Ran Wang
S111503
Exercise 03

Identification

Order of differencing

Apparently, the data has a clear linear trend sand seasonal pattern (see figure1), so it is not stationary.

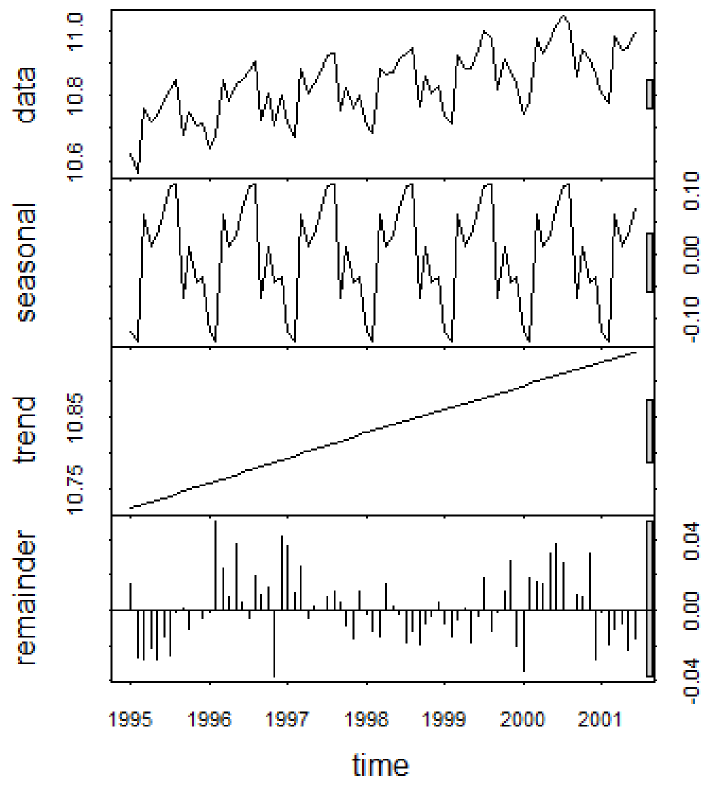


Figure 1 X1 seasonal decomposition

In order to make it stationary, the linear trend and seasonal pattern must be removed by using a suitable differencing is applied to x1. In this case I used one diff for reduce the linear trend and another for seasonality respectively. The result is clear before and after applied differencing that the linear trend is gone, no spike in the PACF, and its residual appears white noise (See figure 2 and 3).

