

# Exam #1

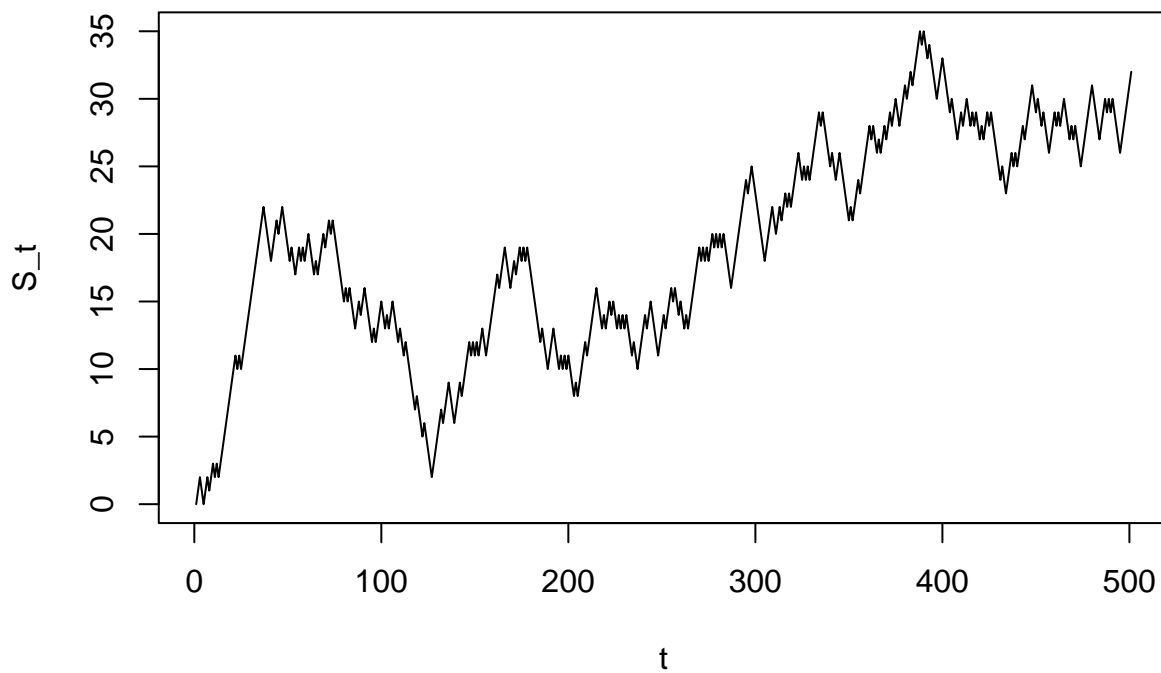
Michael Pena

2024-06-01

## Concepts and Theoretical

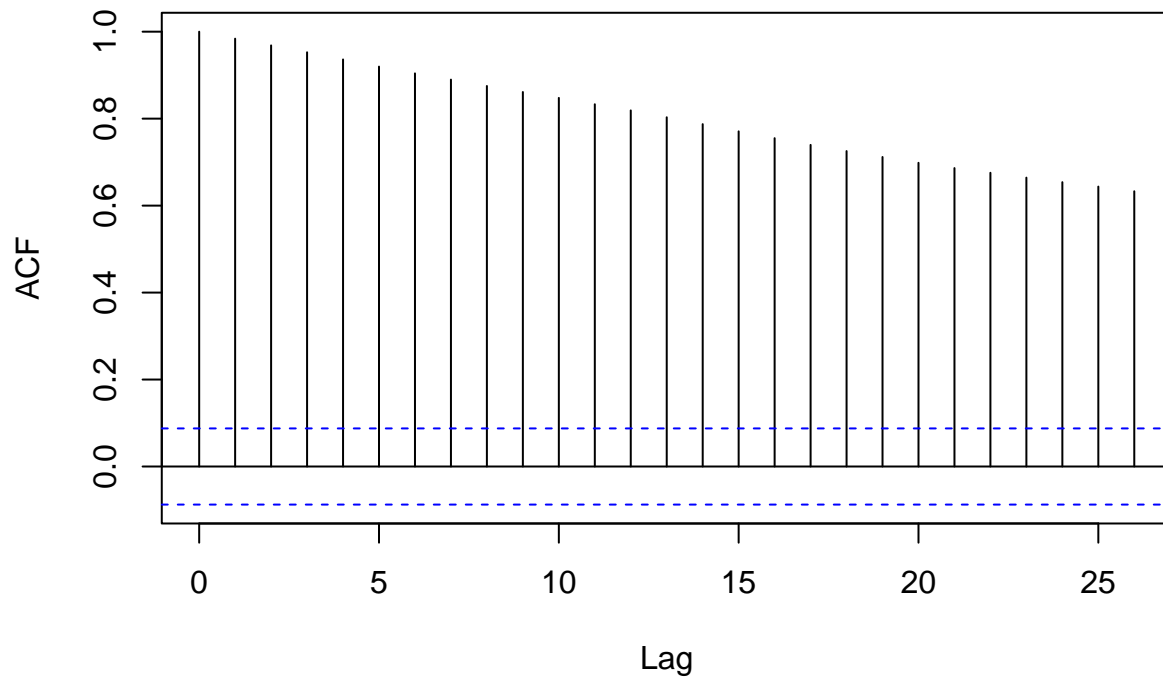
1.

```
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")
```



