Homework Two for Econometrics

Kristopher C. Toll

January 31, 2018

Theory

Problem 1

Part a

This is a Trend Stationary model

Part b

$$E(x_t) = E[\beta_1 + \beta_2 t^2 + w_t] = \beta_1 + \beta_2 E[t^2] = \beta_1 + \beta_2 t^2$$

This model is dependent upon t and thus it is non stationary.

Part c

$$y_t = x_t + x_{t-1} = \beta_1 + \beta_2 t^2 + w_t - [\beta_1 + \beta_2(t-1)^2 + w_{t-1}]$$

$$y_t = \beta_2(t^2 - (t-1)^2) + w_t - w_{t-1}$$

$$E(y_t) = E(\beta_2(t^2 - (t-1)^2) + w_t - w_{t-1}) = E(\beta_2(t^2 - (t-1)^2)) + E(w_t) - E(w_{t-1})$$

$$E(y_t) = \beta_2(t^2 - (t-1)^2) + 0 + 0$$

$$E(y_t) = \beta_2(t^2 - (t-1)^2) = \beta_2(2t-1)$$

The $E[y_t]$ is dependent upon t and thus it is not constant. Therefore, y_t is still non stationary.

part d

$$z_t = y_t - y_{t-1} = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = (x_t - 2(x_{t-1}) + x_{t-2})$$

$$z_t = \beta_1 + \beta_2 t^2 + w_t - 2\beta_1 - 2\beta_2(t-1)^2 - 2w_{t-1} + \beta_1 + \beta_2(t-2)^2 + w_{t-2}$$

$$z_t = \beta_2(t^2 - 2(t-1)^2 + (t-2)^2)) + w_t - 2w_{t-1} + w_{t-2} = 2\beta_2 + w_t - 2w_{t-1} + w_{t-2}$$

$$z_t = \beta_2(t^2 - 2t^2 - 4t - 2 + t^2 - 4t + 4) + w_t - 2w_{t-1} + w_{t-2}$$

$$z_t = 2\beta_2 + w_t - 2w_{t-1} + w_{t-2}$$

$$E(z_t) = E[2\beta_2 + w_t - 2w_{t-1} + w_{t-2}] = 2\beta_2 + 0 - 2(0) + 0$$

$$E(z_t) = 2\beta_2$$

$$var(z_t) = E[(z_t - E(z_t))^2]$$

$$var(z_t) = E[(2\beta_2 + w_t - 2w_{t-1} + w_{t-2} - 2\beta_2)^2] = E[(w_t - 2w_{t-1} + w_{t-2})^2]$$

$$var(z_t) = E[(w_t - 2w_{t-1} + w_{t-2})^2] = \sigma^2 + 4\sigma^2 + \sigma^2 = 6\sigma^2$$

$$cov(z_t, z_{t-h}) = E[(z_t - E(z_t))(z_{t-h} - E(z_{t-h}))]$$

$$cov(z_t, z_{t-h}) = E((w_t - 2w_{t-1} + w_{t-2})(w_t - 2w_{t-1-h} + w_{t-2-h}))$$

$$cov(z_t, z_{t-h} = 6\sigma^2)$$

This is now a stationary model because $E(z_t, var(z_t,$ and the $cov(z_t, z_{t-h})$ do not depend on t.

part e

$$v_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j}$$

$$E[v_t] = E[\frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j}] = \frac{1}{2q+1} \sum_{j=-q}^{q} E[x_{t-j}] = \frac{1}{2q+1} \sum_{j=-q}^{q} E[\beta_1 + \beta_2(t-j)^2 + w_{t-j}]$$

$$E[v_t] = \frac{\sum_{j=-q}^{q} \beta_1}{2q+1} + \frac{\sum_{j=-q}^{q} \beta_2(t^2 - 2tj + j^2)}{2q+1} + \frac{\sum_{j=-q}^{q} E[w_{t-j}]}{2q+1}$$

$$E(y_t) = \frac{(2q+1)\beta_1}{2q+1} + \frac{(2q+1)\beta_2 t^2}{2q+1} - \frac{2(2q+1)\beta_2 t \sum_{j=-q}^{q} j}{2q+1} + \frac{\beta_2}{2q+1} \sum_{j=-q}^{q} j^2 + \frac{0}{2q+1}$$

$$E(y_t) = \frac{(2q + 1)\beta_1}{2q + 1} + \frac{(2q + 1)\beta_2 t^2}{2q + 1} - \frac{2(2q + 1)\beta_2 t * (0)}{2q + 1} + \frac{\beta_2}{2q + 1} \sum_{j=-q}^{q} j^2 + \frac{0}{2q + 1}$$