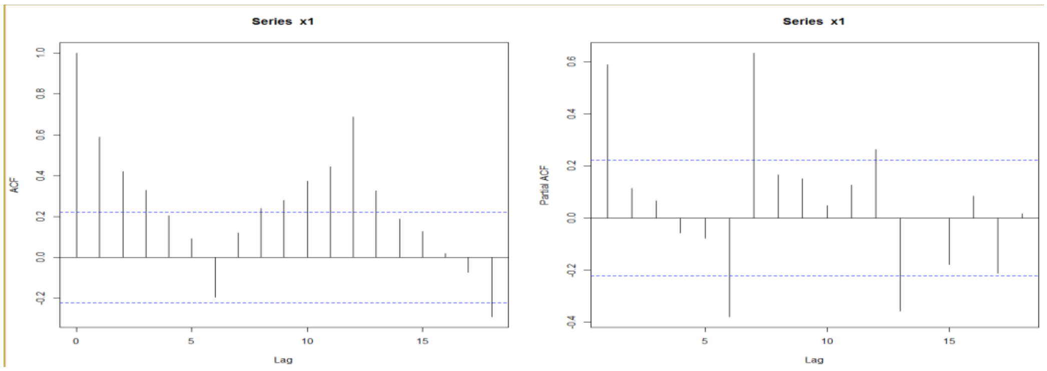


Figure 2 ACF and PACF before differencing

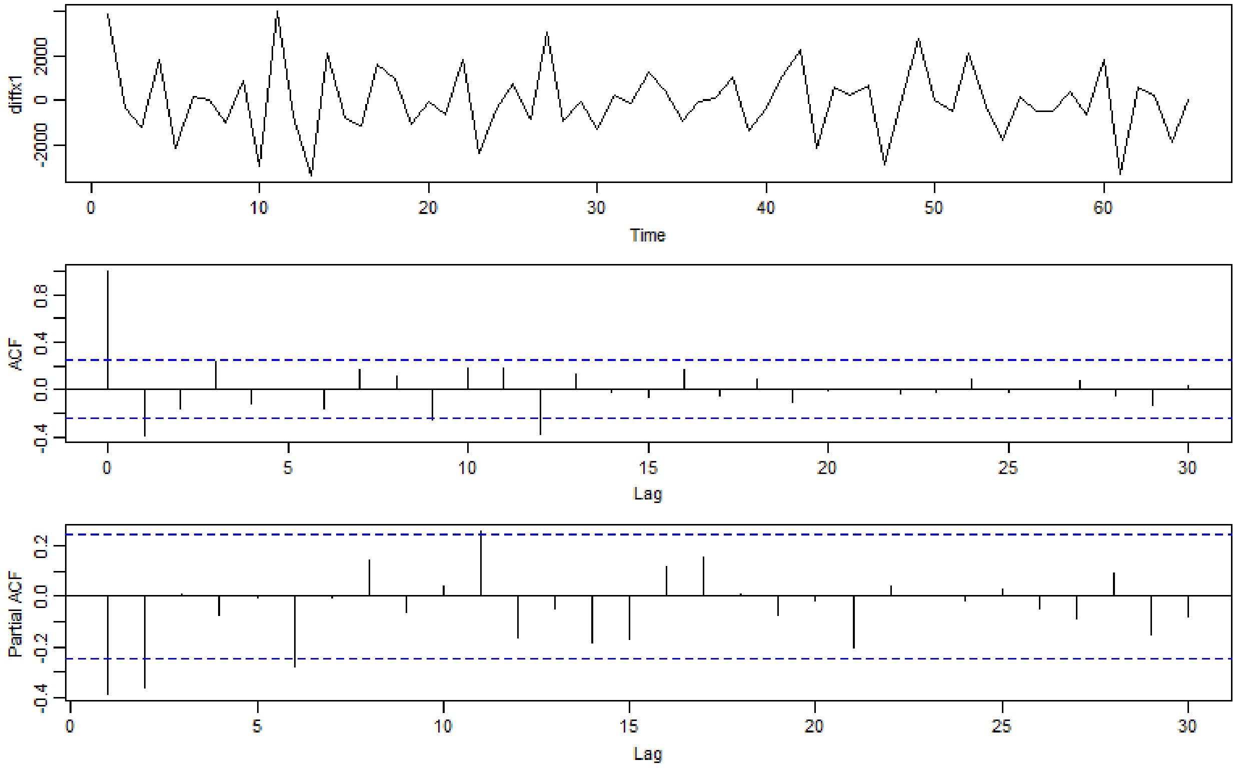


Figure 3 after differencing

Model building and Estimation

Based on the differenced ACF and PACF, I can identify the order of an ARIMA (p,d,q) model. The proposal for both non-seasonal and seasonal parts are given as following:

Non-seasonal part:

Figure 3 PACF has a blank graph after lag 2, which indicates an AR(2) model, as Φ_{kk}=0 for k>p (P155 Table6.1). On the other hand, ACF has a significant spike at lag 1, which implies an MA(1) interpretation.

Seasonal part

ACF shows a couple of significant spikes at lag 1 and 12, nothing after 12, which indicates a $MA(1)_{12}$.

I did a iterative process for choosing the optimal ARIMA order by slightly changing the p,d,q, the best model is (2,1,1) x $(0,1,1)_{12}$, because its variance is the smallest among all other combination of orders. I could also have looked at AIC, but since the data is differenced (in both models), I should not focus so much on AIC. I will give two of the examples comparing model buildings in the following. The residual plots(see figure 4 and 5) and the ACF, PACF look all fine, no significant spikes are out of the bound.

```
Call:
arima(x = x1, order = c(1, 1, 1), seasonal =list(order = c(0, 1, 1), period = 12),method = "ML")
Coefficients:
         ar1      ma1     sma1
      0.1172  -0.7310  -0.4710
s.e.  0.2093   0.1508   0.1286

sigma^2 estimated as 1375379:  log likelihood = -553.41,  aic = 1114.82
```

