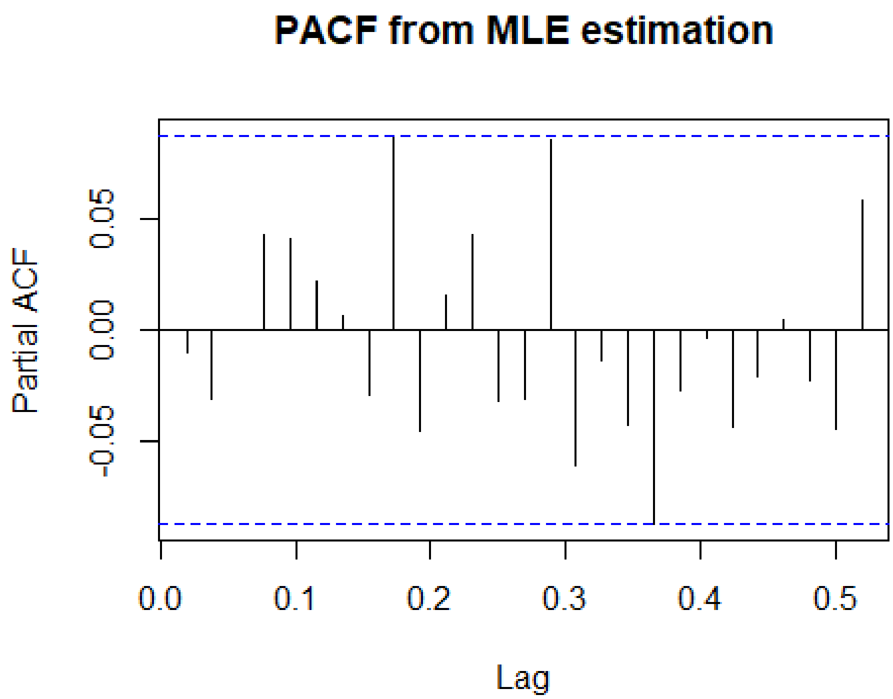
```r
plot(lag2reg_mle$residuals, main = "Residuals from MLE estimation")
```



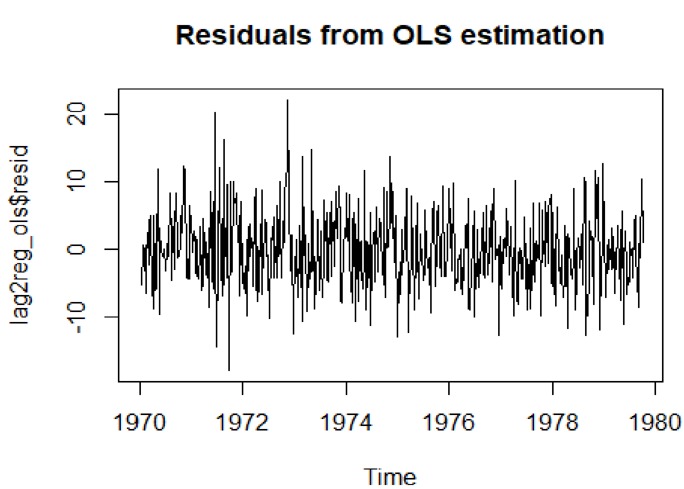

```r
acf(lag2reg_mle$residuals, main = "ACF from MLE estimation")
```



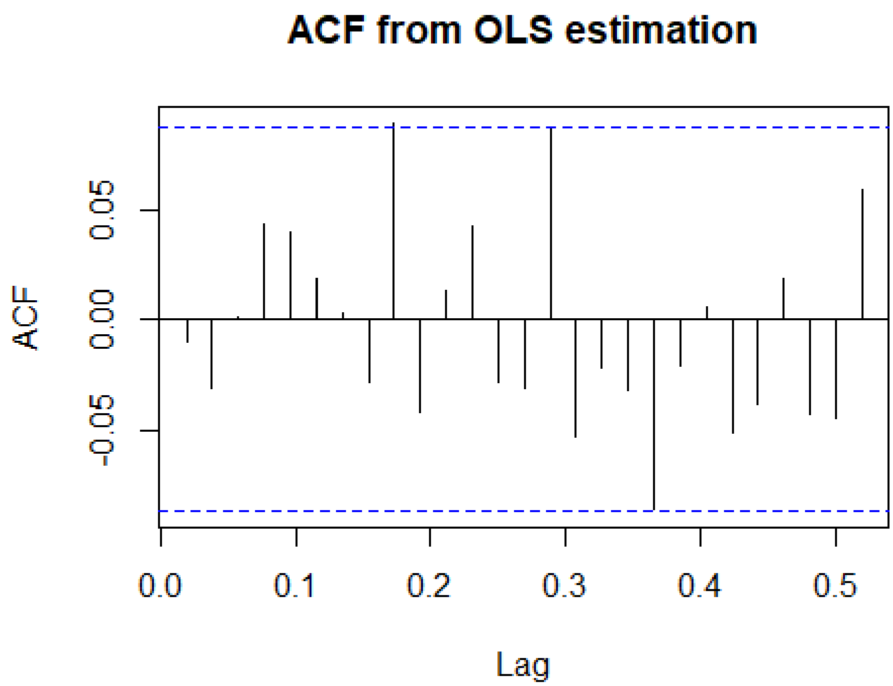
```r
pacf(lag2reg_mle$residuals, main = "PACF from MLE estimation")
```



