STAT 626
Homework 4

Name-Mayank Jaggi

UIN-526005299

Problem 2.8

ii]
$$x_t = w_t$$
$$y_t = w_t - \theta w_{t-1} + u_t$$

$$\gamma(h) = \text{cov}(y_t, y_{t+h})$$

$$= \text{cov}(w_t - \theta w_{t-1} + u_t, w_{t+h} - \theta w_{t+h-1} + v_{t+h})$$

At $h = 0$

$$\gamma(h) = \gamma(0) = \text{cov}(w_t - \theta w_{t-1} + u_t, w_t - \theta w_{t-1} + u_t)$$

$$= \text{var}(w_t - \theta w_{t-1} + u_t)$$

$$= (1+\theta^2)\sigma^2_w + \sigma^2_v$$

At $|h| = 1$

$$\gamma(h) = \text{cov}(w_t - \theta w_{t-1} + v_t, w_{t+1} - \theta w_t + v_{t+1})$$

$$= -\theta^2 \sigma^2_w$$ ($\because$ only 1 overlapping term)

For $|h| > 1$

$\gamma(h) = 0$ $\because$ No overlapping terms

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

$$\therefore \rho(h) = \begin{cases} 1, & h = 0 \\ \dfrac{-\theta\sigma^2_w}{(1+\theta^2)\sigma_w^2 + \sigma_u^2}, & |h| = 1 \\ 0, & |h| > 1 \end{cases}$$

b.]

$$\gamma_{xy}(h) = \text{cov}(x_{t+h}, y_t)$$
$$= \text{cov}(w_{t+h}, w_t - \theta w_{t-1} + u_t)$$

At $h = 0$

$$\gamma_{xy}(0) = \text{cov}(w_t, w_t - \theta w_{t-1} + u_t)$$
$$= \sigma^2_w$$

At $h = 1$

$$\gamma_{xy}(1) = \text{cov}(w_{t+1}, w_t - \theta w_{t-1} + u_t)$$

$$= 0 \quad \because \text{No overlap.}$$

At $h = -1$ ($\because$ CCF is asymmetric)

$$\gamma_{xy}(-1) = \mathrm{cov}(w_{t-1}, w_t - \theta w_{t-1} + u_t)$$
$$= -\theta \sigma_w^2$$

At $h = \pm 2$

$$\gamma_{xy}(h) = 0 \quad (\because \text{no overlap})$$

$$\therefore \gamma_{xy}(h) = \begin{cases} \sigma_w^2 & , h = 0 \\ 0 & , h = +1 \\ -\theta \sigma_w^2 & , h = -1 \\ 0 & , |h| > 1 \end{cases}$$

$$\rho_{xy}(0) = \frac{\sigma_w^2}{\sqrt{\sigma_w^2 \cdot \left[(1+\theta^2)\sigma_w^2 + \sigma_v^2\right]}}$$

$$\rho_{xy}(-1) = \frac{-\theta \sigma_w^2}{\sigma_w^2 \sqrt{(1+\theta^2) + \left(\frac{\sigma_v}{\sigma_w}\right)^2}}$$

$$= \frac{-\theta}{\sqrt{(1+\theta^2) + \left(\frac{\sigma_v}{\sigma_w}\right)^2}}$$