$$E[v_t] = \beta_1 + \beta_2 t^2 + \frac{\beta_2}{2q + 1} \sum_{j=-q}^{q} j^2$$

Problem 2

$$x_t = 2w_{t-1} + w_t + 2w_{t+1}$$

$$E(x_t) = E[2w_{t-1} + w_t + 2w_{t+1}] = 2E[w_{t-1}] + E[w_t] + 2E[w_{t+1}] = 2(0) + (0) + 2(0) = 0$$

$$E(x_t) = \mu = 0$$

$$var(x_t) = var(2w_{t-1} + w_t + 2w_{t+1}) = 4var(w_{t-1}) + var(w_t) + 2var(w_{t+1})$$

$$var(x_t) = 4\sigma^2 + \sigma^2 + 4\sigma^2 = 9\sigma^2$$

$$var(x_t) = 9\sigma^2$$

When $t = t + h, E(x_t x_s) = \theta\sigma^2$

$$cov(x_t, x_t) = \gamma(0) = E[(x_t - E(x_t))(x_t - E(x_t))] = E(x_t^2) = \sigma^2$$

$$cov(x_t, x_t) = \gamma(0) = \sigma^2$$

$$cov(x_t, x_{t+1}) = \gamma(1) = E[(x_t - E(x_t))(x_{t-1} - E(x_{t-1}))] = E[x_t x_{t-1}]$$

$$cov(x_t, x_{t+1}) = \gamma(1) = E[(2w_{t-1} + w_t + 2w_{t+1})(2w_t + w_{t+1} + 2w_{t+2})]$$

$$cov(x_t, x_{t+1}) = \gamma(1) = 2\sigma^2 + 2\sigma^2 = 4\sigma^2$$

$$cov(x_t, x_{t+2}) = \gamma(2) = E[(2w_{t-1} + w_t + 2w_{t+1})(2w_{t+1} + w_{t+2} + 2w_{t+3})]$$

$$cov(x_t, x_{t+2}) = \gamma(2) = 4\sigma^2$$

$$cov(x_t, x_{t+3}) = \gamma(2) = E[(2w_{t-1} + w_t + 2w_{t+1})(2w_{t+2} + w_{t+3} + 2w_{t+4})]$$

$$cov(x_t, x_{t+3}) = \gamma(3) = 0$$

Autocorrelation calculations

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{4\sigma^2}{\sigma^2} = 4$$

$$\rho(2) = \frac{\gamma(2)}{\gamma(0)} = \frac{4\sigma^2}{\sigma^2} = 4$$

$$\rho(3) = \frac{\gamma(3)}{\gamma(0)} = \frac{0}{\sigma^2} = 0$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{0}{\sigma^2} = 0$$

Problem 3

part a

$$x_t = \delta + x_{t-1} + w_t$$

$$x_{t-1} = \delta + x_{t-2} + w_{t-1}$$

After some many lags we will reach.

$$x_{t-h} = \delta + x_0 + w_{t-h}$$

$$x_0 = 0$$

Eventually through substitution of lagged models we will obtain. I would show more steps that involved δ being sum t times but latex was having trouble.

$$x_t = t\delta + \sum_{k=1}^{t} w_k$$

part b

$$E(x_t) = E(\delta t + \sum_{k=1}^{t} w_k) = \delta t + \sum_{k=1}^{t} E(w_k) = \delta t + \sum_{k=1}^{t} (0)$$

$$E(x_t) = \delta t$$

$$var(x_t) = E((\delta t + \sum_{k=1}^{t} (w_k) - \delta t)^2) = \sum_{k=1}^{t} E(w_k)^2 = \sum_{k=1}^{t} \sigma^2$$

$$var(x_t) = t\sigma^2$$

$$cov(x_t, x_s) = E[(x_t - E(x_t))(x_s - E(x_s))] = E[(\delta t + \sum_{k=1}^{t} (w_k) - \delta t)(\delta s \sum_{j=1}^{s} (w_j) - \delta s)]$$

$$cov(x_t, x_s) = E[\sum_{k=1}^{t} (w_k) \sum_{j=1}^{s} (w_j)] = min(t,s)\sigma^2$$

part c

It was shown in part b that $E[x_t]$, $var(x_t)$ and the $cov(x_t, x_s)$ all depend on the time period and thus are not constant which implies nonstationarity.