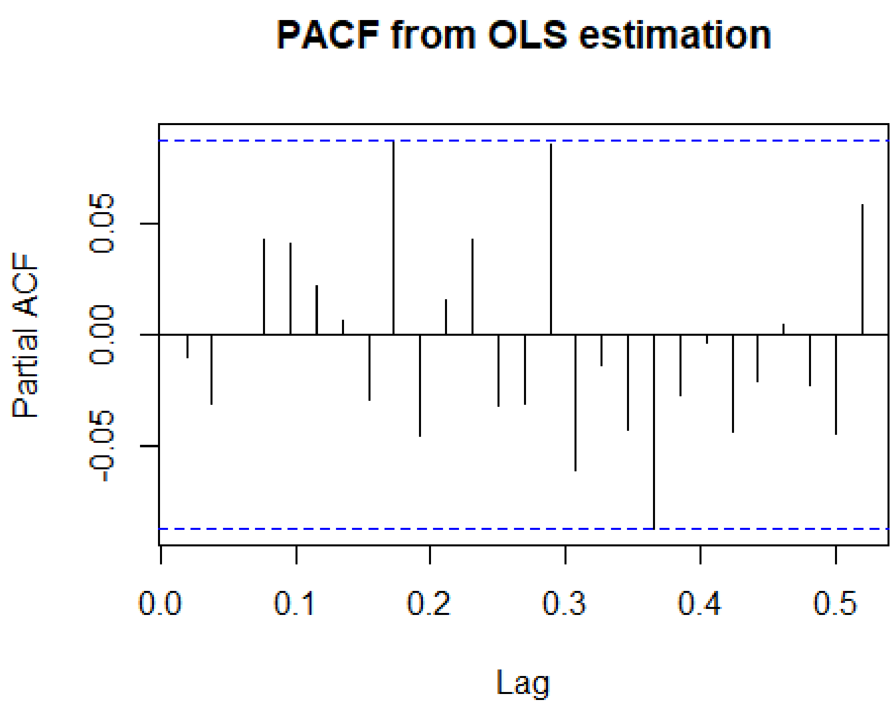

```r
Box.test(lag2reg_mle$residuals, lag = log(508), type = "Ljung")

##
##  Box-Ljung test
##
## data:  lag2reg_mle$residuals
## X-squared = 2.4835, df = 6.2305, p-value = 0.8863

plot(lag2reg_ols$resid, main = "Residuals from OLS estimation")
```



```r
acf(lag2reg_mle$resid, main = "ACF from OLS estimation")
```




```r
pacf(lag2reg_mle$resid, main = "PACF from OLS estimation")
```

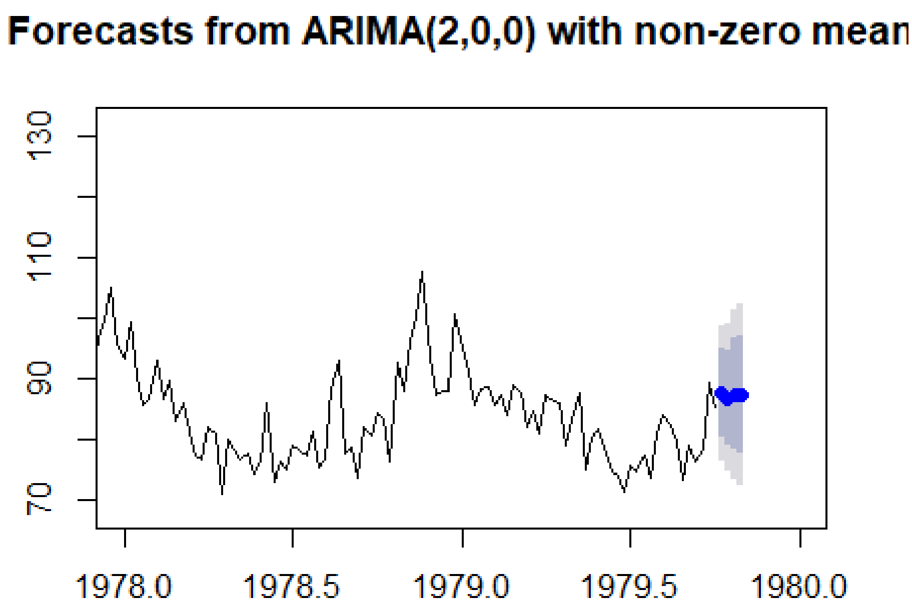


```r
Box.test(lag2reg_mle$resid, lag = log(508), type = "Ljung")

##
##  Box-Ljung test
##
## data:  lag2reg_mle$resid
## X-squared = 2.4835, df = 6.2305, p-value = 0.8863
```

Whether we are examining the MLE model or the OLS model, we can see that there is no autocorrelation in the residuals and that they follow a White noise process. This is confirmed by the box test which shows that there is no autocorrelation and we see no significant spikes in any of ACF or PACF plots. The AR(2) process models the data well.

Part d

```r
predict <- forecast(lag2reg_mle, h = 4)
plot(predict, xlim = c(1978, 1980))
```

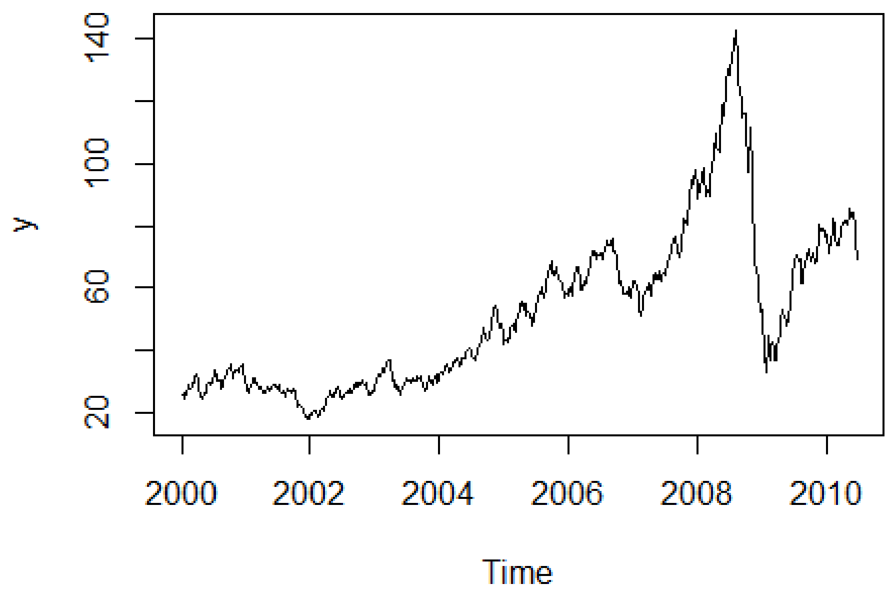


For this plot, I limited the x-axis from 1978 to 1980. Looking ahead we can see that the foretasted mortality rate falls within an expected range. It does not seem unusual given the past data. In other words, death is on the forecast horizon for at least four periods.