$\rho_{xy}(1) = 0$

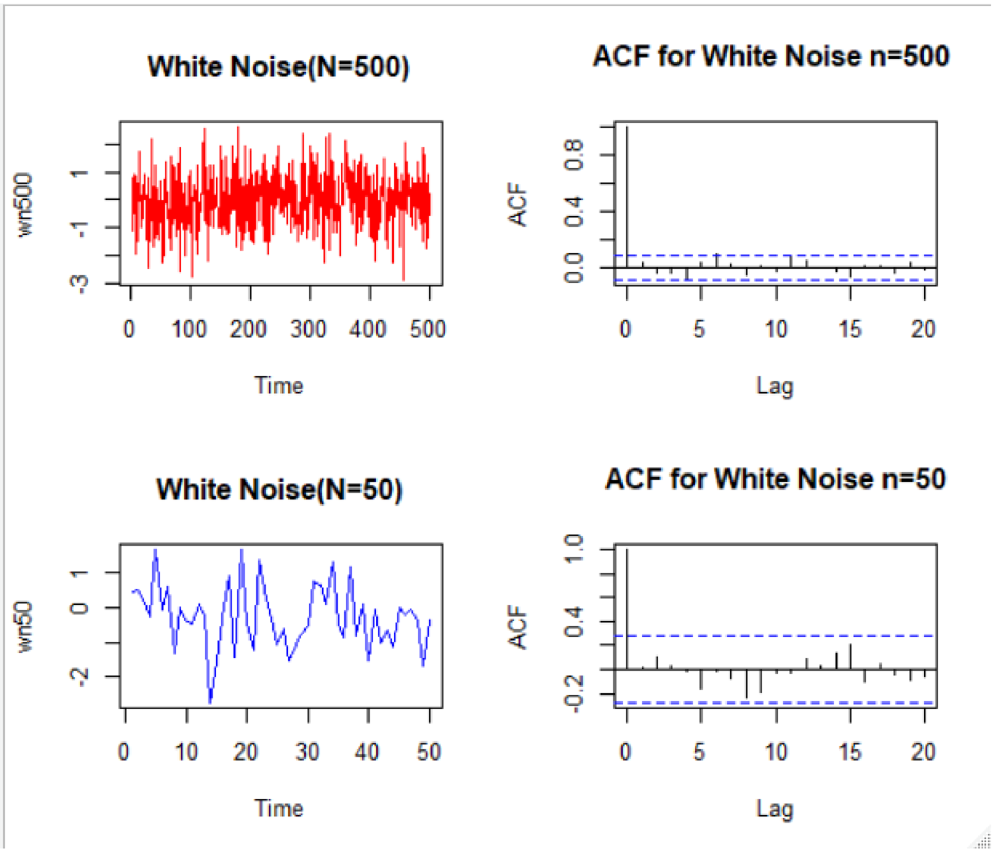
Summarizing,

$$\rho_{xy} = \begin{cases} \dfrac{1}{\sqrt{1+\theta^2+\left(\frac{\sigma_v}{\sigma_w}\right)^2}} & , h = 0 \\[2ex] \dfrac{-\theta}{\sqrt{1+\theta^2+\left(\frac{\sigma_v}{\sigma_w}\right)^2}} & , h = -1 \\[2ex] 0 & , h = 1 \\ 0 & , \text{otherwise} \end{cases}$$

c.] From the previous results mean is constant and cross variance is independent of time. Therefore, the series are jointly stationary.

Problem 2.11

a)



b)

```
Autocorrelations of series 'wn500', by lag

     0      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15     16     17     18
 1.000 -0.023 -0.058  0.002  0.029  0.008  0.035  0.044 -0.016 -0.088  0.029 -0.047  0.016 -0.053 -0.046 -0.004 -0.027 -0.018  0.065
    19     20
-0.046 -0.070
```

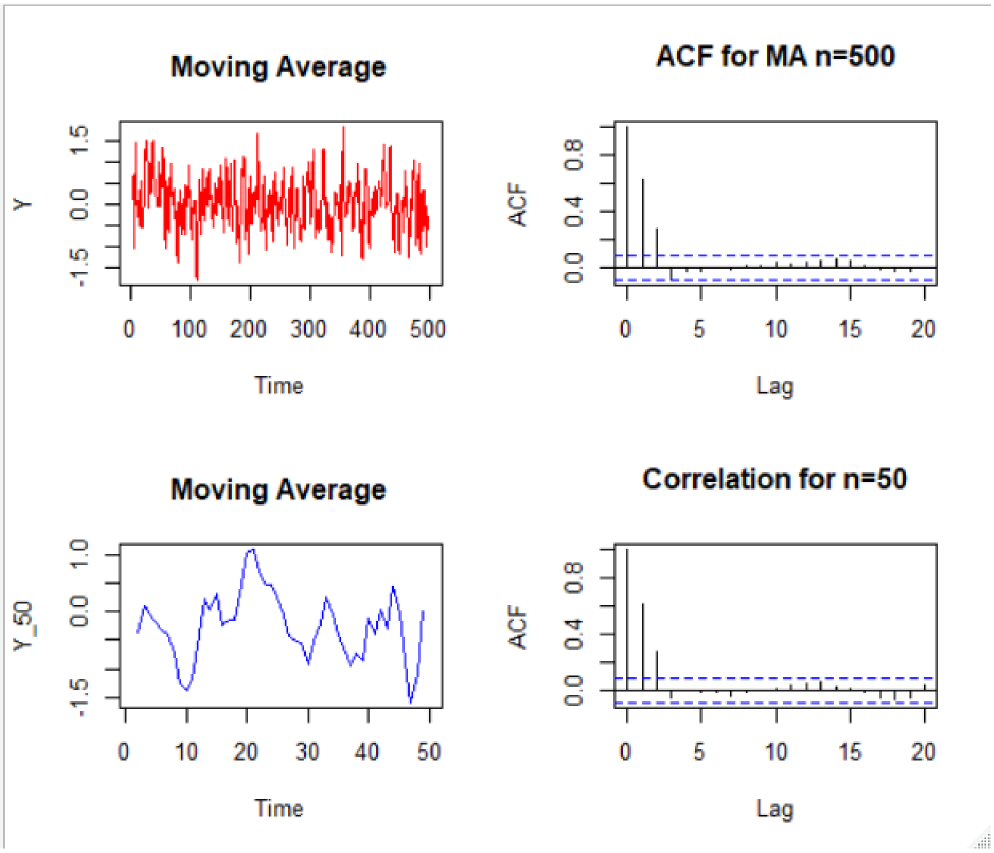
```
Autocorrelations of series 'wn50', by lag

     0      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15     16     17     18
 1.000 -0.058  0.044  0.103 -0.127  0.083 -0.090  0.057 -0.192 -0.087 -0.021 -0.098  0.146 -0.085  0.110  0.043 -0.069  0.055 -0.092
    19     20
 0.083 -0.068
```

By changing n from 500 to 50, we see in the ACF plot, the variation in ACF increases. Also, decreasing the sample size increases the width of the confidence interval.