part d

Let $t - 1 = s$

$$\rho(s,t) = \frac{\gamma_{x_s}\gamma_{x_t}}{\sqrt{\gamma_{x_t}\gamma_{x_s}}} = \frac{min(s,t)\sigma^2}{\sigma^2\sqrt{st}}$$

I had a little trouble showing all my steps in latex so some are missing

$$\rho(s,t) = \frac{(t-1)}{\sqrt{(t-1)t}} = \sqrt{1 - \frac{1}{t}}$$

$$lim_{t->\infty} \sqrt{1 - \frac{1}{t}} = \sqrt{1 - \frac{1}{\infty}} = \sqrt{1-0} = 1$$

This result implies that $t - 1$ can be used to perfectly predict t. They are both perfectly correlated

part e

$$\Delta x_t = x_t - x_{t-1} = \delta t + \sum_{k=1}^{t}(w_k) - \delta(t-1) - \sum_{k=1}^{t-1}(w_k)$$

$$\Delta x_t = \delta t + w_t + \sum_{k=1}^{t-1}(w_k) - \delta(t-1) - \sum_{k=1}^{t-1}(w_k)$$

$$\Delta x_t = \delta + w_t$$

$$E(\Delta x_t) = E(\delta + w_t) = E(\delta) + E(w_t) = \delta$$

$$var(\Delta x_t) = var(\delta + w_t) = var(\delta) + var(w_t) = \sigma^2$$

$$cov(\Delta x_t, \Delta x_{t-h}) = E[(\Delta x_t - E(\Delta x_t)(\Delta x_{t=h} - E(\Delta x_{t-h}))] = E[(\delta + w_t - \delta)(\delta + w_{t-h} - \delta)]$$

$$cov(\Delta x_t, \Delta x_{t-h}) = E(w_t w_{t-h}) = \sigma^2$$

As was shown; the $E[x_t]$, $var(x_t)$ and the $cov(x_t, x_s)$ all do not depend on the time period t and thus they are constant which implies stationarity.

Problem 4

part a

This is an ARMA(2,2) data generating process

part b

$$E(x_t) = E(\alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}) = \alpha + \phi_1 E(x_{t-1}) + \phi_2 E(x_{t-2})$$

$$E(x_t) - \phi_1 E(x_{t-1}) - \phi_2 E(x_{t-2}) = E(x_t)(1 - \phi_1 - \phi_2) = \alpha$$

$$E(x_t) = \frac{\alpha}{(1 - \phi_1 - \phi_2)}$$

Problem 5

part a

This is an ARMA(1,1)

part b

$$E(y_t) = E(\phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}) = \phi E(y_{t-1})$$

$$E(y_t) - \phi E(y_{t-1}) = E(y_t)(1 - \phi) = 0$$

$$E(y_t) = \frac{0}{1 - \phi} = 0$$

$$var(y_t) = E((y_t - E(y_t))^2) = E[(\phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1})^2]$$

$$var(y_t) = E(y_t)^2$$

$$= \phi^2 E(y_{t-1})^2 + E(\varepsilon_t)^2 + \theta^2 E(\varepsilon_{t-1})^2 + 2\phi E(y_t)E(\varepsilon_t) + 2E(\varepsilon_t)\theta E(\varepsilon_{t-1}) + 2\theta \varepsilon_{t-1}\phi y_{t-1}$$

$$var(y_t) = E(y_t)^2 = \phi^2 E(y_{t-1})^2 + \sigma^2 + \theta^2 \sigma^2$$

$$var(y_t) = E(y_t)^2 - \phi^2 E(y_{t-1})^2 = E(y_t)^2(1 - \phi^2) = \sigma^2 + \theta^2 \sigma^2$$

$$E(y_t)^2 = \frac{\sigma^2 + \theta^2 \sigma^2}{1 - \phi^2}$$

Application

Problem 1

part a

```r
# Create the time trend

library(astsa)

# Creating the data set
trend <- time(jj)
q <- factor(rep(1:4, 21))
ln.jj <- log(jj)

# Without the intercept
reg.jj <- lm(ln.jj ~ -1 + trend + q, na.action = NULL)
summary(reg.jj)

##
## Call:
## lm(formula = ln.jj ~ -1 + trend + q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##          Estimate  Std. Error t value          Pr(>|t|)
## trend    0.167172    0.002259   74.00 <0.0000000000000002 ***
## q1    -328.276371    4.450537  -73.76 <0.0000000000000002 ***
## q2    -328.248248    4.451102  -73.75 <0.0000000000000002 ***
## q3    -328.178140    4.451666  -73.72 <0.0000000000000002 ***
## q4    -328.446897    4.452231  -73.77 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic:  2407 on 5 and 79 DF,  p-value: < 0.00000000000000022
```

Looking at the results of the output we can see that without an intercept term we are able to determine what the effects of each quarter are. It seems that in each quarter, there will be a price decrease of 32,800 percent in each quarter. The trend variable means that as each time period moves up by one, we have an average increase of 16% in price. However, by removing the intercept term and including each quarter in the model, we have colinear problem and thus the estimates cannot be taken seriously.