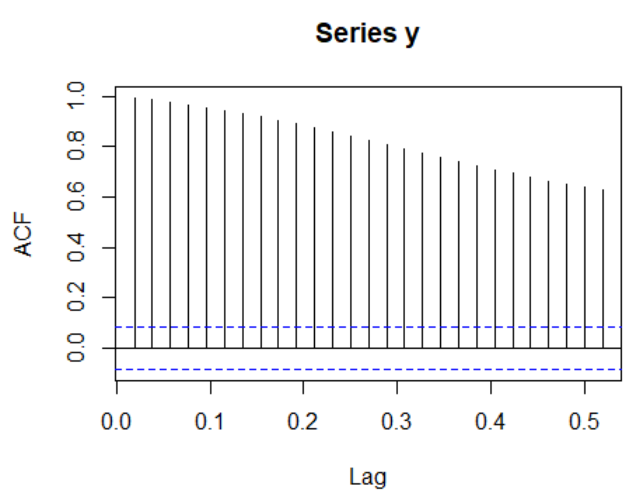
Problem 2

Part a

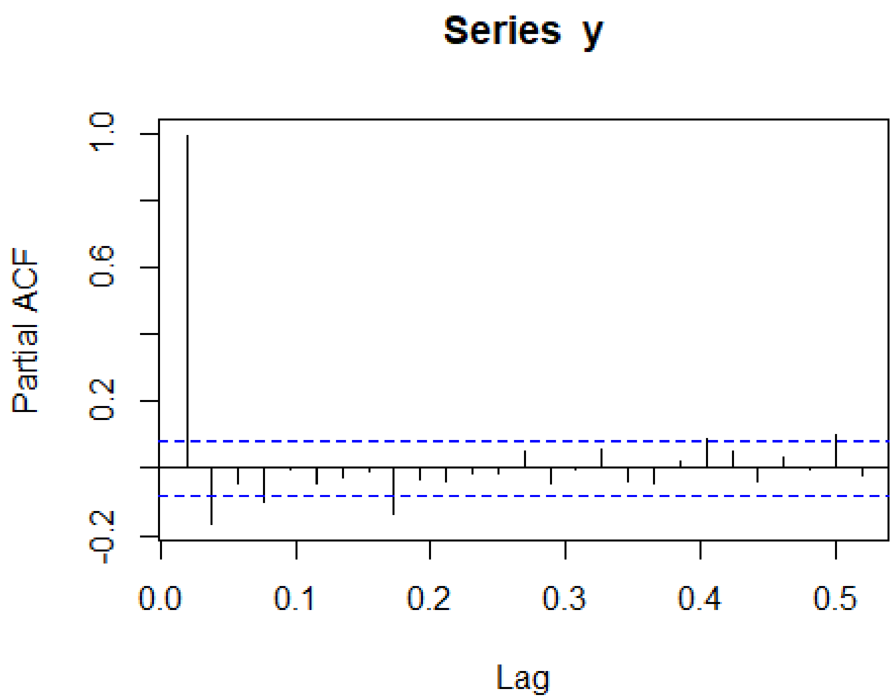
```
y <- oil
plot(y)
```




```
acf(y)
```



```r
pacf(y)
```




```r
adf.test(y, alternative = "stationary", k = 0)

##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -1.7719, Lag order = 0, p-value = 0.6749
## alternative hypothesis: stationary

adf.test(y, alternative = "stationary")

##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -3.4217, Lag order = 8, p-value = 0.04983
## alternative hypothesis: stationary
```

Looking at the ACF plot, we see that there are significant lags in the ACF plot which follow a high persistent pattern. This provides some evidence for a random walk model. The results from the augmented dicky fuller test show that the data is stationary but it is barely significant at the 0.05 level. Running a standard dicky fuller test, we find that the data is non-stationary. Differencing may be needed to correct for this. In part b we pick a differenced model as the best fit.

Part b

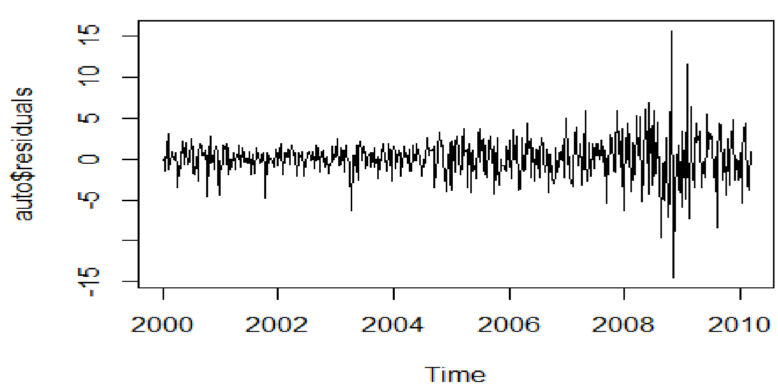
```
y.est <- window(y, end = c(2010, 10))
y.prdct <- window(y, start = c(2010, 11))
auto <- auto.arima(y.est, ic="bic", seasonal =TRUE, stepwise = TRUE)
summary(auto)
```

```
## Series: y.est
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1      ma1
##       0.8856  -0.7887
## s.e.  0.0512   0.0663
##
## sigma^2 estimated as 6.418:  log likelihood=-1241.41
## AIC=2488.81   AICc=2488.86   BIC=2501.63
##
## Training set error measures:
##                      ME     RMSE      MAE        MPE     MAPE      MASE
## Training set 0.04754489 2.526207 1.766768 0.06865754 3.538999 0.1017251
##                    ACF1
## Training set 0.04401942
```

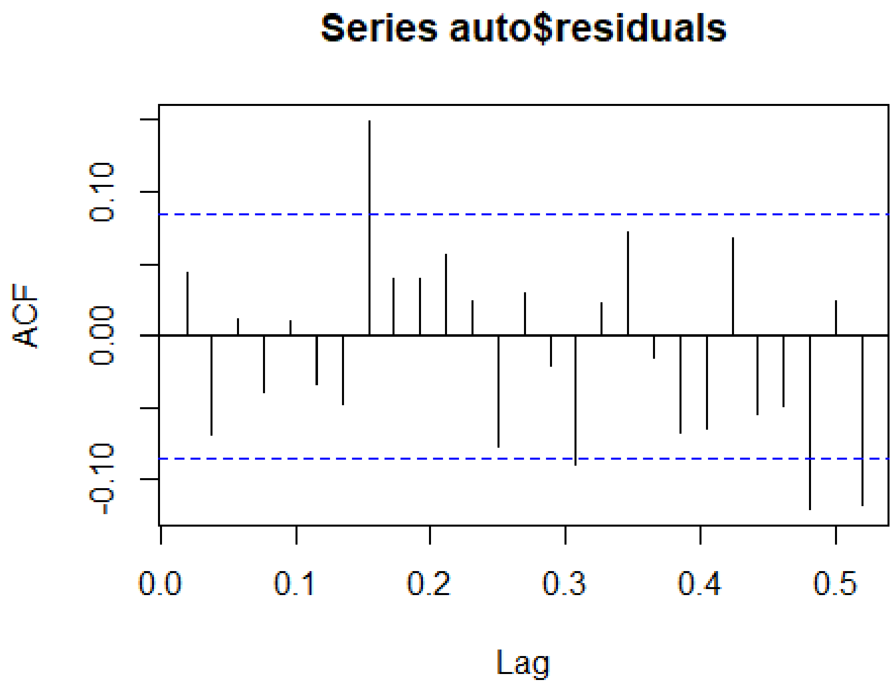
By BIC criteria we will choose an ARIMA(1,1,1) which includes differencing the data. This works around the random walk process. Interestingly, if we go by AIC, we will have a model that has a seasonal component.

