STAT 560: Time Series Analysis Homework 7 (Chapter 5)

Ajoy Kumar Saha, ID 101011922 and Md Mominul Islam, ID: 101009250

Date: 11/23/2020

Exercise 5.12

Table B.3 contains data on chemical process viscosity.

a. Fit an ARIMA model to this time series, excluding the last 20 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 20 observations.

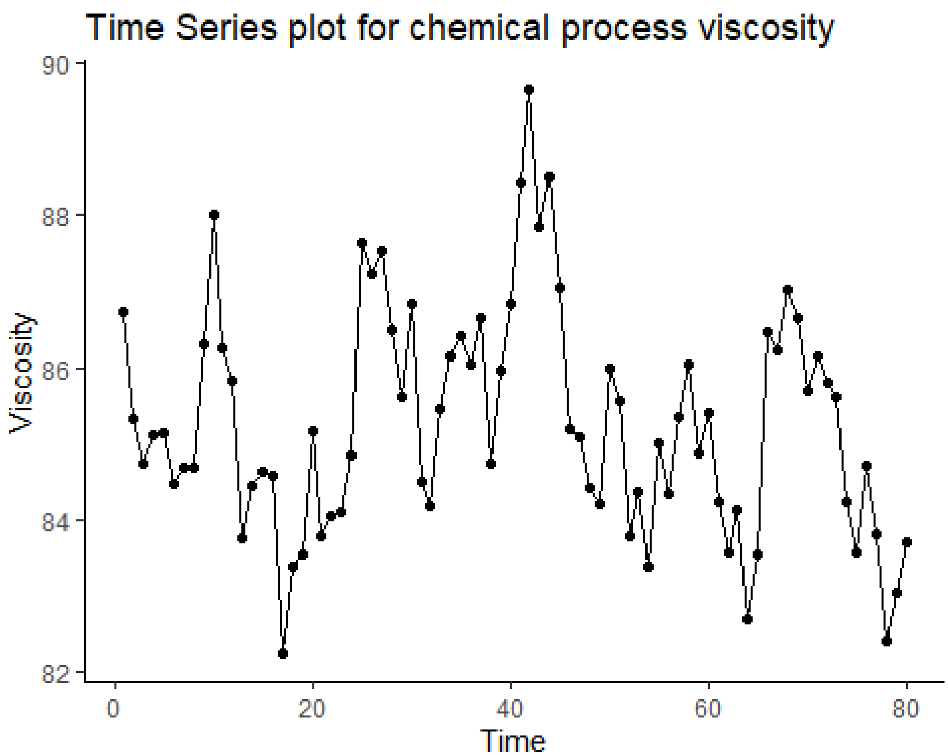
c. Show how to obtain prediction intervals for the forecasts in part b above.

a. Fit an ARIMA model to this time series, excluding the last 20 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

Answer:

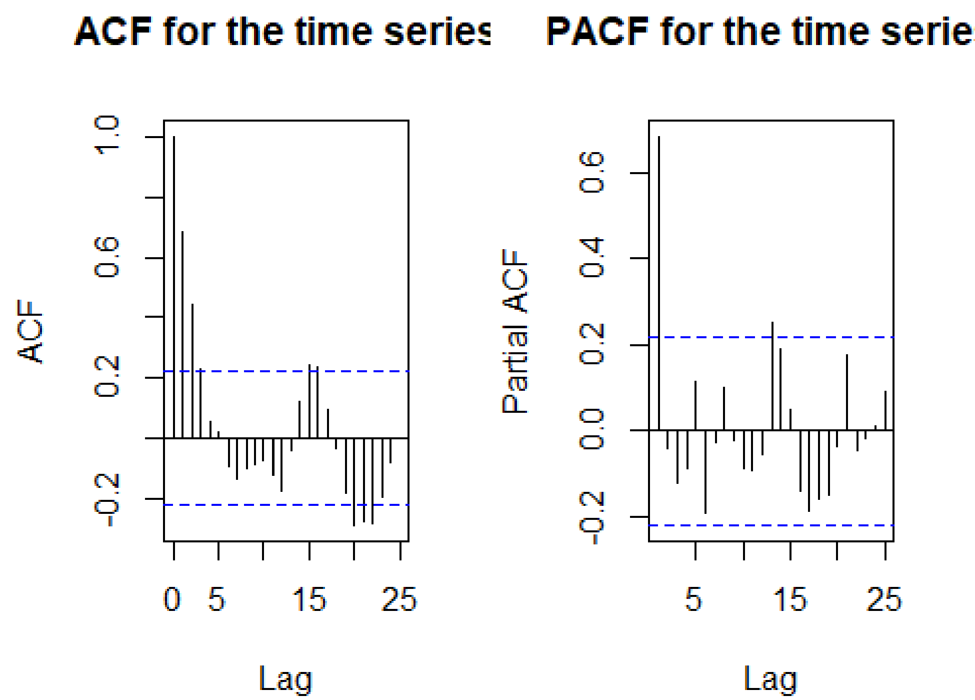
The time series plot for the chemical process viscosity are shown in following figure which excludes last 20 observations. From time serires plot, it seems that this is a stationary time series howerver there might have a little bit downward trend over time.

```
ggplot(data=my.data1)+
  geom_line (aes(x=Time, y = Viscosity))+
  geom_point (aes(x=Time, y = Viscosity))+
  theme_classic()+
labs(x = "Time",
    y = "Viscosity",
    title = "Time Series plot for chemical process viscosity")
```