Part b

```
# With the intercept
reg.jj2 <- lm(ln.jj ~ trend + q, na.action = NULL)
summary(reg.jj2)

##
## Call:
## lm(formula = ln.jj ~ trend + q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##               Estimate  Std. Error t value             Pr(>|t|)
## (Intercept) -328.276371   4.450537 -73.761 < 0.0000000000000002 ***
## trend          0.167172   0.002259  73.999 < 0.0000000000000002 ***
## q2             0.028123   0.038696   0.727               0.4695
## q3             0.098231   0.038708   2.538               0.0131 *
## q4            -0.170527   0.038729  -4.403            0.0000331 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic:  1379 on 4 and 79 DF,  p-value: < 0.00000000000000022
```

By including the intercept parameter, we can now get a more precise estimate for each quarter. The first quarter is part of the intercept term and now the interpretations of each coefficient estimate are accurate and make more sense.

c

```
fitted.ts <- as.ts(reg.jj2)
plot(ln.jj, col="steelblue")
lines(fitted(reg.jj2), col="firebrick")
```
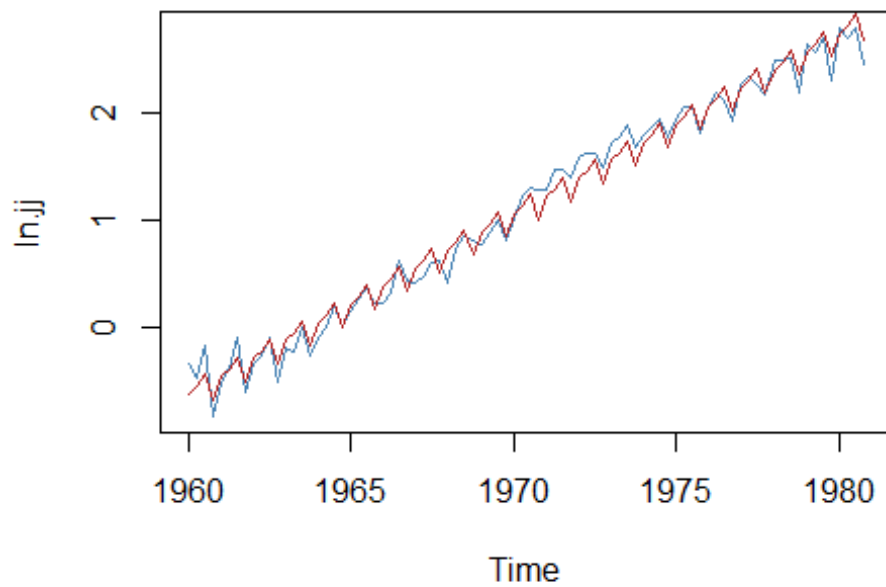
```
plot(reg.jj2$residuals, col="snow4") # I picked snow4 because it is white
just like white noise
```
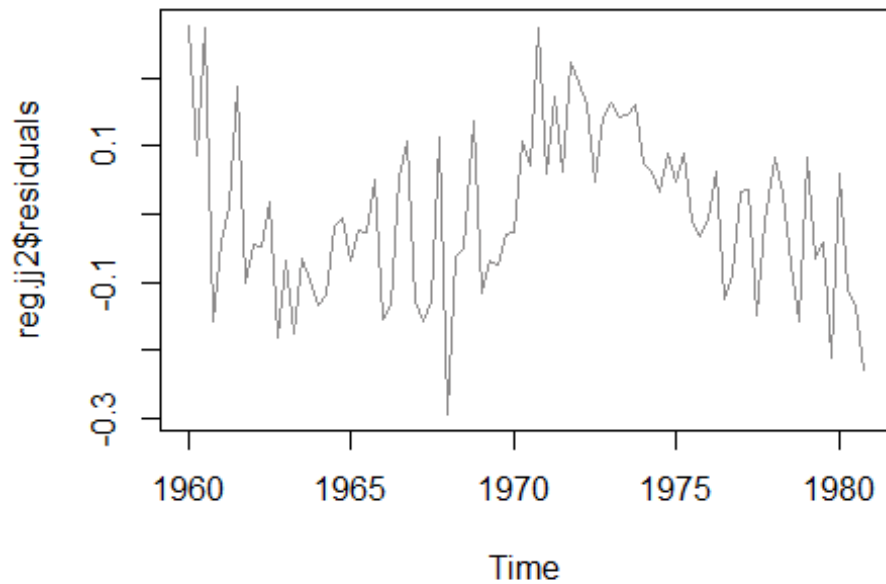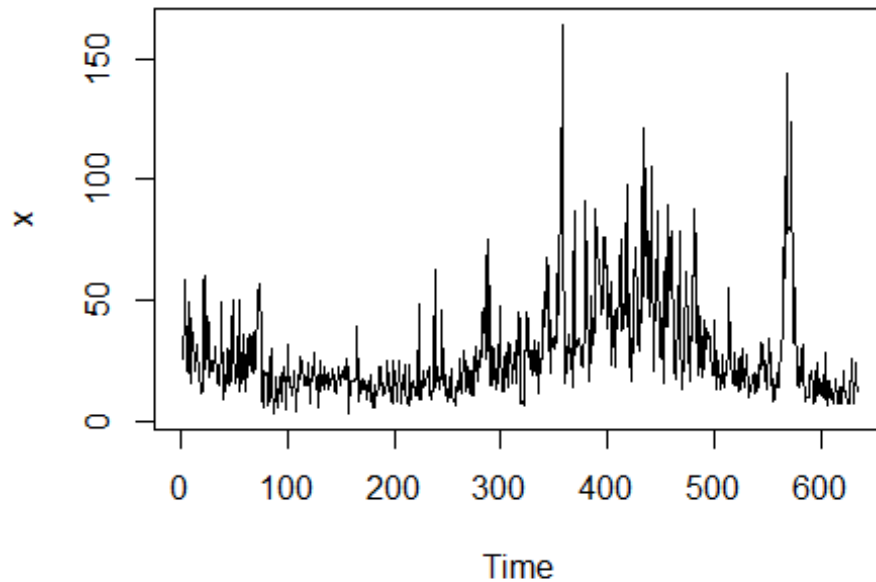