Problem 2.12

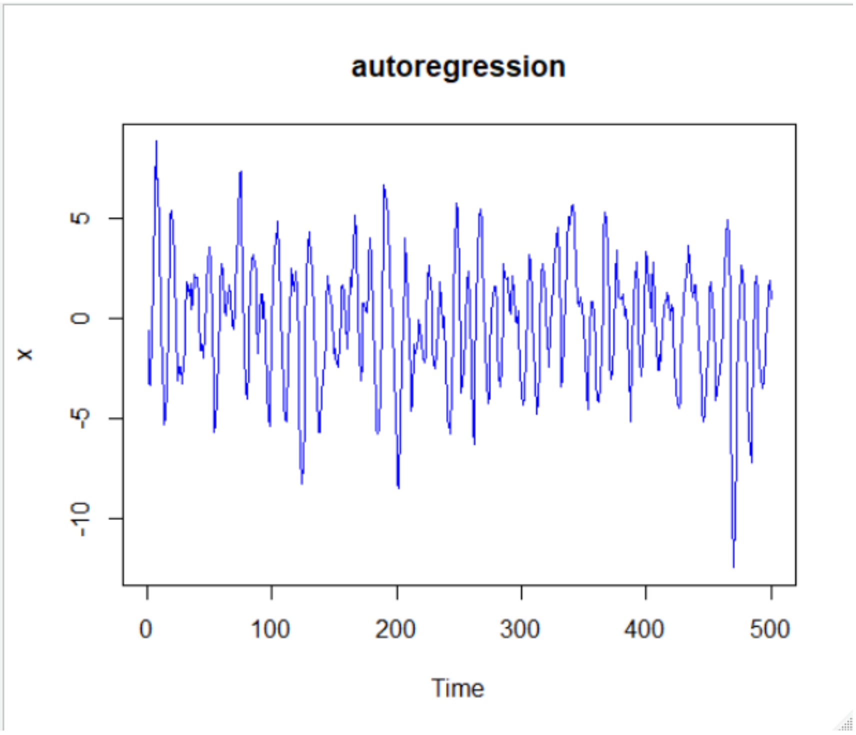
a)

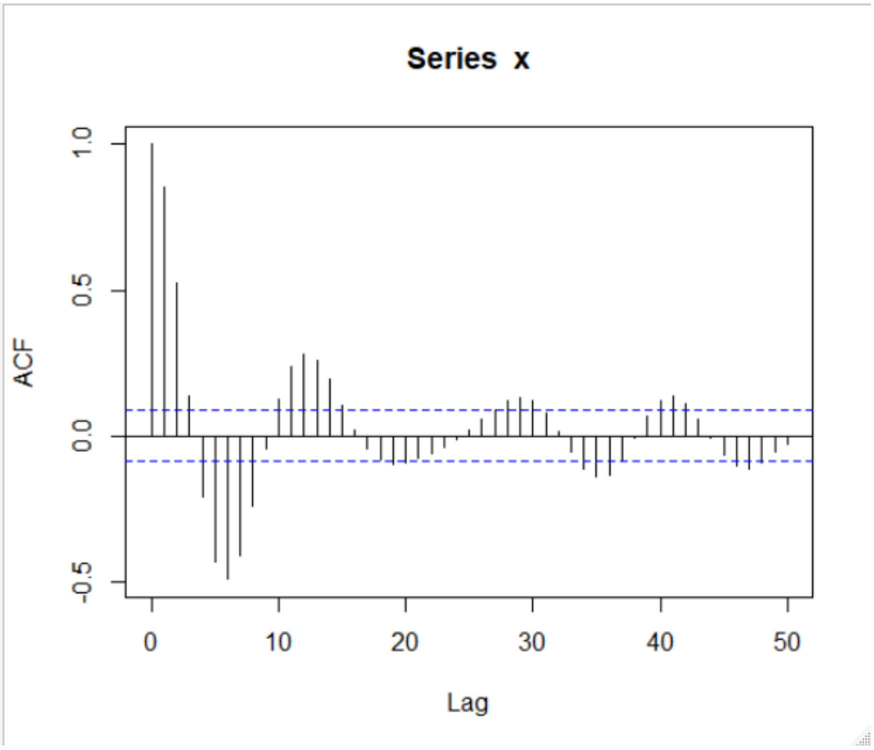


b)

By changing n from 500 to 50, we see in the ACF plot, the variation in ACF doesn't change. Also, decreasing the sample size doesn't change the width of the confidence interval.

