Exam #1

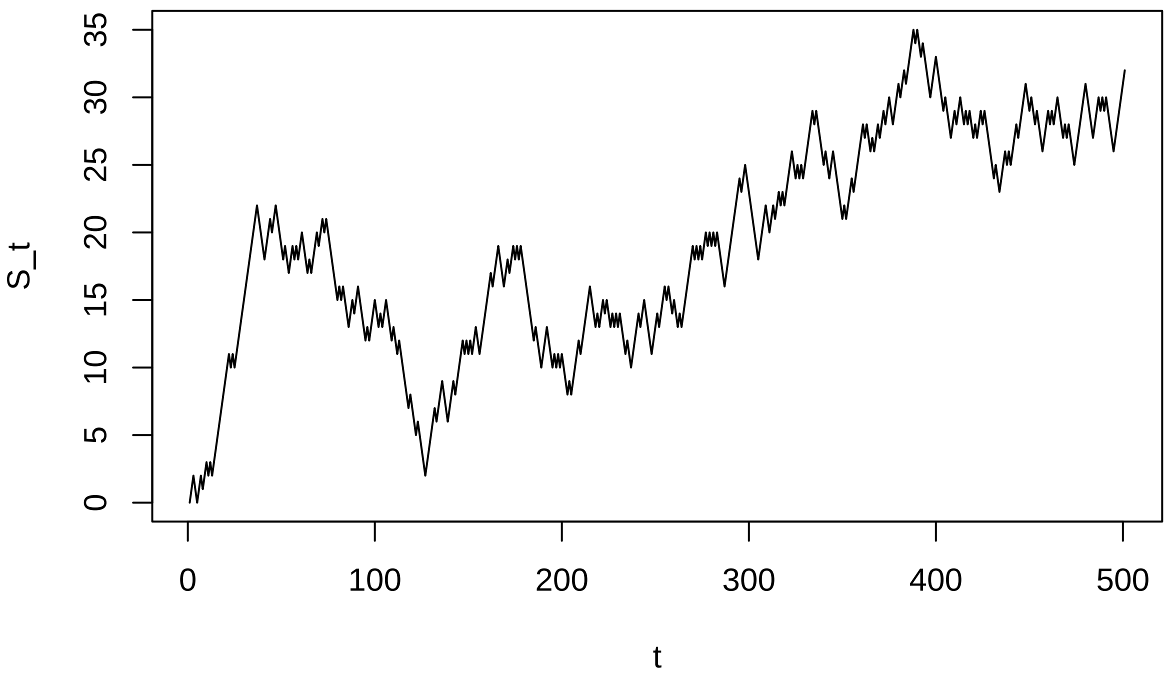
Michael Pena

2024-06-01

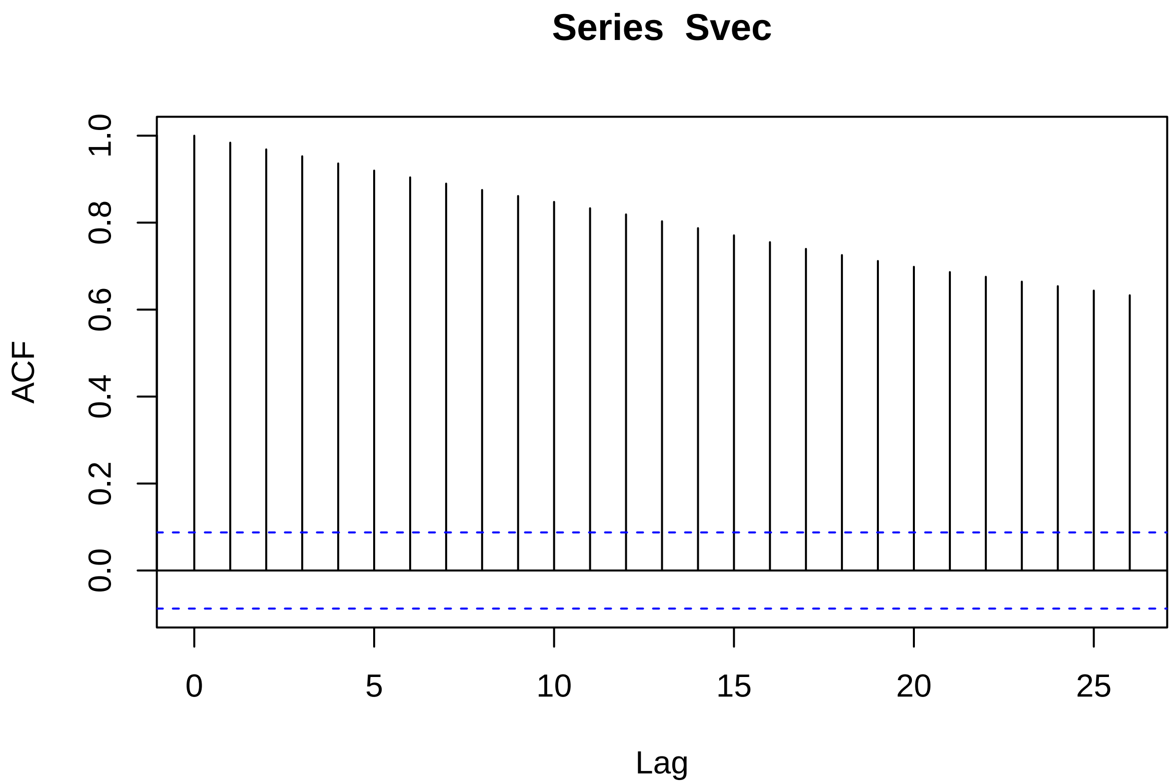
Concepts and Theoretical

1.

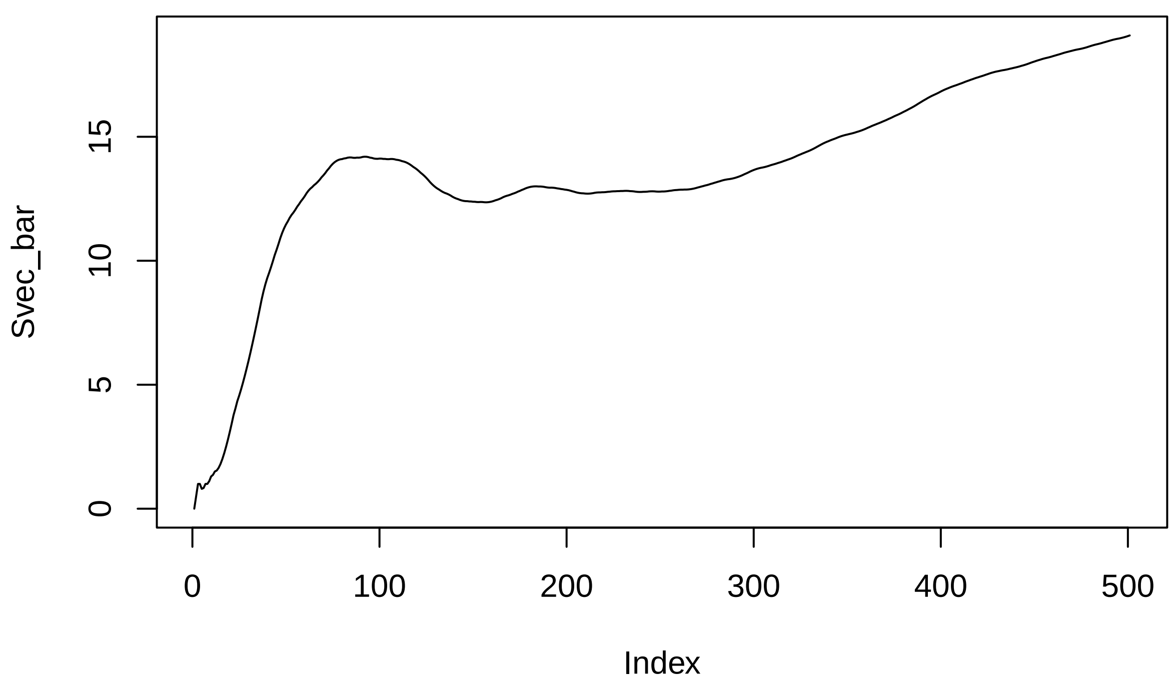
```r
set.seed(536.1)
# initialize X
sample(c(-1,1),size = 500,prob = c(.5,.5),replace = T) -> X
# initialize Svec
Svec = cumsum(c(0,X))
# make graphics
plot.ts(Svec,type = "l",ylab = "S_t", xlab = "t")
```




```r
acf(Svec)
```




```r
# running means
Svec_bar = Svec
for(i in 1:501){
  mean(Svec[1:i]) -> Svec_bar[i]
}
# plot running means
plot(Svec_bar,type = 'l')
```