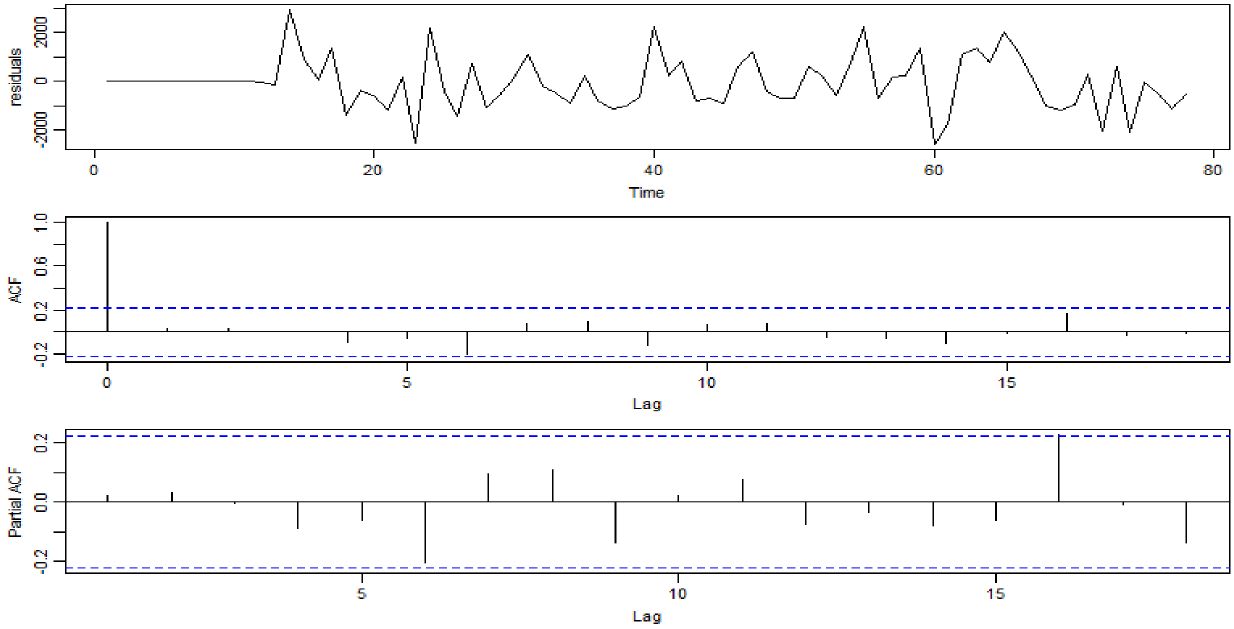


Figure 4 Model 1 residual plot

```
Call:
arima(x = x1, order = c(2, 1, 1), seasonal=list(order = c(0, 1, 1), period = 12), method = "ML")
Coefficients:
          ar1      ar2      ma1     sma1
      -0.2000  -0.2726  -0.3961  -0.4567
s.e.   0.3613   0.2091   0.3810   0.1250

sigma^2 estimated as 1351640:  log likelihood = -552.72,  aic = 1115.44
```