Problem 2.13

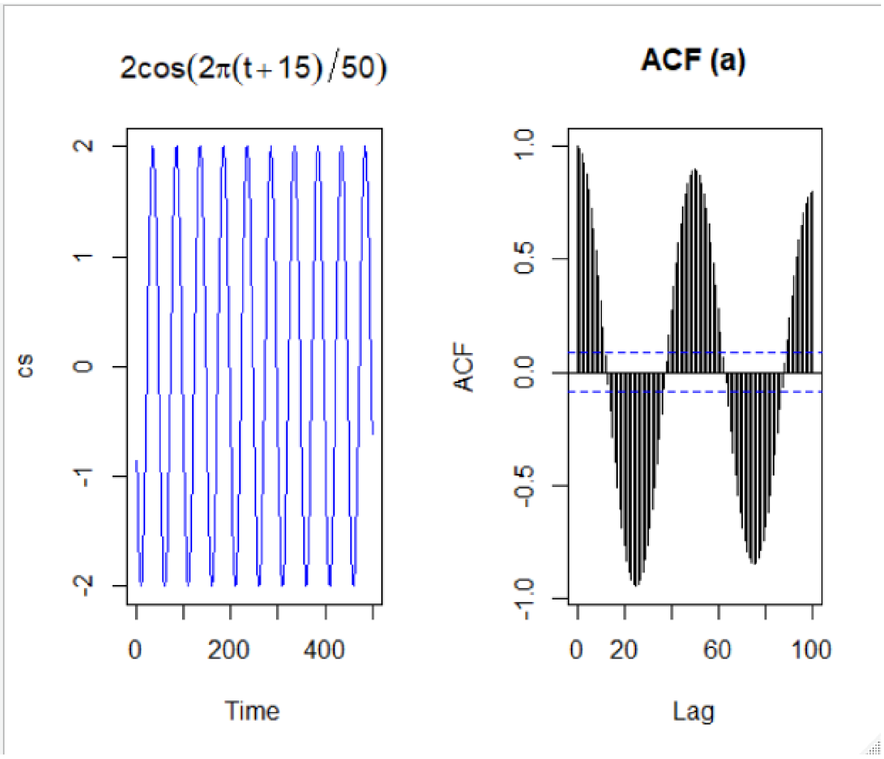




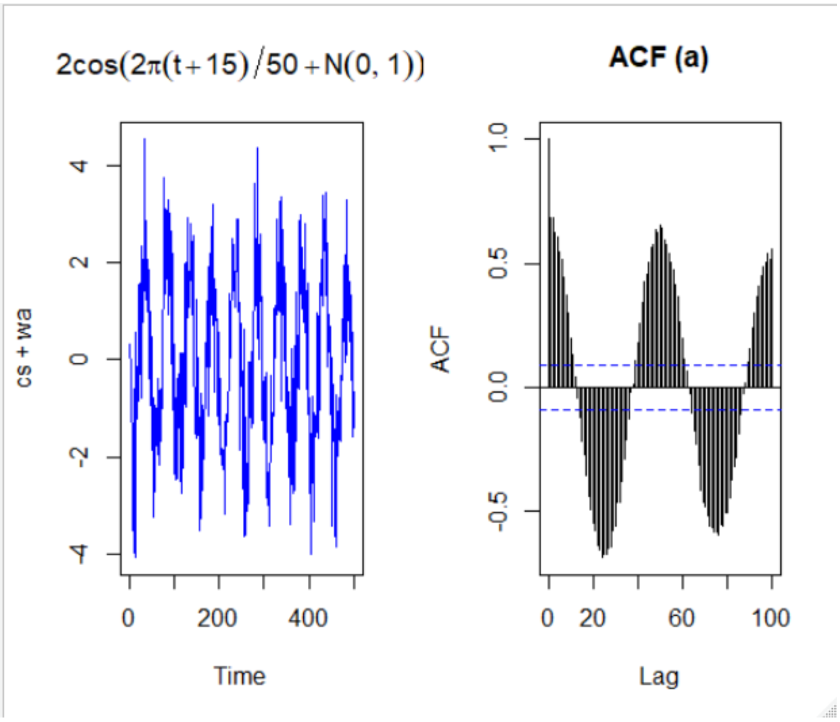
The ACF exhibits periodicity corresponding to the correlation between values separated by approx. 14 units. Observations 14 units apart are strongly positively and negatively correlated.