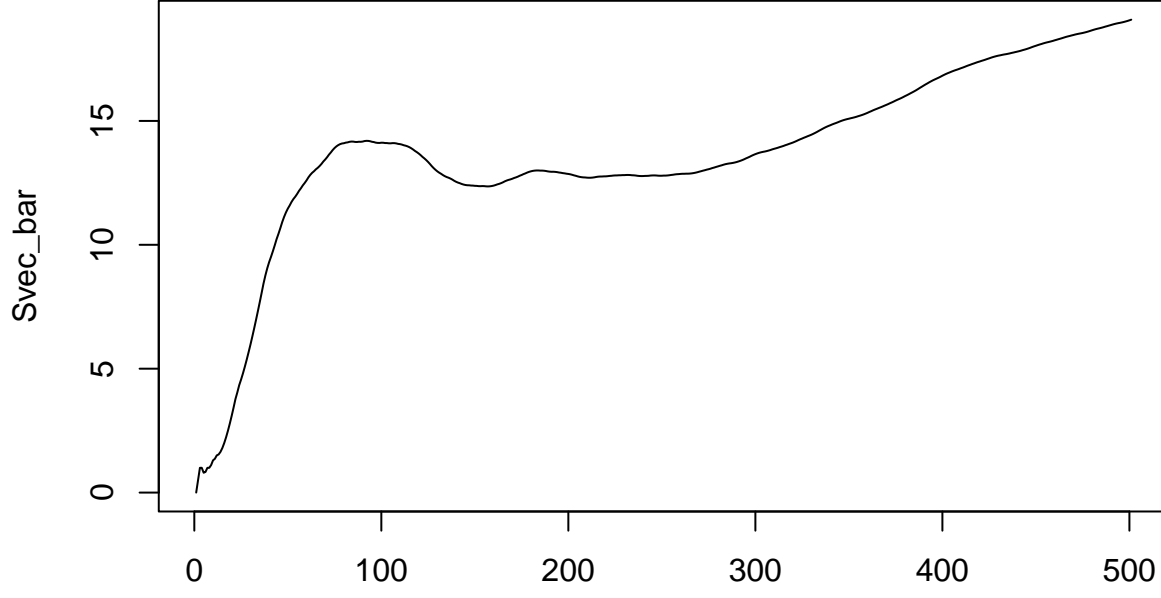
```
acf(Svec)
```

## Series Svec



This slow decay in the ACF may indicate non-stationarity.

```
# 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')
```



Index

Above

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

**2.**

find  $\mu$  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  $\mu = 0$

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

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  $h = 1$ .

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

**3(a).**

$$E[X_t] = E[Z_t] + \theta E[Z_{t-1}] + \theta^2 E[Z_{t-2}] \{Z_t\} \sim WN(0, 9) \Rightarrow E[Z_t] + \theta E[Z_{t-1}] + \theta^2 E[Z_{t-2}] = 0 + \theta \cdot 0 + \theta^2 \cdot 0 = 0$$

$$Cov(X_t, X_{t+h}) = E[(Z_t + \theta Z_{t-1} + \theta^2 Z_{t-2})(Z_{t+h} + \theta Z_{t-1+h} + \theta^2 Z_{t-2+h})]$$

Using the fact that  $Var(X) = E[X^2] - (E[X])^2$  and that  $E[Z_t] = 0$ .