Above graph showing that the expected value will change depending on time. System is not stationary.

2.

find μ because system is stationary, we know that $E[X_t] = E[X_{t-1}] = E[X_{t-2}] = \mu$ but since

$E[X_t] = E[0.2X_{t-1}+0.4X_{t-2}+Z_t]E[X_t] = E[0.2X_{t-1}+0.4X_{t-2}]+0E[X_t] = 0.2E[X_{t-1}]+0.4E[X_{t-2}]\mu = 0.6\mu$

this can only be true if $\mu = 0$

$X_t = 0.2X_{t-1}+0.4X_{t-2}+Z_t \Rightarrow X_{t-h}X_t = 0.2X_{t-1}X_{t-h}+0.4X_{t-2}X_{t-h}+Z_tX_{t-h} \Rightarrow E[X_{t-h}X_t] = E[0.2X_{t-1}X_{t-h}+0.4X_{t-2}X$

knowing $\rho_X(h) = \frac{\gamma(h)}{\gamma(0)}$ divide the last line all by $\gamma(0)$

$$0 = 0.2\rho_X(h-1) + 0.4\rho_X(h-2) - \rho_X(h)$$

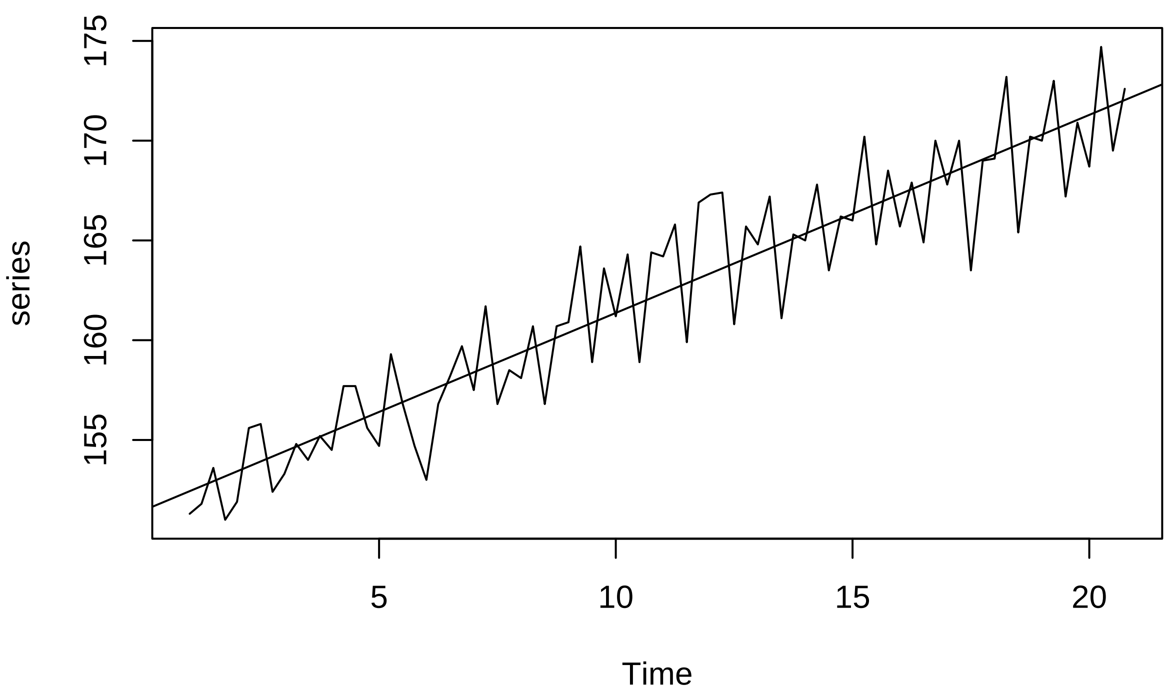
knowing this system is stationary than we know that $\rho_X(0) = 1$, set $h = 1$.