Problem 2.14

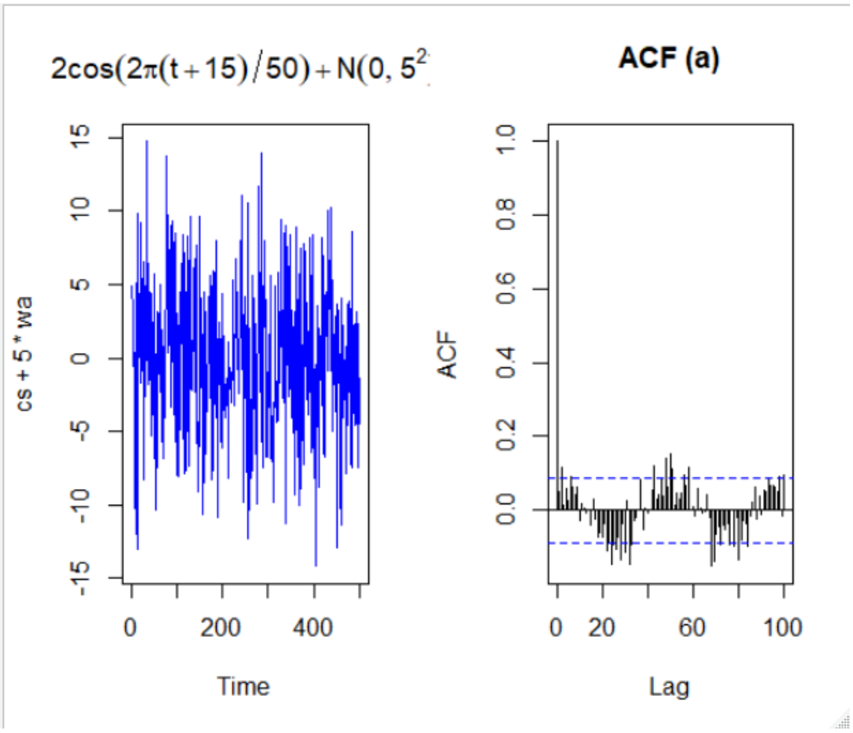
a)



b)



c)



The increase in variance of noise, degrades the signal and thus leading to decrease in ACF of the signal. As shown in the above graphs, the magnitude of the ACF is decreasing with the increase in variance of signal**.**

Problem 2.15

$$y_t = 5 + n_t - 0.5\, n_{t-1}$$

$$\gamma(0) = \operatorname{cov}(y_t, y_t)$$

$$= E\left[(5 + n_t - 0.5\, n_{t-1} - 5)^2\right]$$

$$= E(n_t - 0.5\, n_{t-1})^2$$

$$= E\left[n_t^2 + 0.25\, n_{t-1}^2 - n_t n_{t-1}\right]$$

$$= 1 + 0.25 - 0$$

$$= 1.25$$

$$\left[E(n_t^2) = \operatorname{Var}(n_t) + (E(n_t))^2\right]$$

$$\gamma(1) = \operatorname{cov}(y_{t+1}, y_t)$$

$$= E\left[(5 + n_{t+1} - 0.5\, n_t - 5)(5 + n_t - 0.5\, n_{t-1} - 5)\right]$$

$$= E\left[(n_{t+1} - 0.5\, n_t)(n_t - 0.5\, n_{t-1})\right]$$

$= -0.5$ (Only 1 Overlapping Term)

$\rho(1) = \frac{-0.5}{1.25} = -0.4$

For $h > 1$ there are no Overlapping Terms.

$\therefore \rho(h) = 0$ for $h > 1$

Problem 3.2

a)

Summary of Model 3.17 with AIC and BIC values

```
Call:
lm(formula = cmort ~ trend + temp + temp2 + part, na.action = NULL)

Residuals:
     Min       1Q   Median       3Q      Max 
-19.0760  -4.2153  -0.4878   3.7435  29.2448 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.831e+03  1.996e+02   14.19  < 2e-16 ***
trend       -1.396e+00  1.010e-01  -13.82  < 2e-16 ***
temp        -4.725e-01  3.162e-02  -14.94  < 2e-16 ***
temp2        2.259e-02  2.827e-03    7.99 9.26e-15 ***
part         2.554e-01  1.886e-02   13.54  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.385 on 503 degrees of freedom
Multiple R-squared:  0.5954,	Adjusted R-squared:  0.5922 
F-statistic:   185 on 4 and 503 DF,  p-value: < 2.2e-16

> num=length(cmort)
> AIC(fit1)/num-log(2*pi)   #AIC
[1] 4.721732
> BIC(fit1)/num-log(2*pi)   #BIC
[1] 4.771699
```

Summary of Model with P_{t-4} with AIC and BIC values

```
Call:
lm(formula = cmort ~ trend + temp + temp2 + part + part4, data = mort, 
    na.action = NULL)

Residuals:
    Min      1Q  Median      3Q     Max 
-18.228  -4.314  -0.614   3.713  27.800 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.808e+03  1.989e+02  14.123  < 2e-16 ***
trend       -1.385e+00  1.006e-01 -13.765  < 2e-16 ***
temp        -4.058e-01  3.528e-02 -11.503  < 2e-16 ***
temp2        2.155e-02  2.803e-03   7.688 8.02e-14 ***
part         2.029e-01  2.266e-02   8.954  < 2e-16 ***
part4        1.030e-01  2.485e-02   4.147 3.96e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.287 on 498 degrees of freedom
Multiple R-squared:  0.608,	Adjusted R-squared:  0.6041 
F-statistic: 154.5 on 5 and 498 DF,  p-value: < 2.2e-16

> 
> 
> num=length(cmort)
> AIC(fit)/num-log(2*pi)   #AIC
[1] 4.641492
> BIC(fit)/num-log(2*pi)   #BIC
[1] 4.699677
```