if  $h = 0$  then  $\gamma(0) = 9(1 + \theta^2 + \theta^4)$   $|h| = 1$  then  $\gamma(1) = 9\theta(1 + \theta^2)$   $|h| = 2$  then  $\gamma(2) = 9\theta^2$   $|h| \geq 3$  then  $\gamma(h) = 0$

$$\rho_X(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta(1+\theta^2)}{1+\theta^2+\theta^4} & h = \pm 1 \\ \frac{\theta^2}{1+\theta^2+\theta^4} & h = \pm 2 \\ 0 & \text{otherwise} \end{cases}$$

3(b).

$$\text{corr}(X_t, X_{t-2}) = \rho_X(-2) = \frac{\theta^2}{1 + \theta^2 + \theta^4}$$

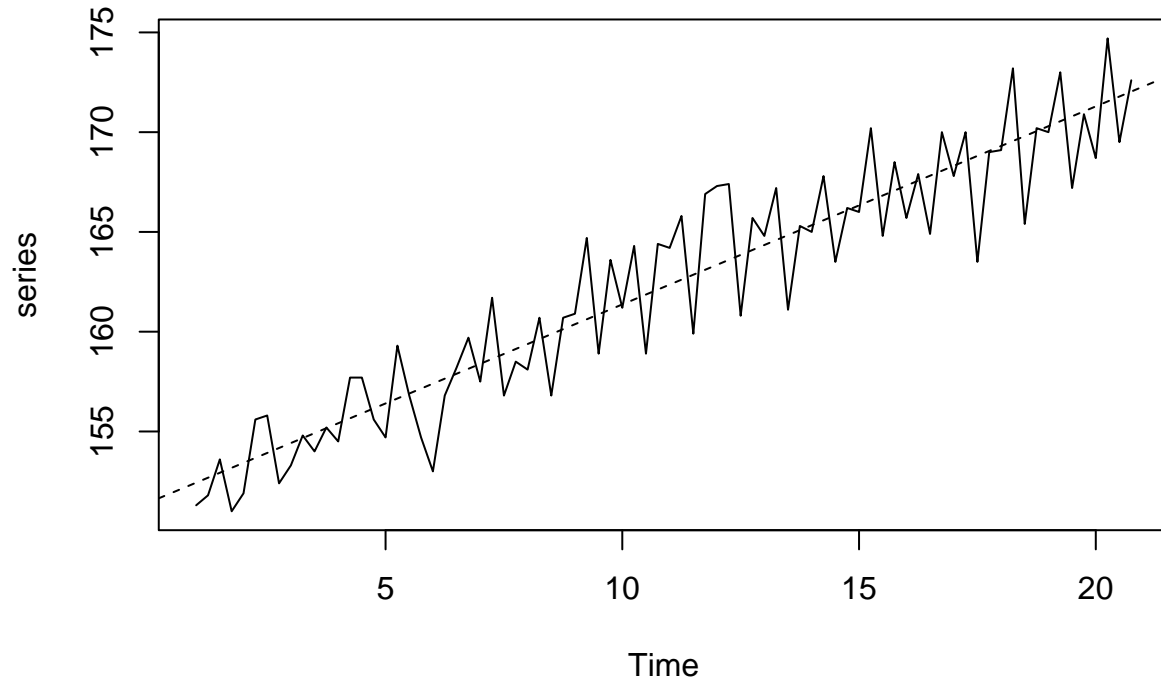
Allow some  $f$  such that  $f(\theta) = \frac{\theta^2}{1+\theta^2+\theta^4}$  knowing  $\theta < \infty$

$$f'(\theta) = \frac{-2\theta(\theta^4 - 1)}{(1 + \theta^2 + \theta^4)^2} \Rightarrow 0 = -2\theta(\theta^4 - 1) \Rightarrow \theta = \{-1, 0, 1\} f(\pm 1) = 1/3 f(0) = 0$$