Part c

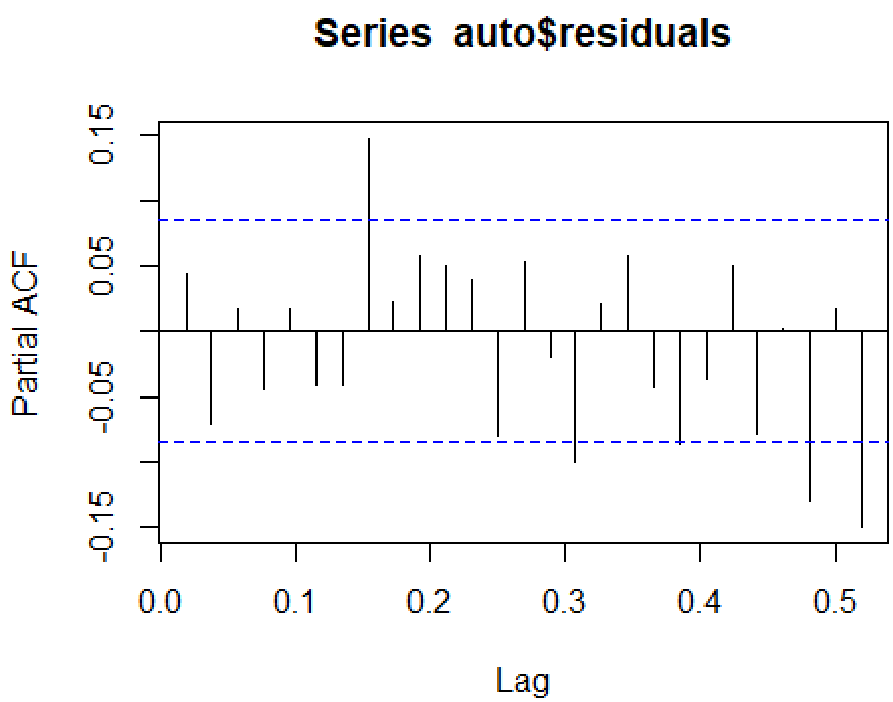
```
plot(auto$residuals)
```



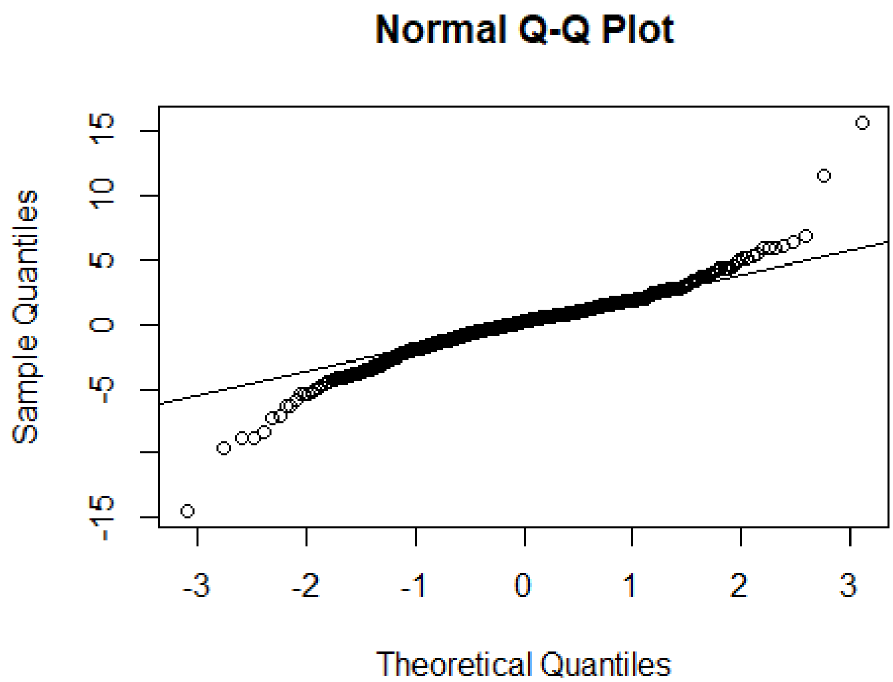
```r
acf(auto$residuals)
```




```r
pacf(auto$residuals)
```



```r
adf.test(auto$residuals, alternative = "stationary", k = 0)

## Warning in adf.test(auto$residuals, alternative = "stationary", k = 0): p-
## value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  auto$residuals
## Dickey-Fuller = -21.944, Lag order = 0, p-value = 0.01
## alternative hypothesis: stationary

adf.test(auto$residuals, alternative = "stationary")

## Warning in adf.test(auto$residuals, alternative = "stationary"): p-value
## smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  auto$residuals
## Dickey-Fuller = -6.9018, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary

Box.test(auto$residuals, lag = log(545), type = "Ljung")

##
##  Box-Ljung test
##
## data:  auto$residuals
## X-squared = 5.0495, df = 6.3008, p-value = 0.5741

qqnorm(auto$residuals)
qqline(auto$residuals)
```