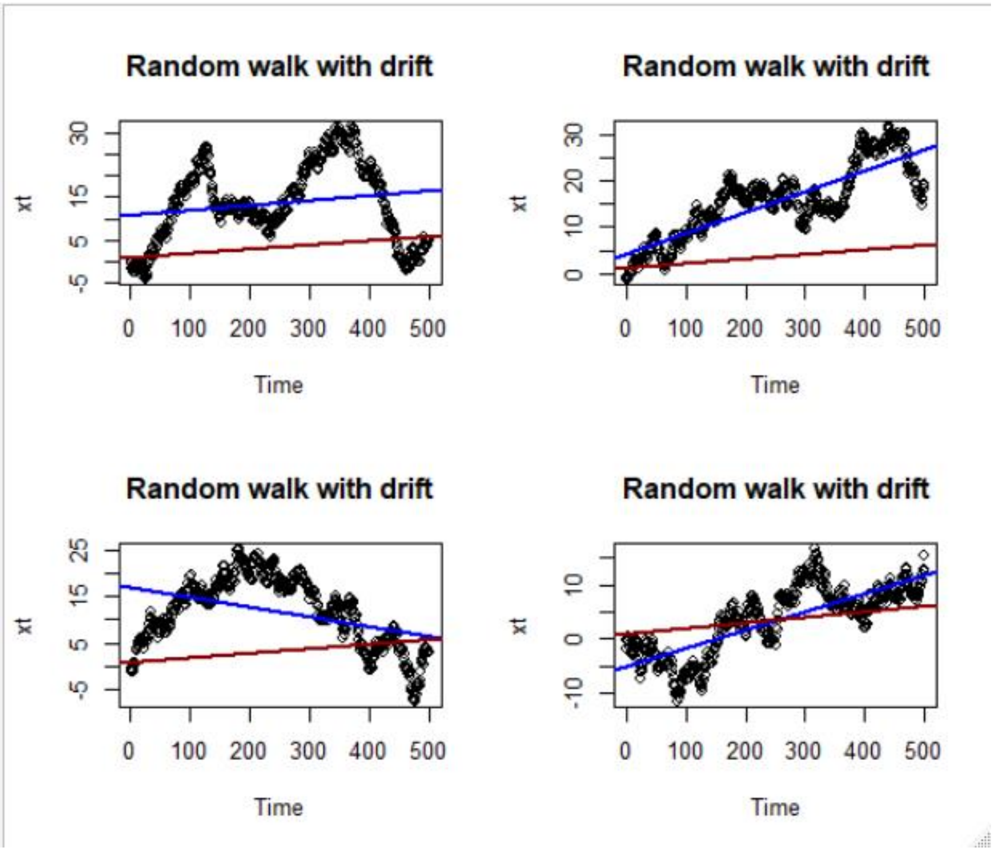
The new model has an Adjusted R-squared value larger than model 3.17, hence improving the fit.

b)

From the AIC and BIC values, the values are smaller in the new model. Therefore, there is an improvement over the model 3.17.

Problem 3.3

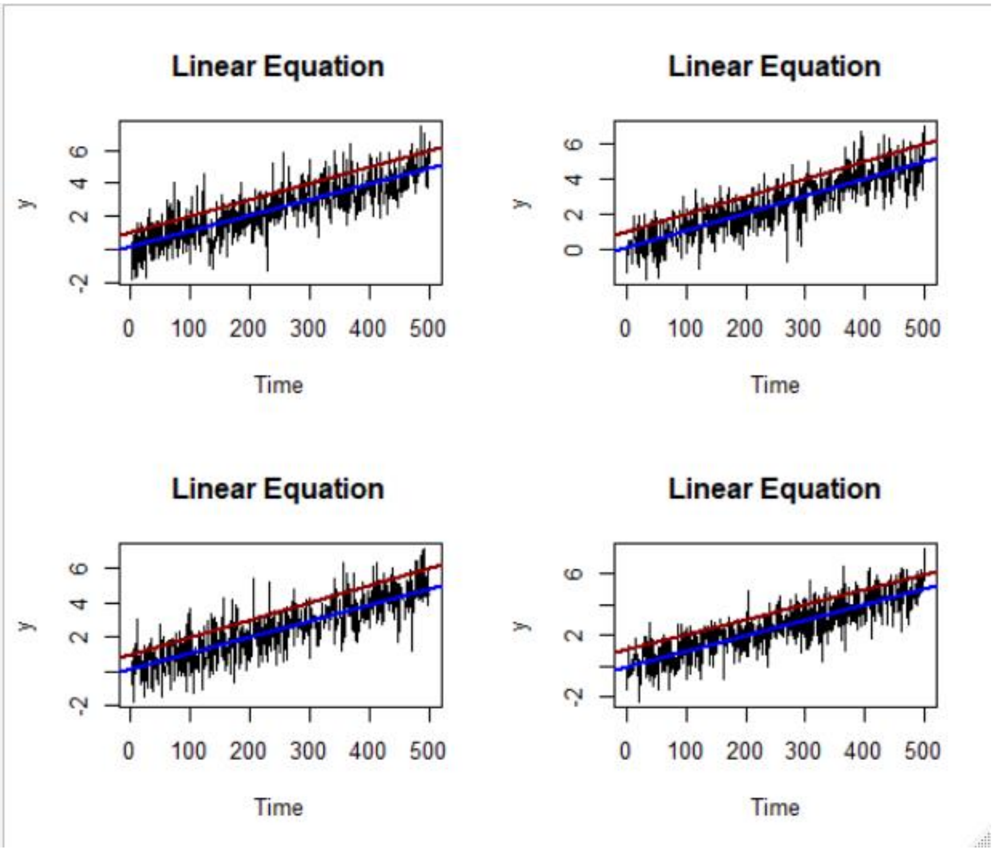
a)