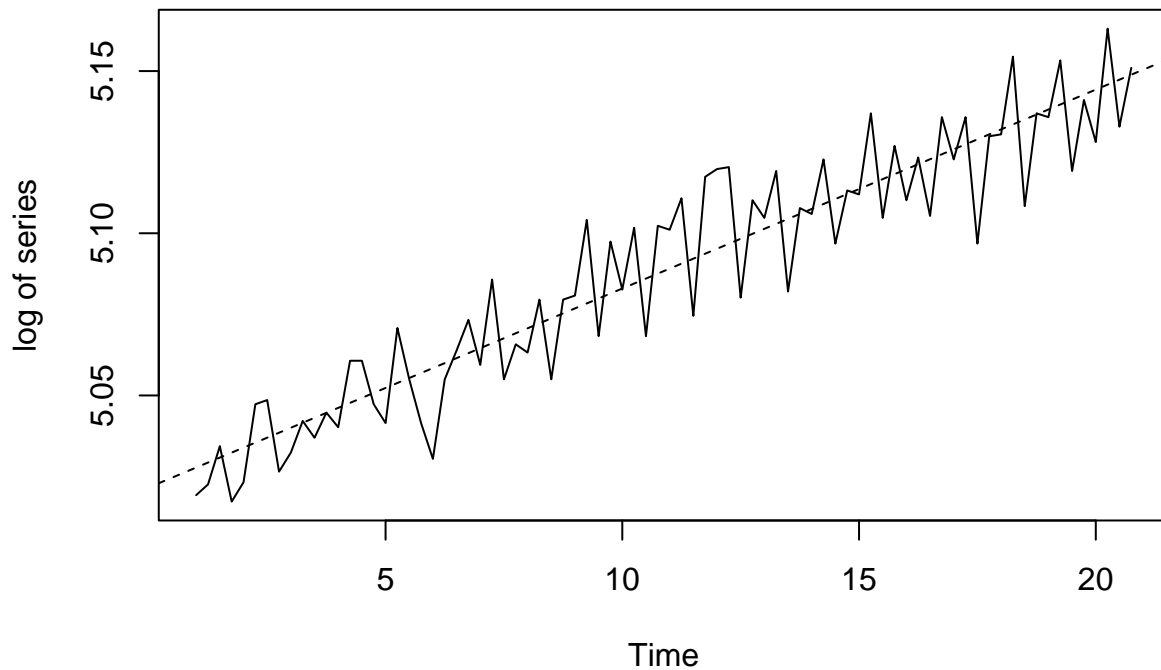
this makes that maximum  $\frac{1}{3}$  and the minimum is 0.

## Data Analysis

```
plot.ts(series,type = 'l',ylab = "series")
abline(lm(series ~ time(series),data = data),lty = 2)
```



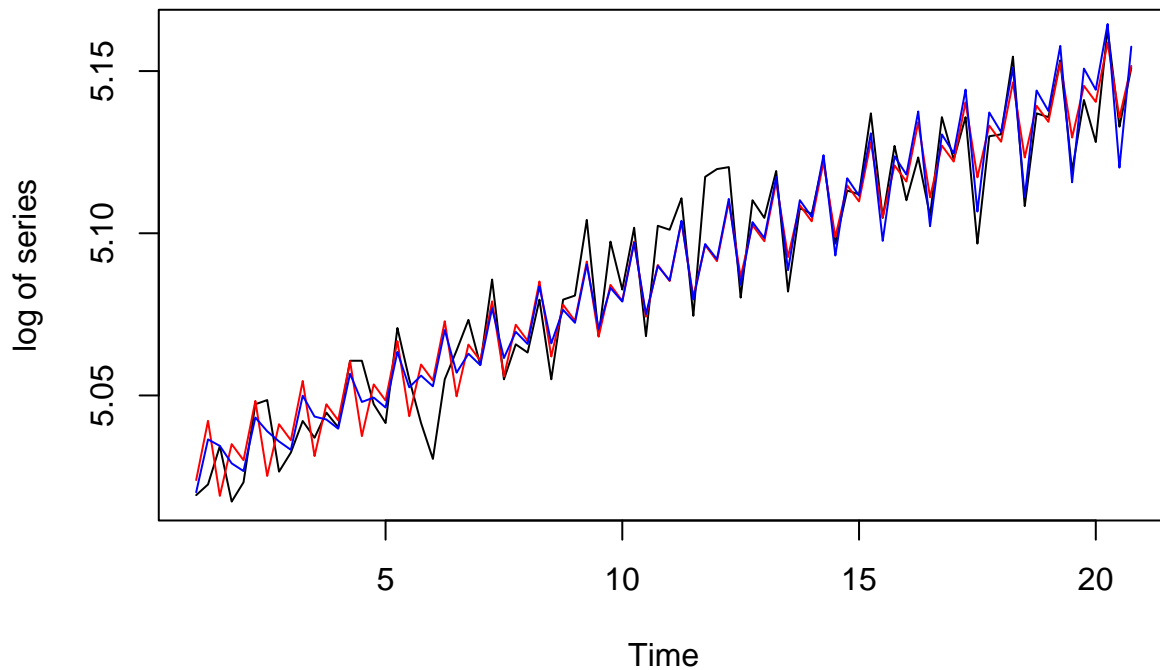
```
plot.ts(log(series),type = "l",ylab="log of series")
abline(lm(log(series) ~ time(series),data = data), lty =2)
```