The model with the intercept fits the data very and well. The adjusted r squared value is .9852 and the plotted residuals do appear to be random which suggests white noise.

Problem 2

Part a

```
x <- varve
plot(x)
```


There do seem to be some trends from year to year. Which suggests nonstationarity but we can't verify from this plot alone.

Part b

```
half <- .5*length(x)
x.1 <- x[1:half]
varx.1 <- var(x.1)

x.2 <- x[(half+1):length(x)]
varx.2 <- var(x.2)

print(c(varx.1, "variance of first half"))

## [1] "133.457415667053"      "variance of first half"

print(c(varx.2, "variance of second half"))

## [1] "594.490438823224"       "variance of second half"
```

We can see that there is a difference between the variance of the first half of the series and the second half.

Part c

```r
y <- log(x)
y.1 <- y[1:half]
y.2 <- y[(half+1):length(y)]

print(c(var(y.1), "variance of first half"))

## [1] "0.270721652653357"        "variance of first half"

print(c(var(y.2), "variance of second half"))

## [1] "0.451371011716303"        "variance of second half"
```
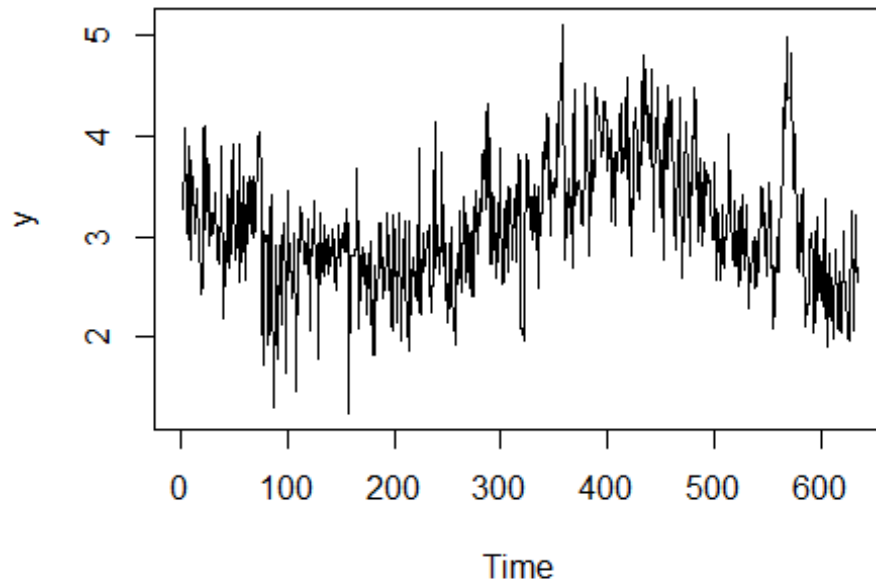
Here we can see that the difference in variance between the first half and second half are not as exaggerated and this series is more stable. Heterogeneity has be removed.