Note:

- Blue line is fitted line
- True mean function is the red line

b)



Note:

- Blue line is fitted line
- True mean function is the red line

c)

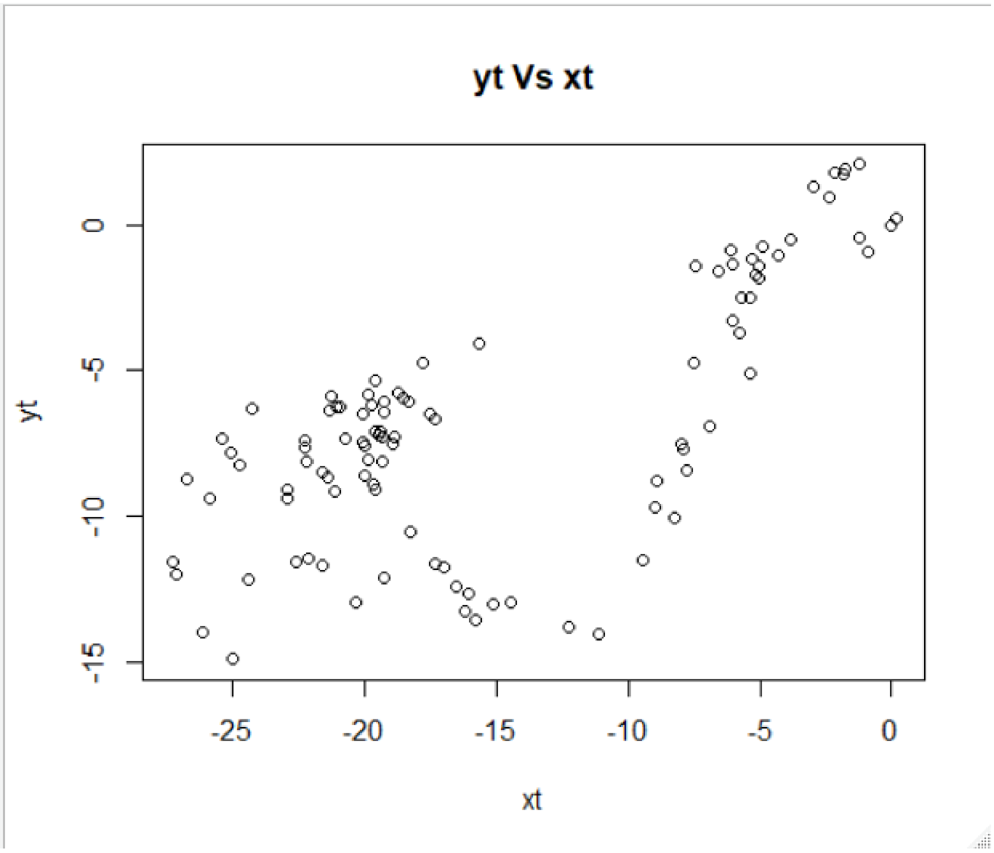
In random walk model, the trend line and the fitted line intersect each other and do not follow the random walk data points, as the predictors here are dependent.

In **b)** part model, the trend line justifies the increase in the mean value in the data and the fit line passes through the center of the data.