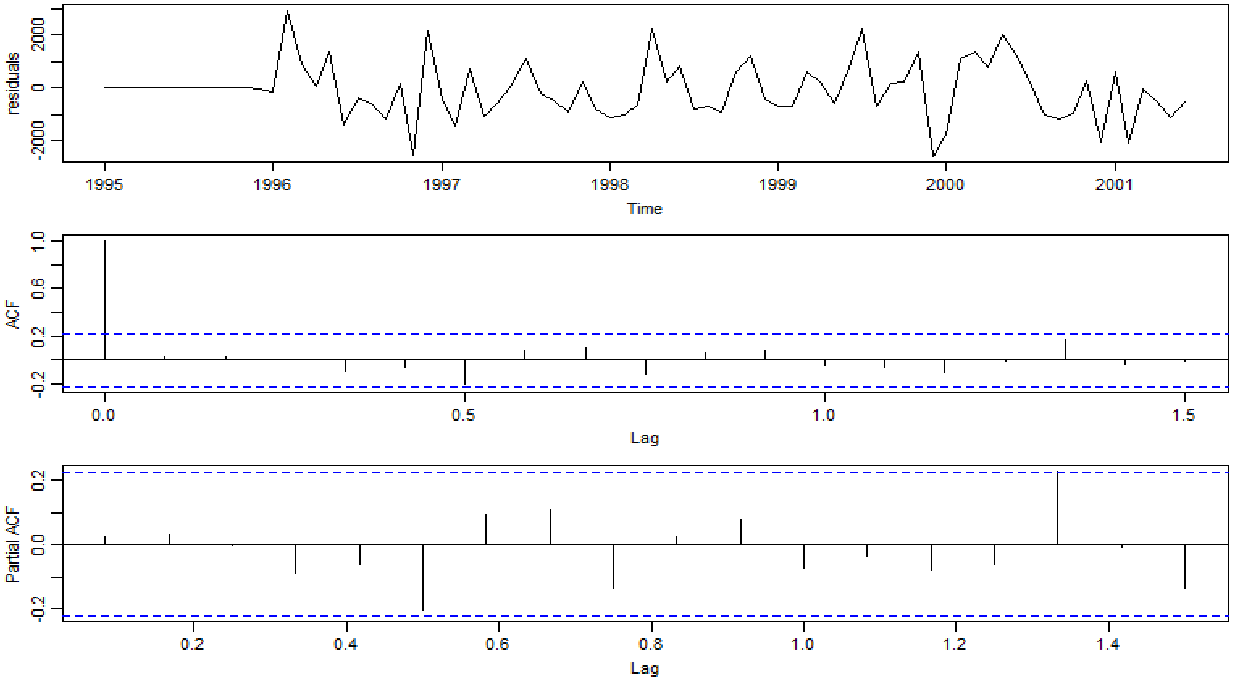


Figure 5 Model 2 residual plot

Model checking

Plot checking

Figure6 shows that the model error is identically, independently and normally distributed white noise. The errors are in the 95% CI around the theoretical white noise (the diagonal line from 0,0 to 0.5,1) (P176). P value for Ljung Box are all well above 0.05, which mean non-significant autocorrelation in residuals.