$0 = 0.2\rho_X(0)+0.4\rho_X(-1)-\rho_X(1) \to 0 = 0.2+0.4\rho_X(-1)-\rho_X(1) \to 0 = 0.2-0.6\rho_X(\pm1) \to \frac{1}{3} = \rho_X(\pm1)\rho_X(h) = \begin{cases} 1 \\ 1/3 \\ 0.2\rho_X(h- \end{cases}$

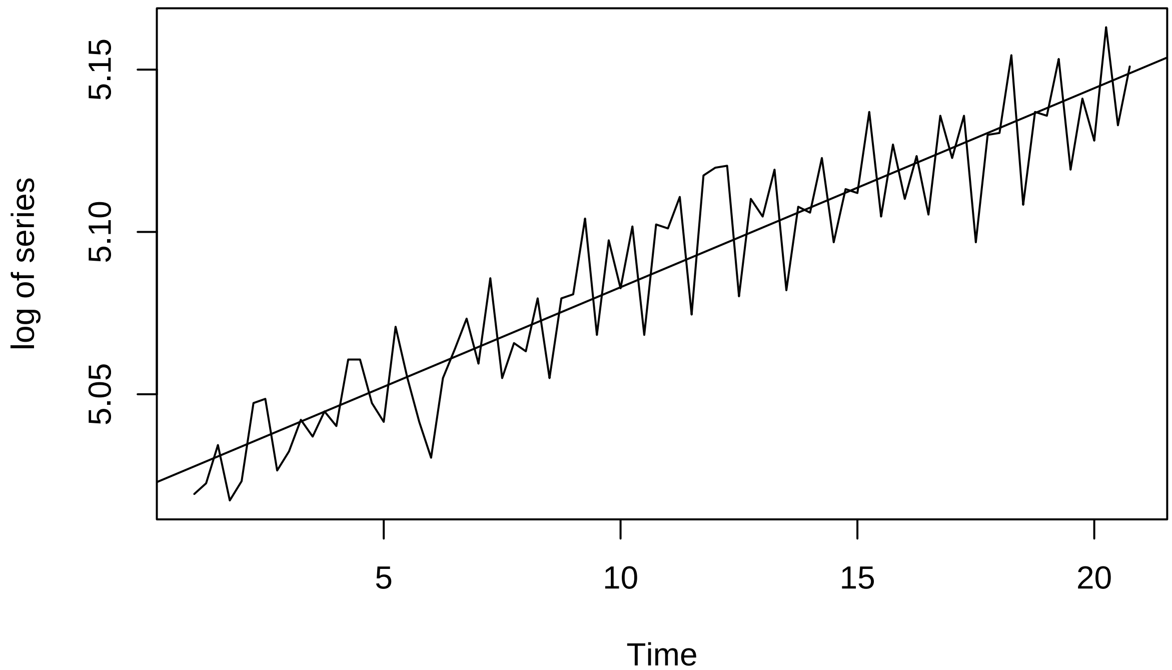
3(a).