II

a)

i)



From the above plot, it's difficult to identify any distinguishable pattern. There is some relation between xt, yt which will be difficult to explain in a linear model.

ii)

No, I expect we would fail to reject the null hypothesis. Because ideally, two random walks are independent.

iii)

```
Call:
lm(formula = yt ~ xt)

Residuals:
   Min     1Q Median     3Q    Max
-8.712 -1.923  1.149  2.210  4.015

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.14335    0.66292  -1.725   0.0877 .
xt           0.37632    0.03914   9.615 7.57e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.106 on 99 degrees of freedom
Multiple R-squared:  0.4829,    Adjusted R-squared:  0.4777
F-statistic: 92.45 on 1 and 99 DF,  p-value: 7.572e-16
```

From the above summary of the linear regression model p-value for beta1 is less than 0.05, hence beta1 is significant, rejecting the null hypothesis. Correlation coefficient is -0.32, suggesting a weak linear relation between xt and yt.

b)

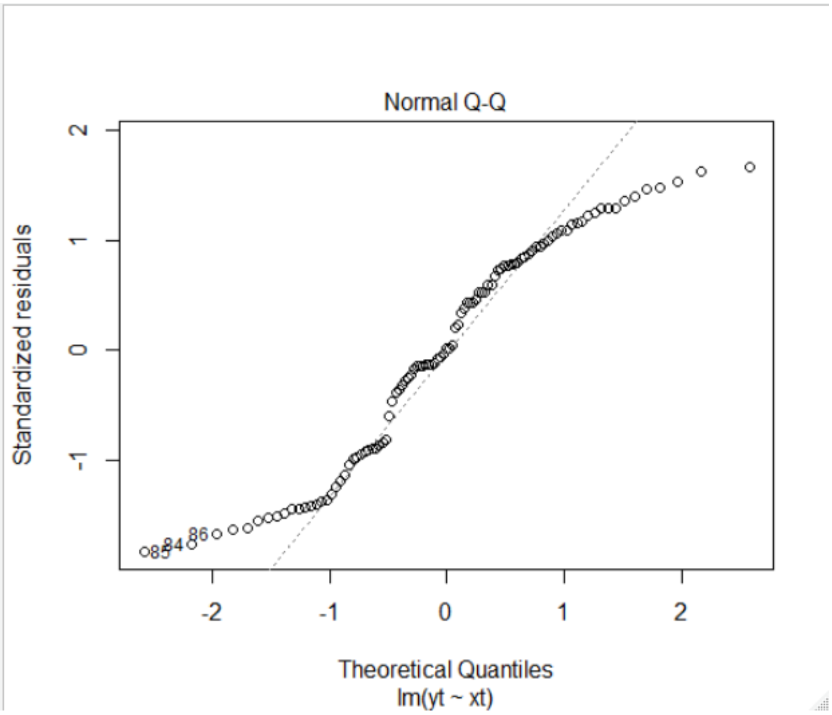
After repeating the experiment 1000 times, 758 times the null hypothesis gets rejected.

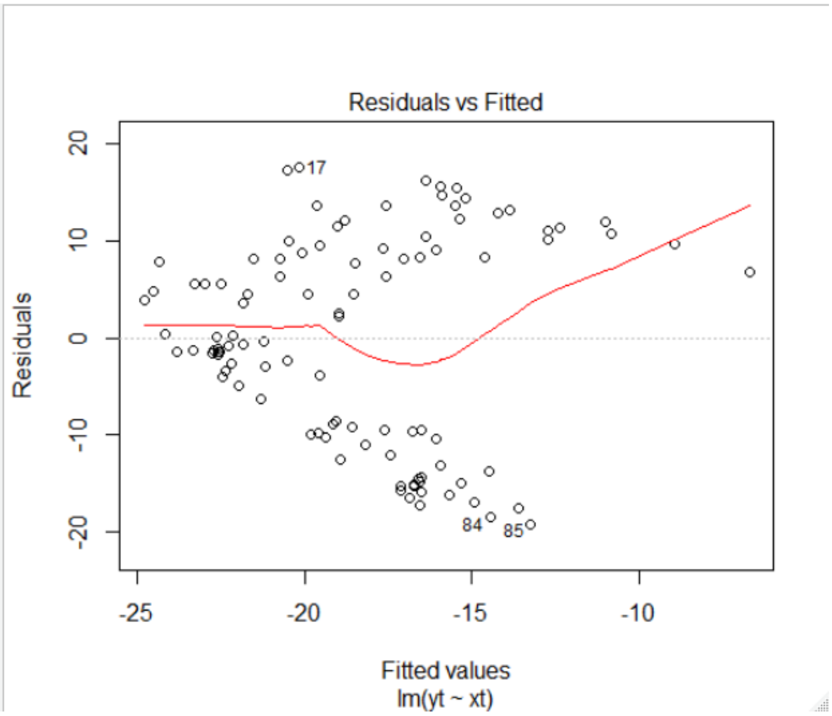
This does not support my expectation in part i.

There are 3 assumptions of inference of regression models: Independence, homogeneity of variances and normality.

From the Normal Q-Q plot (as shown below), the random walks xt, yt violate normality.

From the residual plot, we see the data is fanning out, and thus the variance is not uniform.





III

Multiple Choice Questions

1. A, D
2. B
3. D
4. D
5. B
6. C
7. A
8. A
9. E
10. D