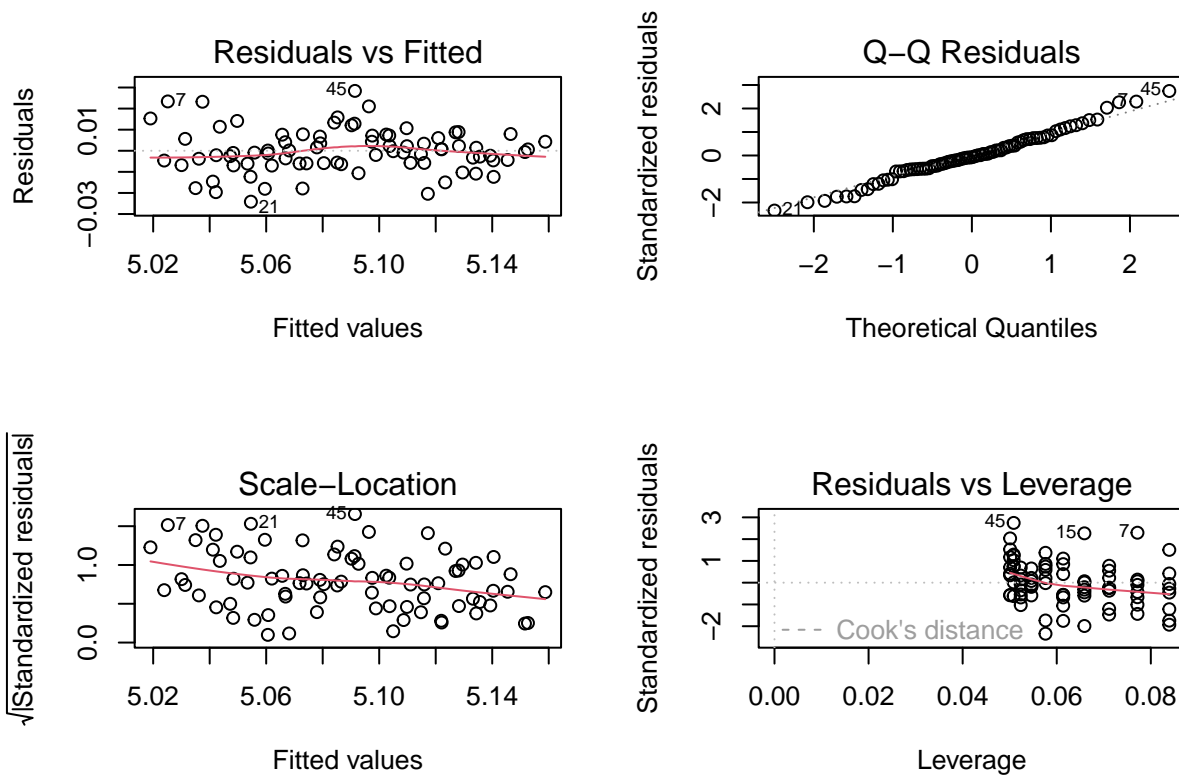
Looking at both plots, there is a clear increase after each cycle. While this business seems to experience drops in sales on a seasonal basis, there seems to be a season in which the sales increase greater than in the last cycle, causing an overall upward trend.