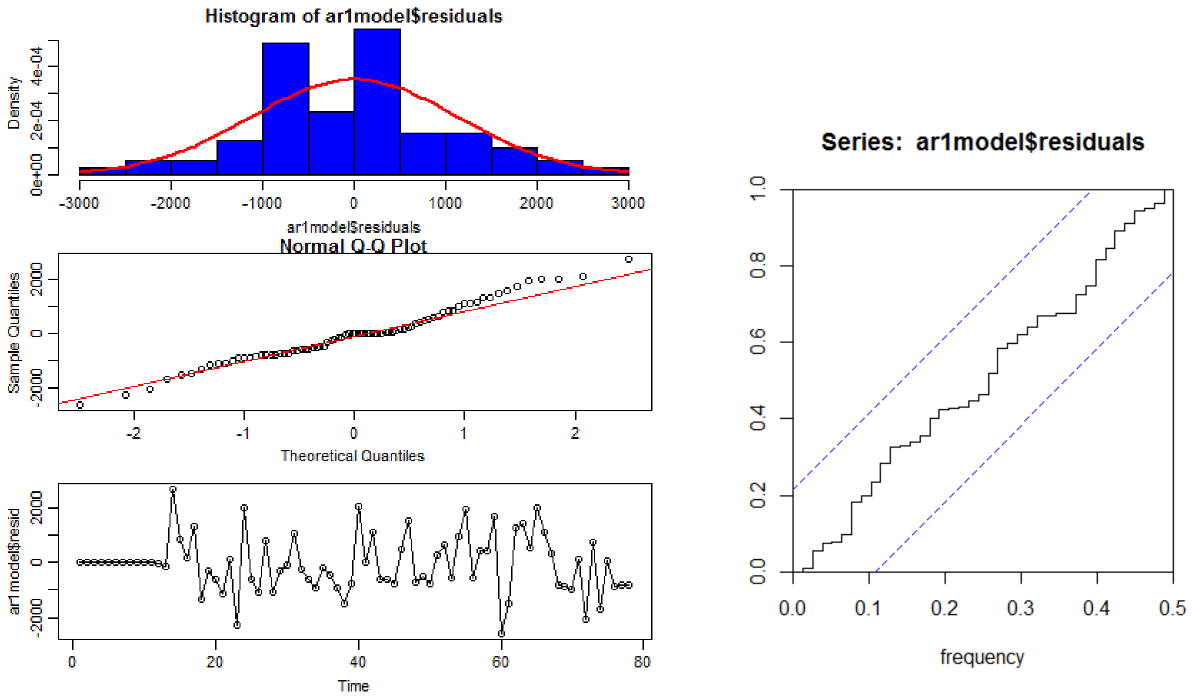


Figure 6 Model error distribution

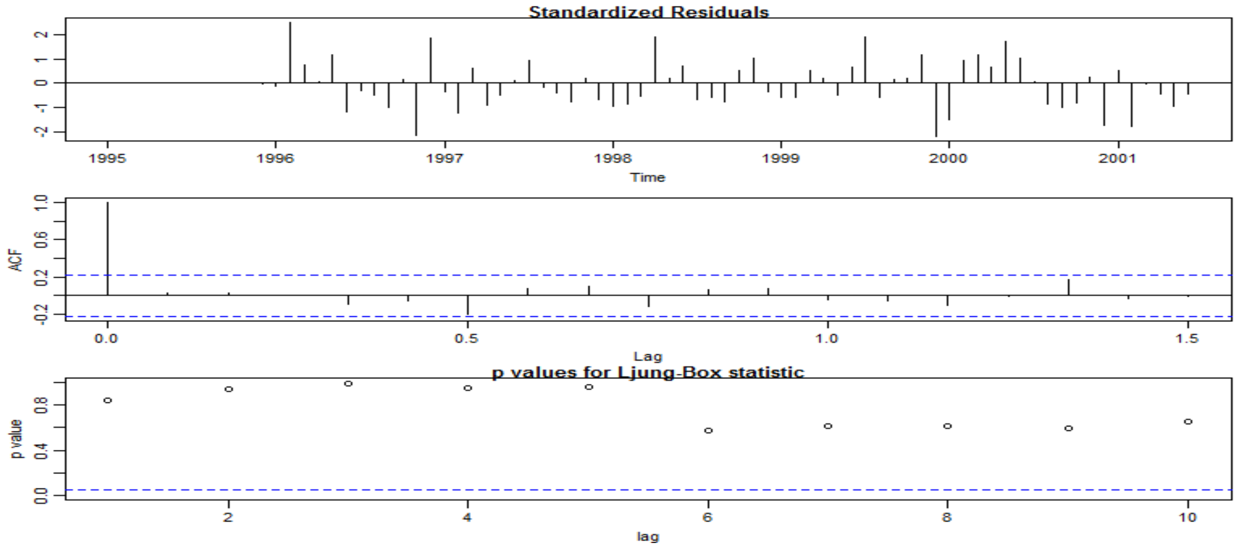


Figure 7 standard residual and Ljung Box

Test of change in signs:

By using R code (See Appendix) firstly I get the mean of the times of sign changes is 38.5, sd=4.3874. Then have its 95% CI 29.90 to 47.09, the final sign changes is 29 that falls into the 95% interval. Thus it proves the error being white noise.

Tests for lower model order

If I try to lower model (0, 1, 0) x seasonal(0, 1, 1) gives slightly higher variance (1850250

) even though the residual, ACF, PACF histogram plots look all fine (see figure 8 and 9), but the Ljung Box gives the p values that are all below 0.05, which indicates significant autocorrelation in the residual series. And the qq plot looks slightly worse than the previous two models. All in all, the lower order model has significant evidence showing it is not a pure white noise.