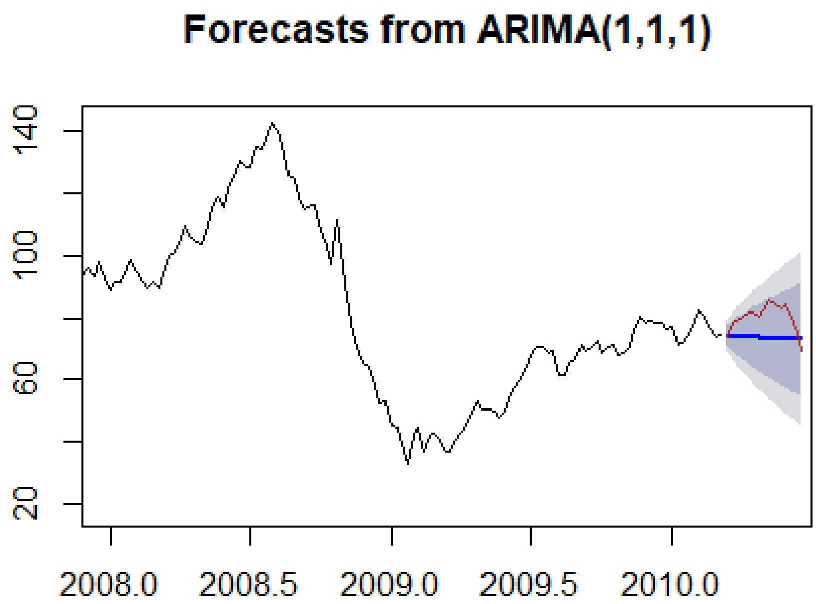



```
shapiro.test(auto$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  auto$residuals
## W = 0.93127, p-value = 7.013e-15
```

We can see from the Augmented Dicky Fuller test that the data is stationary. However, in examining the ACF Plots, we do have a significant spike around the eighth lag or so. It is possible that a seasonal component is missed in the ARIMA(1,1,1) model. Our tests for normality provide evidence that our data is skewed. It may suffer from conditional heteroscedasticity.

Part d

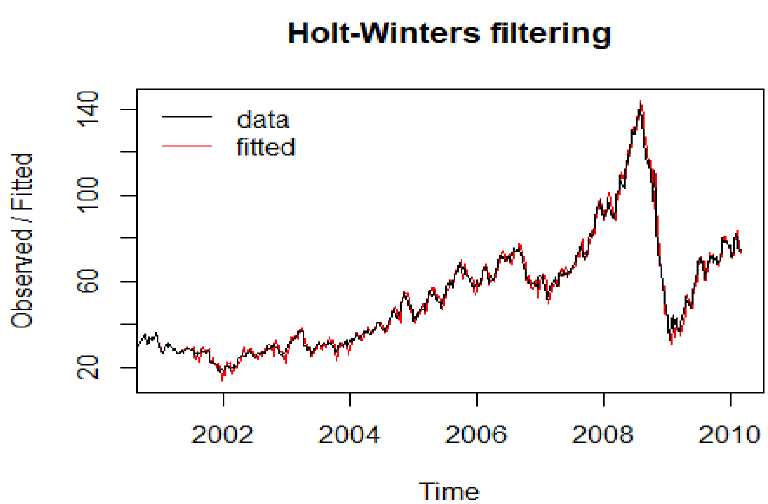
```
predict.2 <- forecast(auto, h = 15)
plot(predict.2, xlim = c(2008, 2010.4))
lines(y.prdct, col = "firebrick")
```



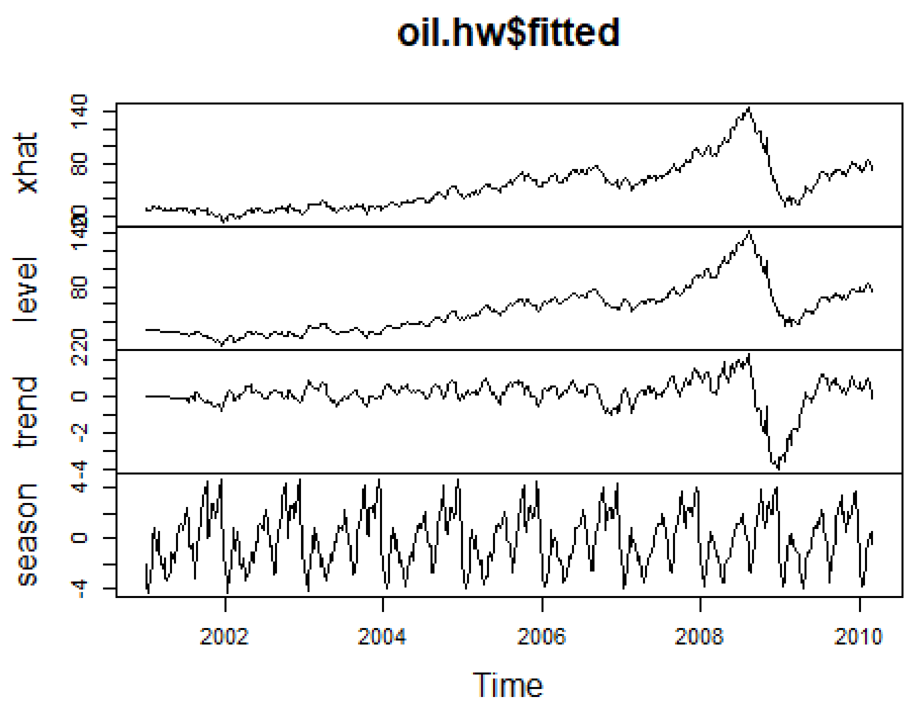
The predicted sample falls within the foretasted confidence interval. This model seems to work reasonably well.

Part e

```
oil.hw <- HoltWinters(y.est)
plot(oil.hw, col = 1:2)
legend("topleft", c("data", "fitted"), col = c(1,2), lty = 1, bty = "n")
```