Data Analysis

```r
plot.ts(series,type = 'l',ylab = "series")
abline(lm(series ~ time(series),data = data))
```

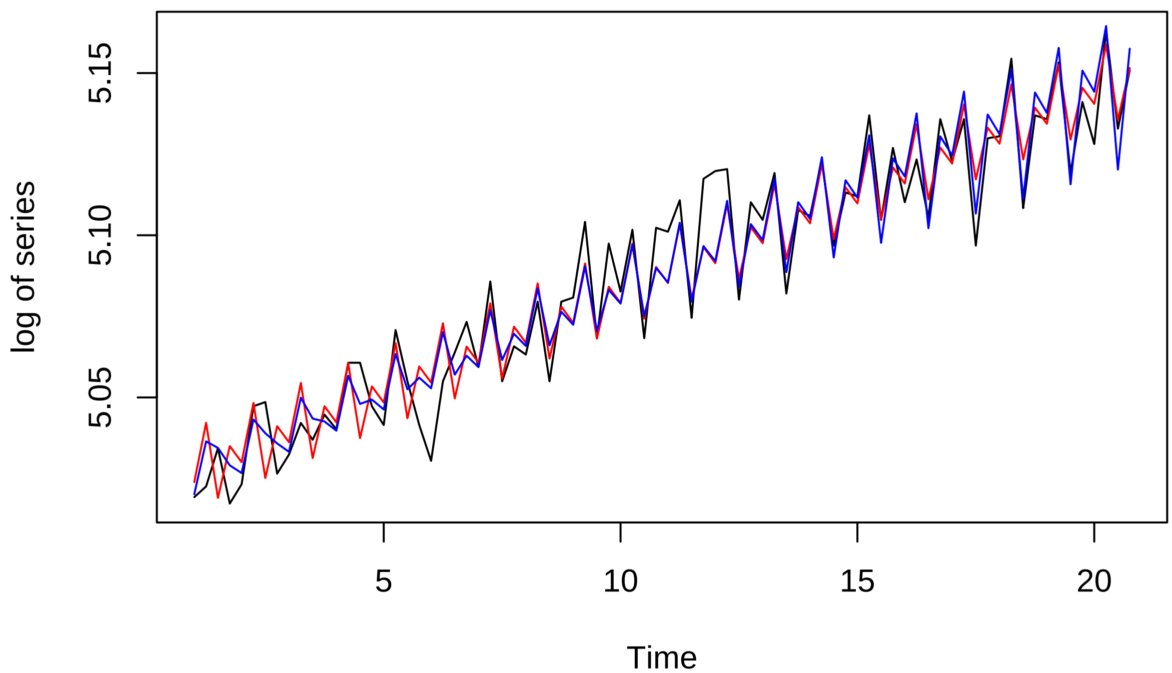



```r
plot.ts(log(series),type = "l",ylab="log of series")
abline(lm(log(series) ~ time(series),data = data))
```