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

# 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")
```

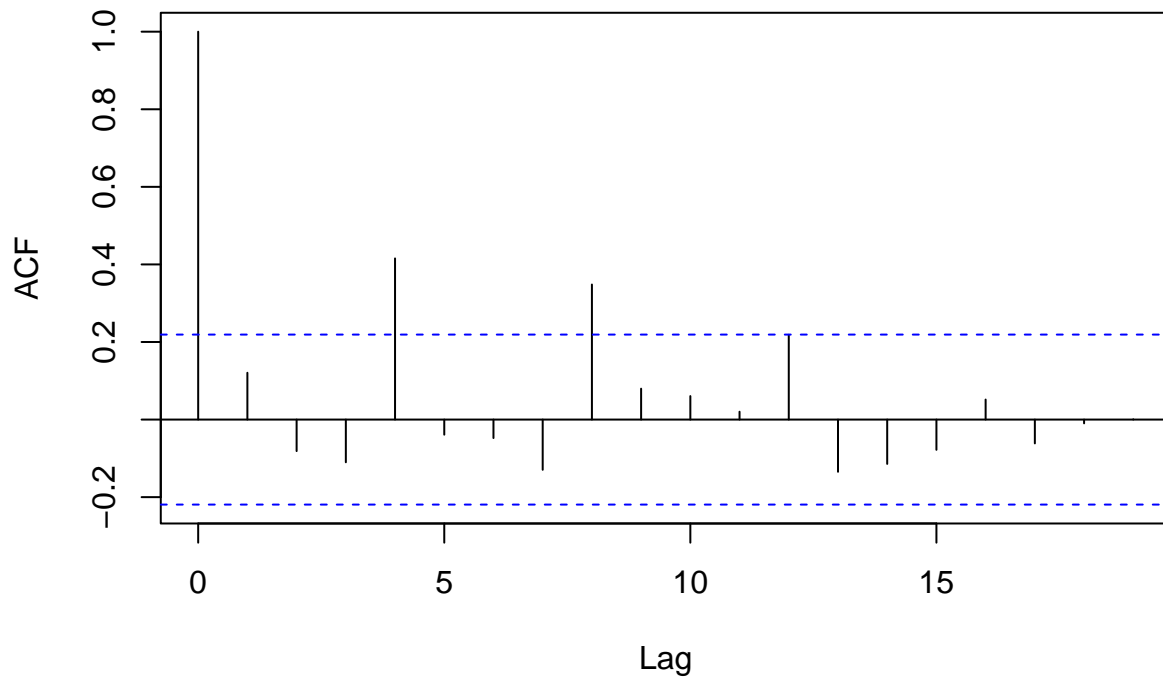


```
# diagnostics for fit 1
par(mfrow=c(2,2)) # Dividing the plotting page into 4 panels
plot(fit1)
```

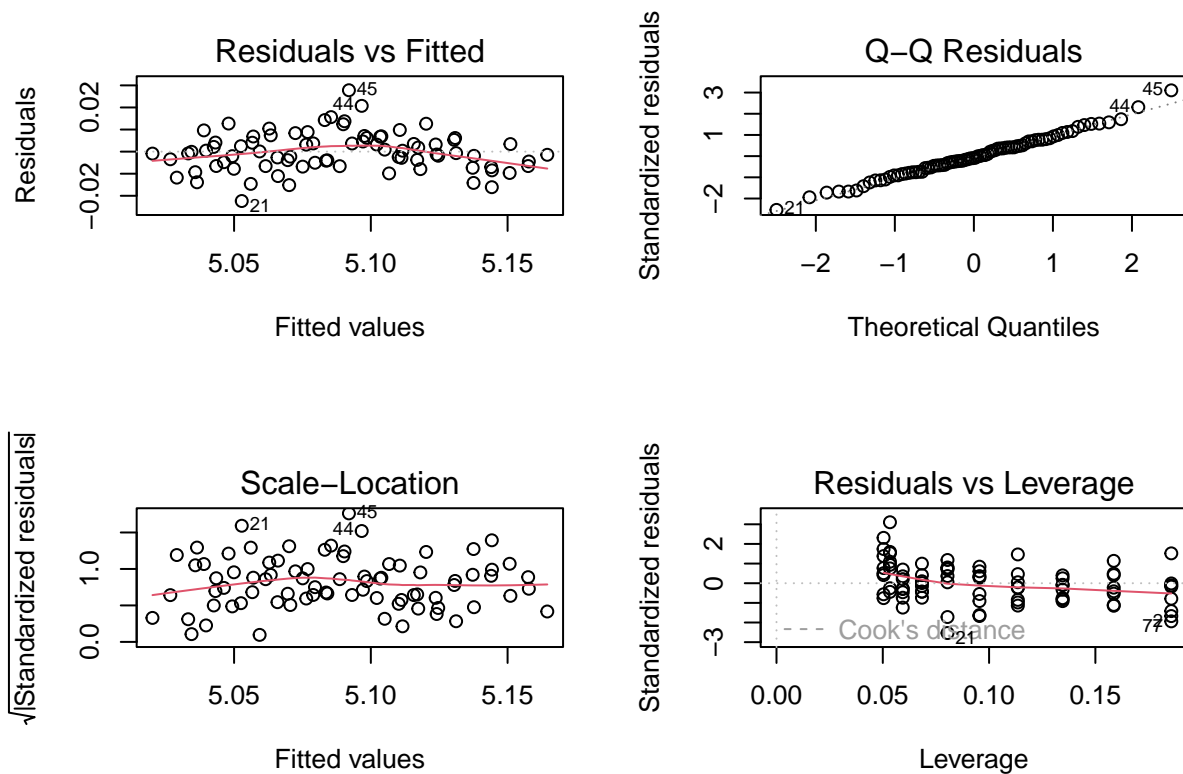