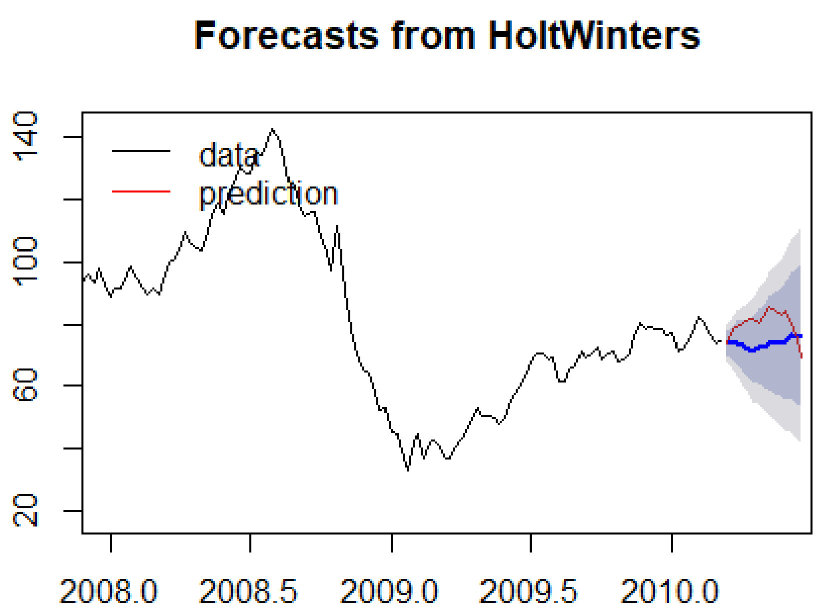
```r
plot(oil.hw$fitted)
```




```r
print(oil.hw)

## Holt-Winters exponential smoothing with trend and additive seasonal
component.
##
## Call:
## HoltWinters(x = y.est)
##
## Smoothing parameters:
##  alpha: 0.926495
##  beta : 0.0910864
##  gamma: 1
##
```

The exponential smoothing model does seem to capture seasonal trends that may have been missed in the ARIMA(1,1,1). Looking at the plot it is pleasing to the eye so the fit is good.

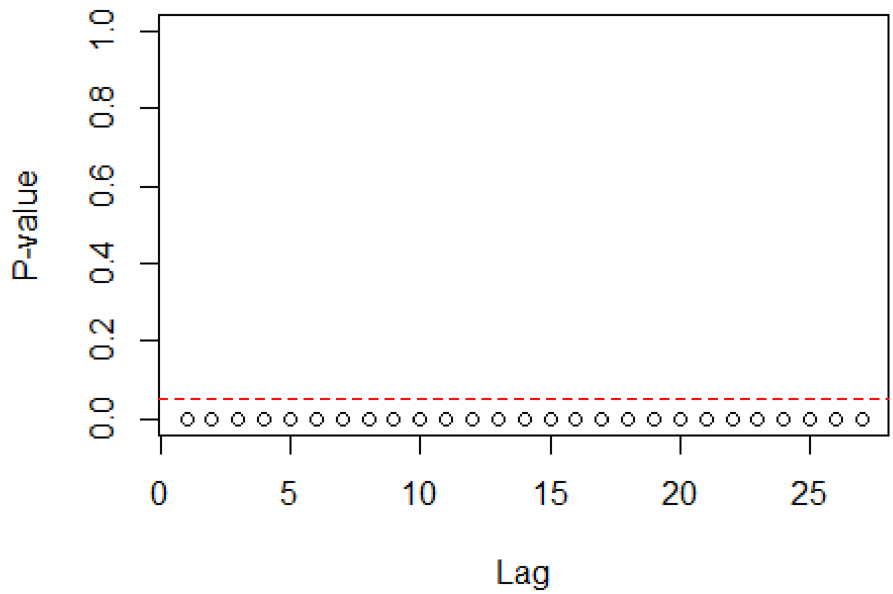
```
oil.predict <- forecast(oil.hw, h = 15)
plot(oil.predict, xlim =c(2008, 2010.4))
lines(y.prdct, col="firebrick")
#ts.plot(y, oil.predict, col = 1:2, xlab = "Time", ylab = "Observed/Predict")
legend("topleft", c("data", "prediction"), col = c(1,2), lty = 1, bty = "n")
```



The predicted sample falls within the foretasted range generated by the Holt Winters model. Although we can see that the forecast in showing an increase while the predicted is decreasing.

Part f

```
Box.test(auto$fitted^2, lag = log(545), type = "Ljung")

##
##  Box-Ljung test
##
## data:  auto$fitted^2
## X-squared = 2945.4, df = 6.3008, p-value < 2.2e-16

library(TSA)
McLeod.Li.test(auto, y.est)
```



By the box test and the McLeod test on the squared residuals, we find that there is conditional heteroscedasticity. In order to account for dependent volatility in the model, we need to run a Garch model.

Part g

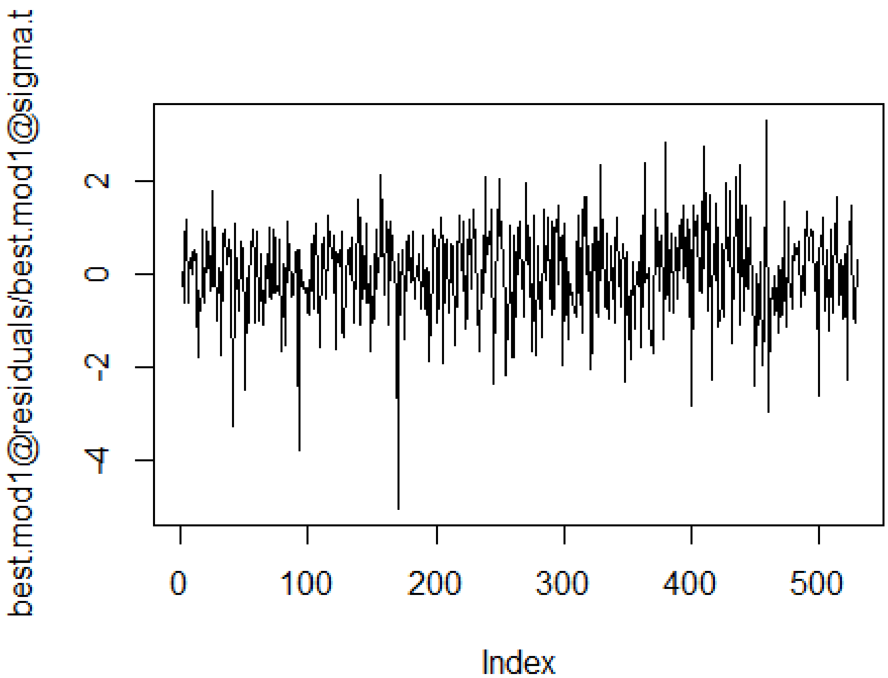
```
best.mod1 <- garchFit(~ arma(1,1) + garch(1,1), data = diff(y.est), cond.dist
= "norm", trace = FALSE)
best.mod2 <- garchFit(~ arma(1,1) + garch(1,2), data = diff(y.est), cond.dist
= "norm", trace = FALSE)

## Warning in sqrt(diag(fit$cvar)): NaNs produced

best.mod3 <- garchFit(~ arma(1,1) + garch(2,1), data = diff(y.est), cond.dist
= "norm", trace = FALSE)
summary(best.mod1)

##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~arma(1, 1) + garch(1, 1), data = diff(y.est),
##     cond.dist = "norm", trace = FALSE)
##
## Mean and Variance Equation:
##  data ~ arma(1, 1) + garch(1, 1)
## <environment: 0x0000000023f39160>
```

```
##  [data = diff(y.est)]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##        mu        ar1        ma1      omega     alpha1      beta1
##  0.091502  -0.443716   0.619195   0.056629   0.077663   0.912519
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu       0.09150     0.11900    0.769 0.441940
## ar1     -0.44372     0.13386   -3.315 0.000917 ***
## ma1      0.61919     0.11608    5.334 9.60e-08 ***
## omega    0.05663     0.02772    2.043 0.041097 *
## alpha1   0.07766     0.01728    4.495 6.96e-06 ***
## beta1    0.91252     0.01914   47.665  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  -1116.709    normalized:  -2.110982
##
## Description:
##  Sun Apr 08 17:27:24 2018 by user: Kristopher
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  79.65849  0
##  Shapiro-Wilk Test  R    W      0.9811542 2.402027e-06
##  Ljung-Box Test     R    Q(10)  10.54221  0.3942767
##  Ljung-Box Test     R    Q(15)  15.72279  0.4007107
##  Ljung-Box Test     R    Q(20)  21.76071  0.3536351
##  Ljung-Box Test     R^2  Q(10)  4.575141  0.9176956
##  Ljung-Box Test     R^2  Q(15)  5.857387  0.9820449
##  Ljung-Box Test     R^2  Q(20)  7.971572  0.9920542
##  LM Arch Test       R    TR^2   4.953987  0.9594982
##
## Information Criterion Statistics:
##      AIC      BIC      SIC     HQIC
## 4.244648 4.293090 4.244394 4.263610

#summary(best.mod2)
#summary(best.mod3)
```