Part d

```r
hist(x, probability = TRUE, breaks = 20, ylim = c(0, 0.035))
lines(density(x), col = "firebrick") # I overlaid the kernel on the histogram
```
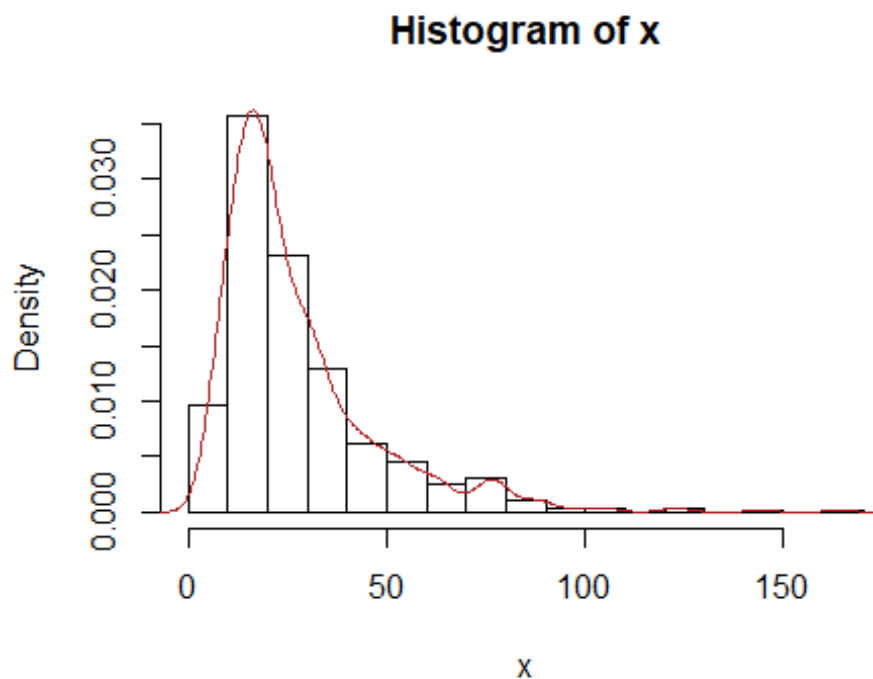
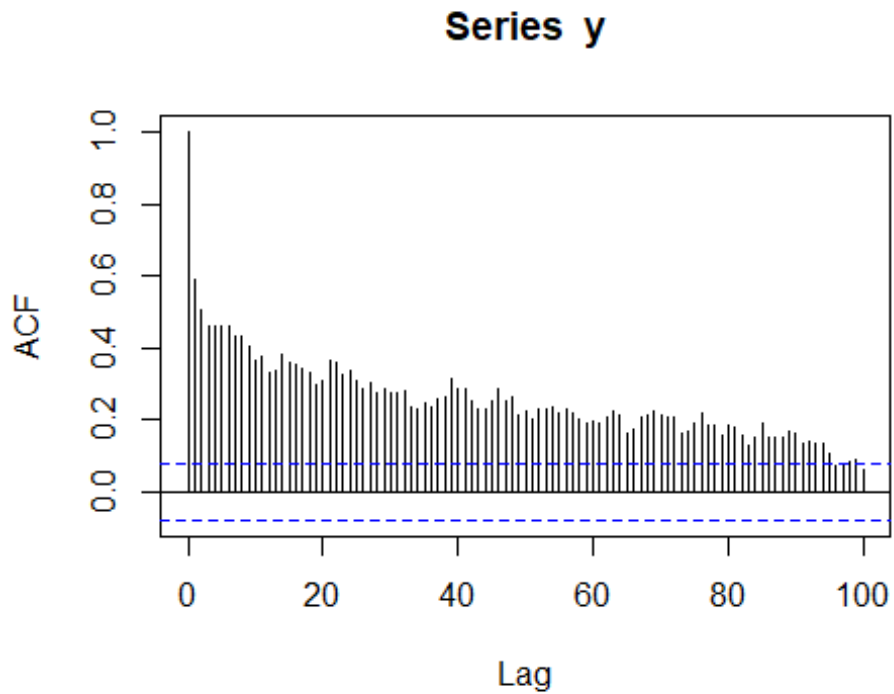**Histogram of x**

```
plot(y)
```



```
hist(y, probability = TRUE, breaks = 20, ylim = c(0, 0.65))
lines(density(y), col = "firebrick") # I overlaid the kernel on the histogram
```
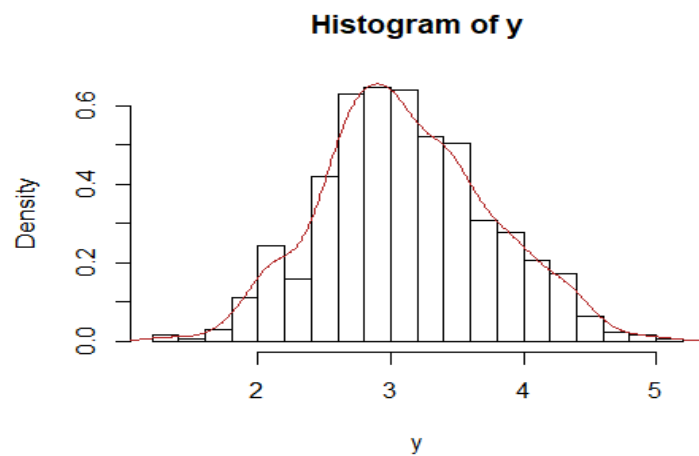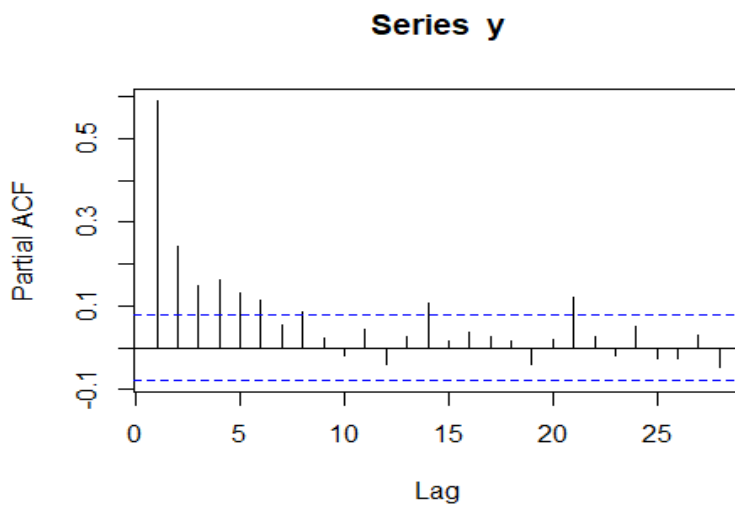