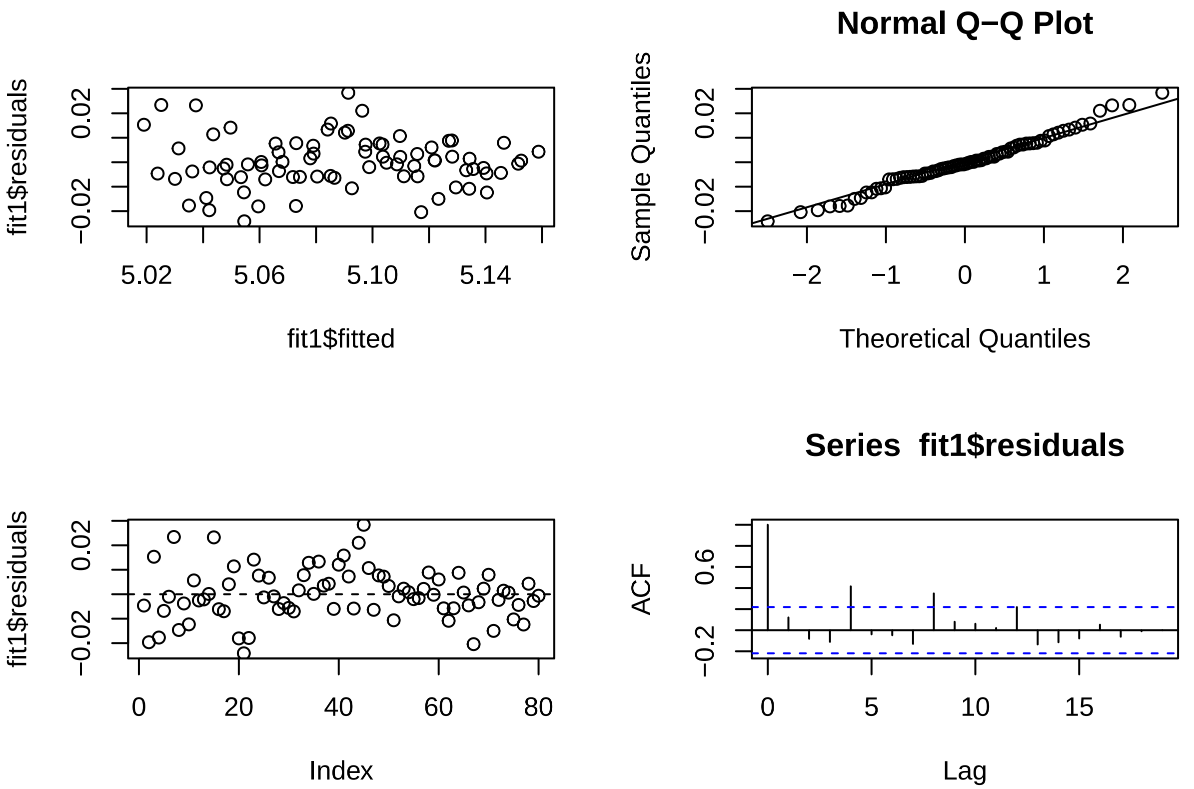



```r
#define a t and cycle
t = time(series)
cyc = as.factor(cycle(series))
```