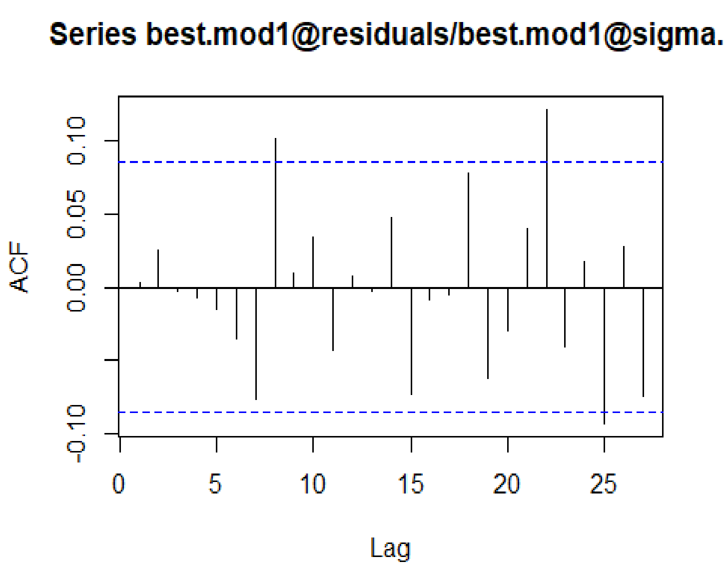
By the AIC and BIC values, the best model is an ARIMA-GARCH(1,1,1)(1,1,1).

Part h

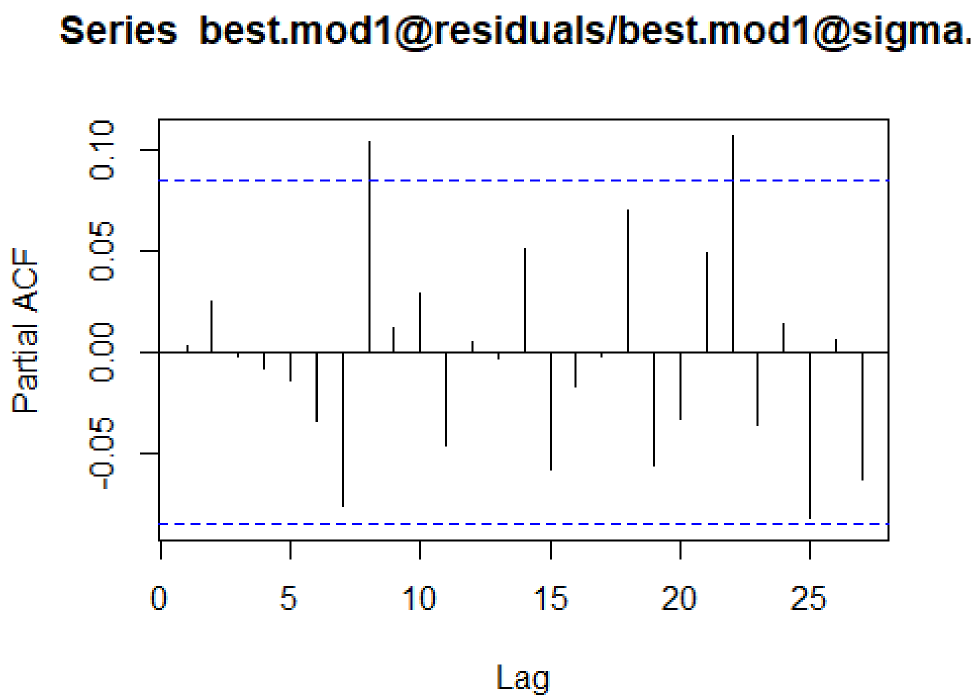
```
plot(best.mod1@residuals/best.mod1@sigma.t, type = "l")
```




```
acf(best.mod1@residuals/best.mod1@sigma.t)
```



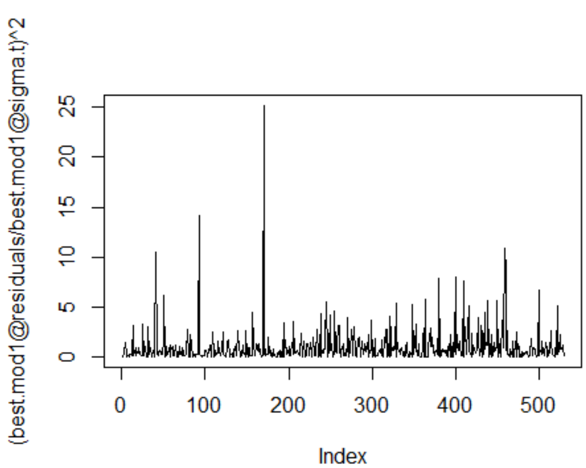
```r
pacf(best.mod1@residuals/best.mod1@sigma.t)
```