```
acf(fit1$residuals) #sample acf plot of residuals
```

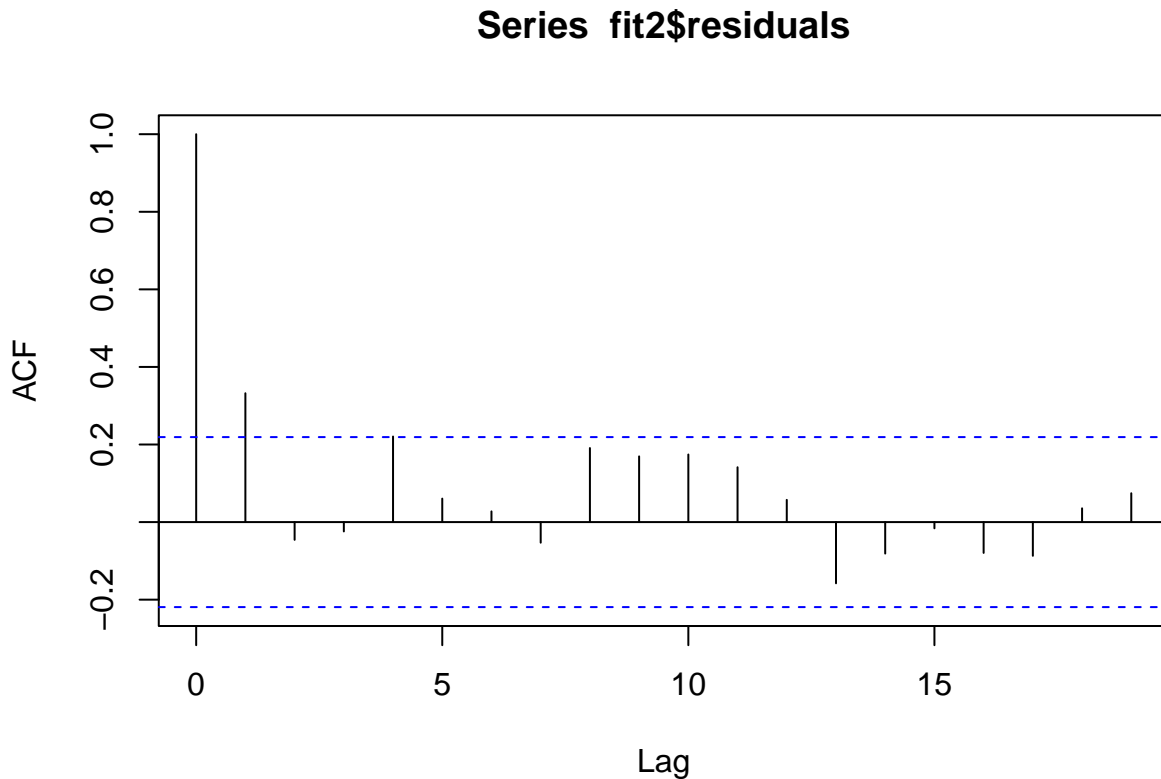
## Series fit1\$residuals



```
# diagnostics for fit 1
par(mfrow=c(2,2)) # Dividing the plotting page into 4 panels
plot(fit2)
```



```
acf(fit2$residuals) #sample acf plot of residuals
```



Our second model seems to fair better in the diagnostics. Model 2 exhibits more normality. ACF has little correlation by the 3rd lag.

```
# 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"
```

The MSE seems to indicate that the 2nd model is only slightly better at fitting the series than the 1st model. As for the AIC and BIC, model 1 scores lower and is thus the better model to balance out the overfitting. Because the difference in MSE is negligible (0.009695), but an obvious difference in AIC and BIC, it seems more reasonable for the business to choose model 1 for prediction.