The log transformation does a lot of improve normality. It isn't perfect but it is much more usable.

```
acf(y, lag.max = 100)
```
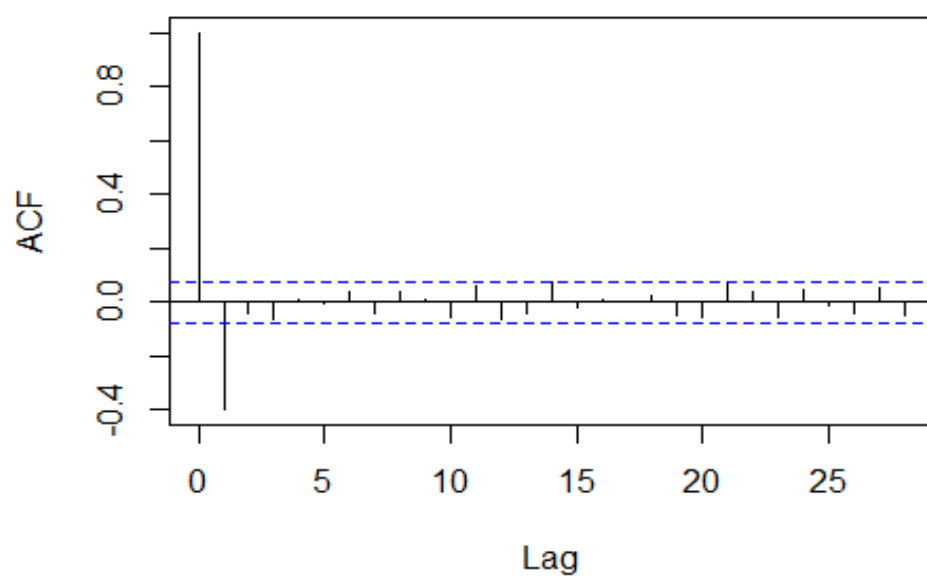
Series y


```
pacf(y)
```