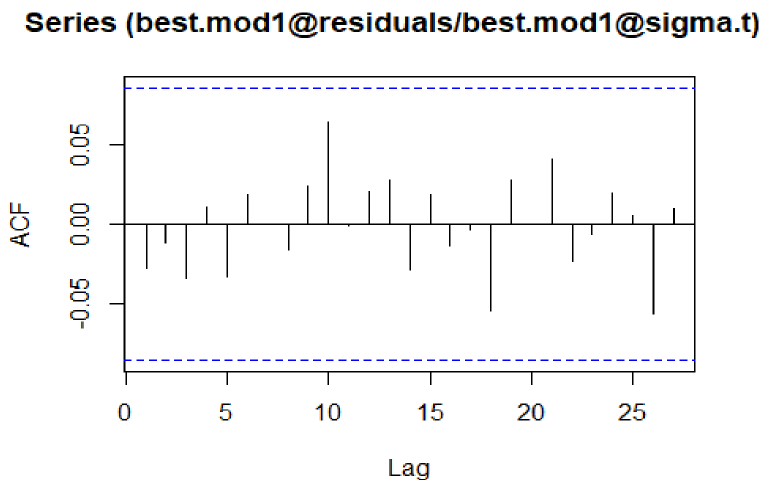

```r
Box.test(best.mod1@residuals/best.mod1@sigma.t, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  best.mod1@residuals/best.mod1@sigma.t
## X-squared = 0.0047513, df = 1, p-value = 0.945
```

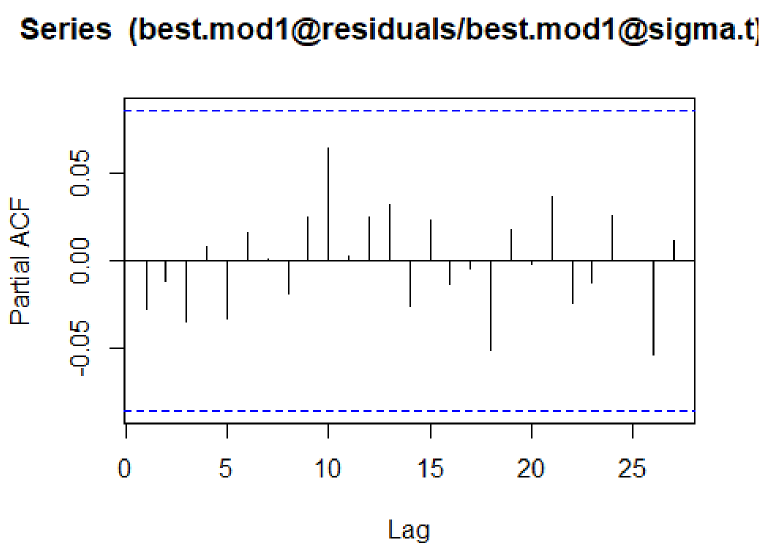
```r
plot((best.mod1@residuals/best.mod1@sigma.t)^2, type = "l")
```



```r
acf((best.mod1@residuals/best.mod1@sigma.t)^2)
```




```r
pacf((best.mod1@residuals/best.mod1@sigma.t)^2)
```

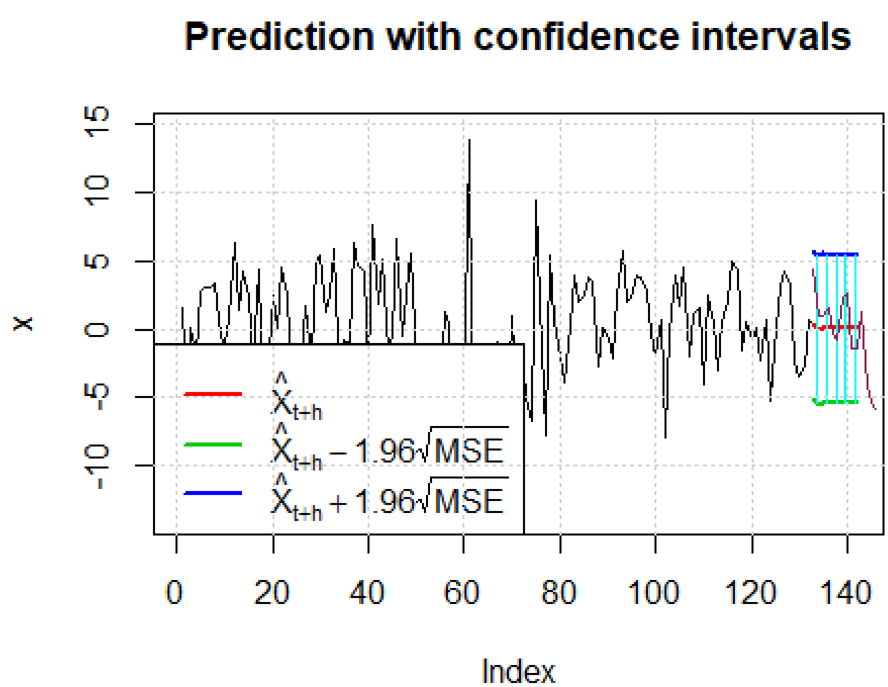


```r
Box.test((best.mod1@residuals/best.mod1@sigma.t)^2, type = "Ljung")

##
##  Box-Ljung test
##
## data:  (best.mod1@residuals/best.mod1@sigma.t)^2
## X-squared = 0.39986, df = 1, p-value = 0.5272
```

Looking at the ACF and PACF for the standardized residuals, we do see a few significant spikes but they don't fall that far outside the confidence interval. In examining the ACF and PACF for the squared standardized residuals, we see no significant spikes. Autocorrelation has been removed from σ_t^2 in the model and our box test confirms it.

Part i

```r
mod.forecast <- predict(best.mod1, h=15, plot=TRUE)
lines(ts(diff(y.prdct), start = 133), col = "violetred4")
```



The prediction sample does lie within the 95% confidence interval of the foretasted Garch model. So this is a decent fit.