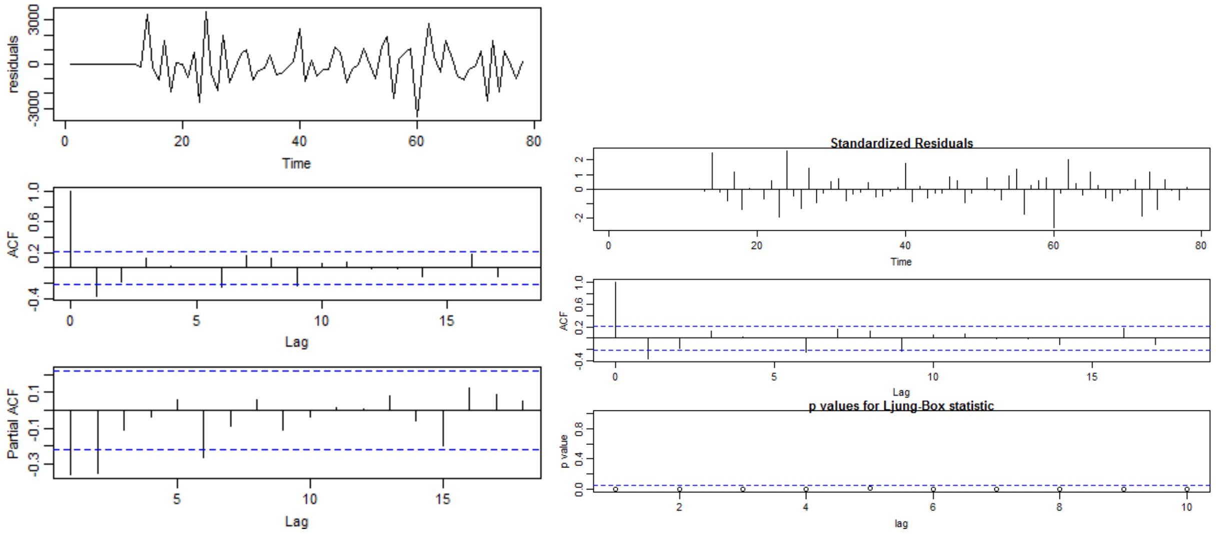


Figure 8 Low order arima plot (0,1,0) x (0,1,1)s

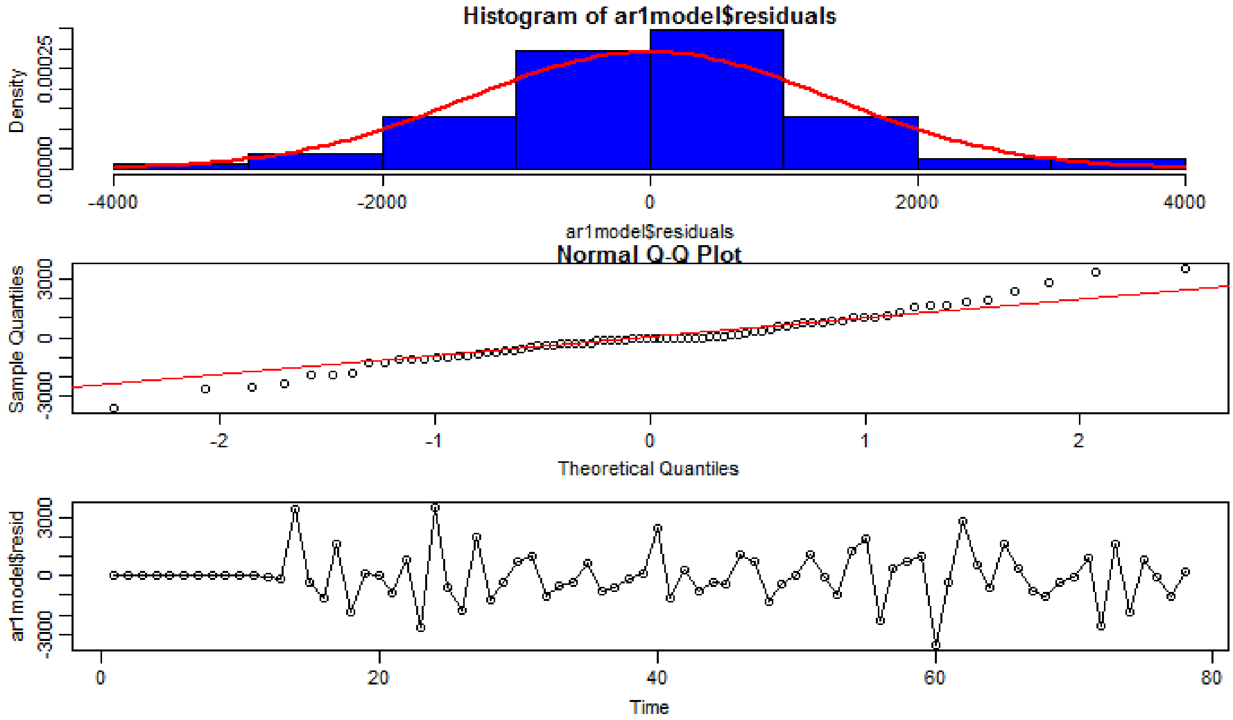


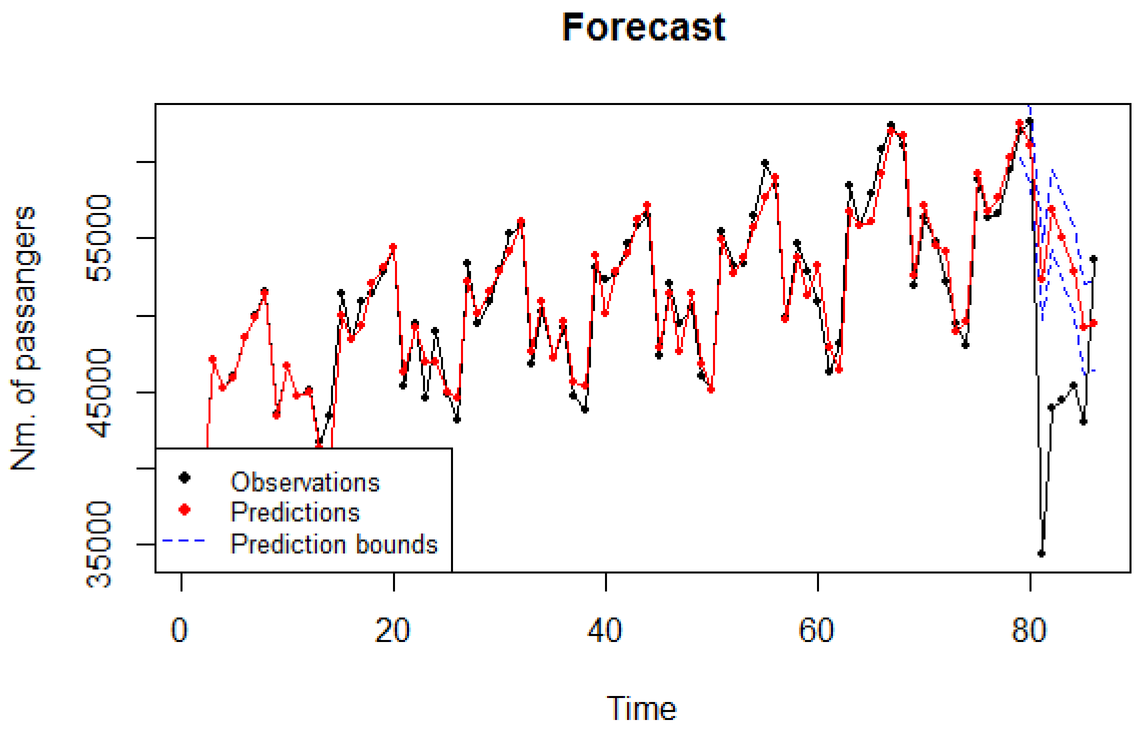
Figure 9 Normal distribution

Prediction

The prediction could not foresee terror attack based on the previous observations, this element is not in the model. However, the predictions tend to converge with the observations except September 2001 (see figure 10), which indicates that the model is still durable provided that the event happens regularly.

```
$pred
          Jan      Feb Mar Apr May Jun      Jul      Aug      Sep      Oct
   Nov      Dec
2001                                   62035.18 61002.04 51912.84 56418.06 54
482.27 52963.67
2002 49384.40 48898.45


$se
          Jan      Feb Mar Apr May Jun      Jul      Aug      Sep      Oct
   Nov      Dec
2001                                   1162.638 1253.875 1287.256 1386.772 14
80.894 1549.972
2002 1619.356 1690.013
```