Series y

By examining the acf we notice that there is a decrease in autocovariance. However, some of the future lags are higher than past lags and it doesn't decrease to zero till the 96 lag. It is difficult to determine what kind of process this is. Pacf doesn't help us either.

```
t <- 2:length(y)
u <- (y[t] - y[t-1])
plot(u, type = "l", col = "royalblue3", main = "Plot of One Period Lagged Log
Values")
```
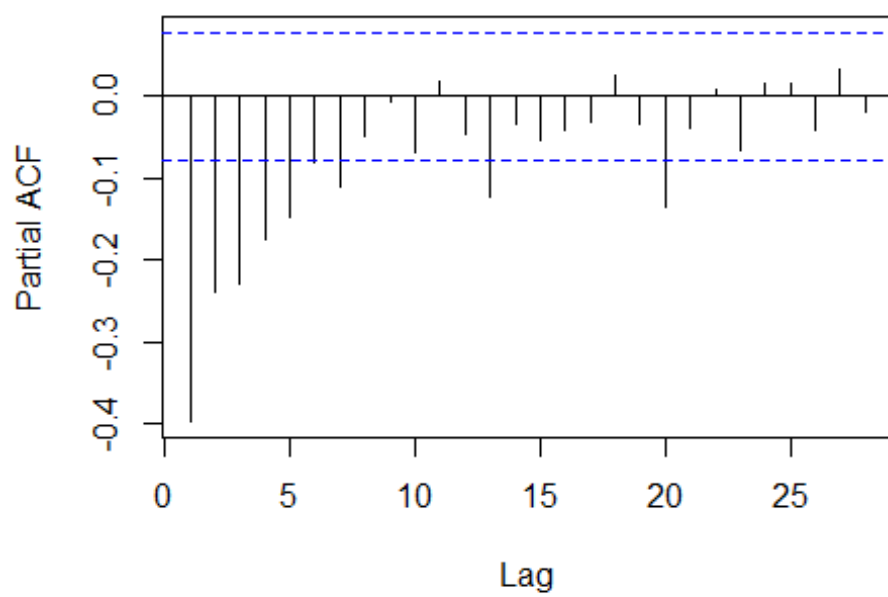

Plot of One Period Lagged Log Values
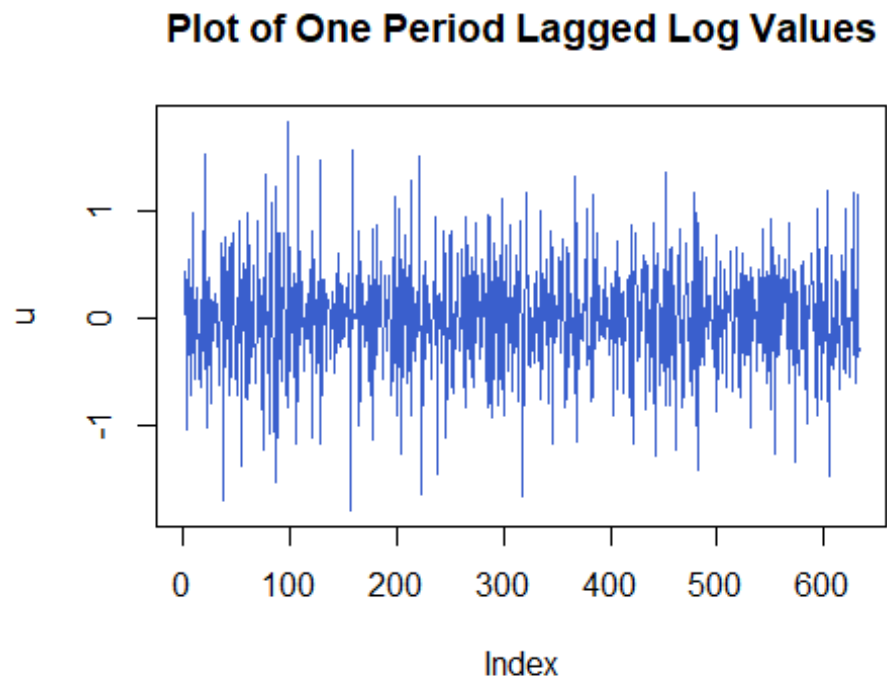
```
acf(u)
```

Series u


```
pacf(u)
```

Series u

Part g

In examining the plots we see that acf cuts off after one lag and partial acf seems to decay exponentially. This suggests an MA(1) process.