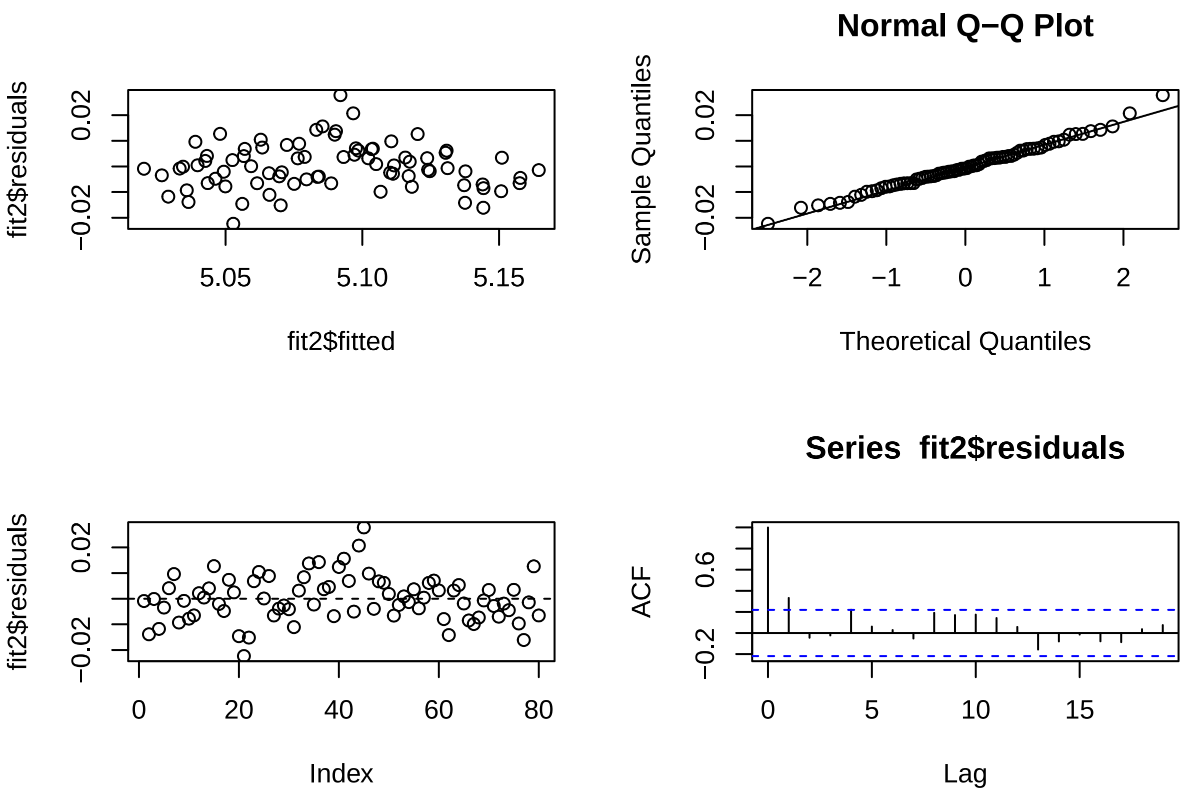
```r
# test two regressions
fit1 <- lm(log(series) ~ cyc + t)
fit2 = lm(log(series) ~ (cyc*t)^3 + cyc*t+ cyc)
plot.ts(log(series),type = "l",ylab="log of series")
points(t,predict.lm(fit1),type = "l", col = "red" )
points(t,predict.lm(fit2),type = "l", col = "blue" )
```



```
# diagnostics for fit 1
par(mfrow=c(2,2)) # Dividing the plotting page into 4 panels
plot(fit1$fitted, fit1$residuals) # plot of fitted values vs residuals
qqnorm(fit1$residuals) #qq-plot of residuals
qqline(fit1$residuals) # plotting the line, along which the dots in qq-plot should lie
plot(fit1$residuals) # plotting the residuals vs time
abline(h=0,lty=2) # plotting a horizontal line at 0
acf(fit1$residuals) #sample acf plot of residuals
```




```
# diagnostics for fit 1
par(mfrow=c(2,2)) # Dividing the plotting page into 4 panels
plot(fit2$fitted, fit2$residuals) # plot of fitted values vs residuals
qqnorm(fit2$residuals) #qq-plot of residuals
qqline(fit2$residuals) # plotting the line, along which the dots in qq-plot should lie
plot(fit2$residuals) # plotting the residuals vs time
abline(h=0,lty=2) # plotting a horizontal line at 0
acf(fit2$residuals) #sample acf plot of residuals
```



Our second model seems to fair better in the diagnostics

```
# get MSE for both
MSE <- function(fit,yhat){
  y <- predict.lm(fit)
  dif <- abs(y - yhat)
  mse <- (1/length(y)) * sum(dif^2)
  return(mse)
}

# get AIC/BIC for both models
ABIC <- function(fit){
  n <- length(predict.lm(fit))
  k <- length(fit$coefficients) - 1
  L <- as.numeric(logLik(fit))
  a <- 2*k - log(L)*2
  b <- k*log(n) - 2*log(L)
  return(list(AIC = a, BIC = b))
}
```

```
sprintf("first model's MSE is %f while the second model has an MSE of %f",MSE(fit1,series),MSE(fit2,ser
```

```
## [1] "first model's MSE is 24732.397190 while the second model has an MSE of 24732.387495"
```

```
sprintf("first model AIC is %f while BIC is %f", ABIC(fit1)$AIC, ABIC(fit1)$BIC)
```

```
## [1] "first model AIC is -3.064345 while BIC is 6.463762"
```

```
sprintf("second model AIC is %f while BIC is %f", ABIC(fit2)$AIC, ABIC(fit2)$BIC)
```

```
## [1] "second model AIC is 2.834924 while BIC is 19.509111"
```

T