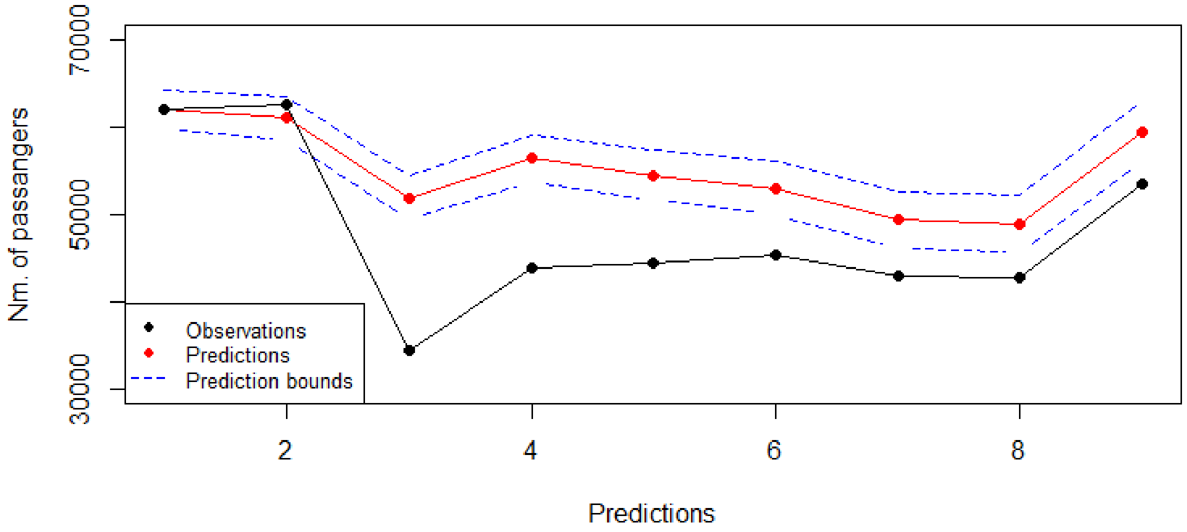


Figure 10 Forecast with limit compare with xc2

Appendix

```
dat <- read.table("assignment3data.txt")
x <- ts(dat,frequency=12,start=1995)
x <- x[,1]
x1 <- x[1:78]
x2 <- x[-(1:78)]
x3 <- ts(x1,frequency=12,start=1995)
ts.plot(x1)
var(x1)
mean(x1)
#seasonal decomposition
plot(stl(log(x3),s.window = "per", t.window = 101,na.action = na.omit))
par(mfrow=c(1,2),mar=c(3,3,1,1),mgp=c(2,0.7,0))
acf(x1)
pacf(x1)
#differencing
diffx12=diff(x1,1,1) #Wt
diffx1=diff(diffx12,12,1)
par(mfrow=c(3,1))
ts.plot(diffx1)
acf(diffx1, lag.max = 30,na.action = na.omit)
pacf(diffx1,lag.max = 30, na.action = na.omit)

diagtool <- function(residuals){
  par(mfrow=c(3,1),mar=c(3,3,1,1),mgp=c(2,0.7,0))
  plot(residuals)
  acf(residuals)
  pacf(residuals)
}
diagtool(diffx1)

ar1model = arima(x1, order = c(0,1,0), method ="ML",seasonal = list(order = c(0, 1, 0), period = 12),)
ar2model = arima(x1, order = c(2,1,1), method ="ML",seasonal = list(order = c(0, 1, 1), period = 12),)
ar2model
ar1model
#model1
diagtool(ar1model$residuals)
tsdiag(ar1model)
hist(ar1model$residuals,probability=T,col='blue')
curve(dnorm(x,sd=sqrt(ar1model$sigma2)), col=2, lwd=2, add = TRUE)
qqnorm(ar1model$residuals)
qqline(ar1model$residuals,col=2)
plot(ar1model$resid,type="o")
par(mfrow=c(1,2))
cpgram(x1)
cpgram(ar1model$residuals)
#model2
diagtool(ar2model$residuals)
tsdiag(ar2model)
hist(ar2model$residuals,probability=T,col='blue')
curve(dnorm(x,sd=sqrt(ar2model$sigma2)), col=2, lwd=2, add = TRUE)
qqnorm(ar2model$residuals)
qqline(ar2model$residuals,col=2)
plot(ar2model$resid,type="o")
par(mfrow=c(1,2))
cpgram(x1)
cpgram(ar2model$residuals)

# sign test mean and sd:
(length(ar1model$residuals)-1)/2
#binom.test((length(ar1model$residuals)-1)/2,length(ar1model$residuals)-1)
### sd:
sqrt((length(ar1model$residuals)-1)/4)
### 95% interval:
(length(ar1model$residuals)-1)/2 + 1.96 * sqrt((length(ar1model$residuals)-1)/4) * c(-1,1)
### test:
res <- ar1model$residuals
(N.sign.changes <- sum( res[-1] * res[-length(res)]<0 ))

# sign test mean and sd:
(length(ar2model$residuals)-1)/2
```

```
### sd:
sqrt((length(ar2model$residuals)-1)/4)
### 95% interval:
(length(ar2model$residuals)-1)/2 + 1.96 * sqrt((length(ar2model$residuals)-1)/4) * c(-1,1)
### test:
res <- ar2model$residuals
(N.sign.changes <- sum( res[-1] * res[-length(res)]<0 ))

# Likelihood ratio test
pchisq(-2* ( ar1model$loglik - ar2model$loglik ), df=1,lower.tail = FALSE)

# F-test
s1 <- sum(ar1model$residuals^2)
s2 <- sum(ar2model$residuals^2)
n1 <- 2
n2 <- 6

pf( (s1-s2)/(n2-n1) / (s2/(length(ar2model$residuals)-n2)), df1 = n2 - n1, df2 =
(length(ar2model$residuals)-n2), lower.tail = FALSE)
#prediction
p<-predict(ar2model, n.ahead = 9, level=95)


#95%limit
upper<-p$pred + (p$se *1.96)
lower<-p$pred - (p$se *1.96)

plot(x, type="o", xlab="Time", ylab="Nm. of passangers", main="Forecast")
plot(x, type="l", col="red", pch=16,xlab="Time", ylab="Nm. of passangers",ylim=c(30000, 70000),
main="Forecast")
plot(x, type="o", col="red", pch=16,xlab="Predictions", ylab="Nm. of passangers",ylim=c(30000,
70000))points(p$pred, type="o", col="black",pch=16)
points(p$pred[1:9], type="o", col="blue", lty=2,ylim=c(30000, 70000),xlim=c(1995,2003))
points(p$pred[1:9], type="l", col="blue", lty=2)


legend("bottomleft", c("Observations", "Predictions", "Prediction bounds"), cex=0.8,
col=c("black","red","blue"), pch=c(16,16,-1), lty=c(0,0,2))


plot(1:9,p$pred, type="o", col="red", pch=16,ylim=c(30000, 70000),xlab="Predictions", ylab="Nm. of
passangers",)
lines(upper[1:9], type="l", col="blue", lty=2)
lines(lower[1:9], type="c", col="blue", lty=2)
points(x2, type="o", col="black",pch=16)
legend("bottomleft", c("Observations", "Predictions", "Prediction bounds"), cex=0.8,
col=c("black","red","blue"), pch=c(16,16,-1), lty=c(0,0,2))
```

Ran Wang
S111503
Exercise 03