To find the appropriate ARIMA model for this time series, at first sample ACF and PACF plot has been generated from the given time series. After examine sample ACF and PACF, it seems that the best model is AR(1) model for this time series. Here, the sample ACF shows the exponential decay of ACF values and there are some ACF values are lies outside the boundary limits (for lag 15,16,20 etc). Whereas PACF shows the decent pattern as it gives cut off at 1. However in this plot there are a bit confusion for the lag 13 which produces bit high PACF values and lies outside the boundary. Therefore, we decided that AR(1) model for this time series.

```
par(mfrow=c(1,2))
my.data1.acf <- acf(my.data1[,2],lag.max=25,type="correlation",main="ACF for
the time series")
my.data1.pacf <- acf(my.data1[,2],lag.max=25,type="partial",main="PACF for
the time series")
```



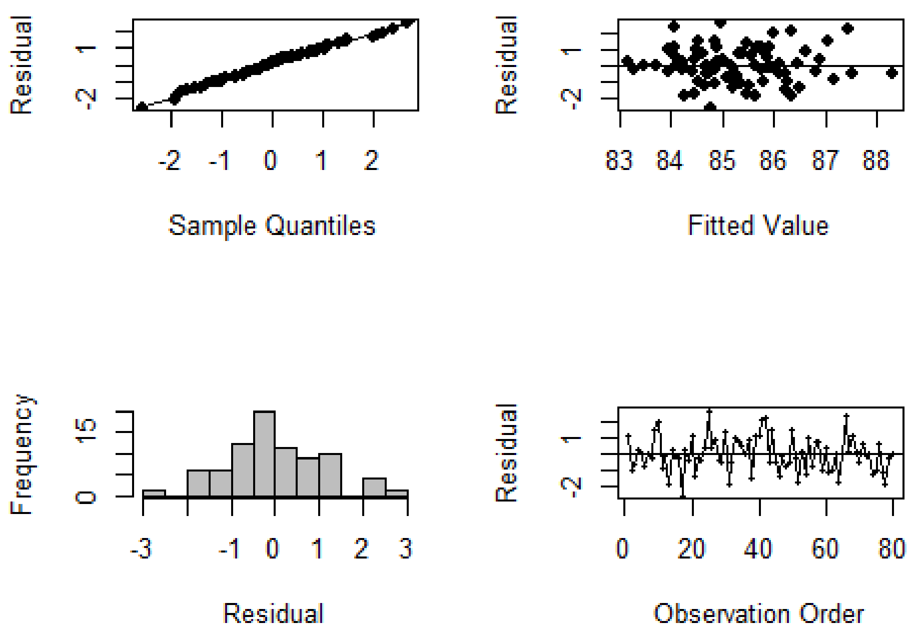

```
# AR(1) model create with code ARIMA (1,0,0)
viscosity.model.ar1<-arima(my.data1[,2],order=c(1, 0, 0))
viscosity.model.ar1

##
## Call:
## arima(x = my.data1[, 2], order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.6934    85.2721
## s.e.  0.0802     0.3756
##
## sigma^2 estimated as 1.121:  log likelihood = -118.42,  aic = 242.84
```

The coefficients of this AR(1) models are μ= 85.2721, ϕ = 0.6934. So that δ = μ(1- ϕ) = 26.1443 Therefore, the fit model (general form of AR(1) is y_t = δ + ϕ y_{t-1} + ε_t) of this time series is

Model Adequacy:

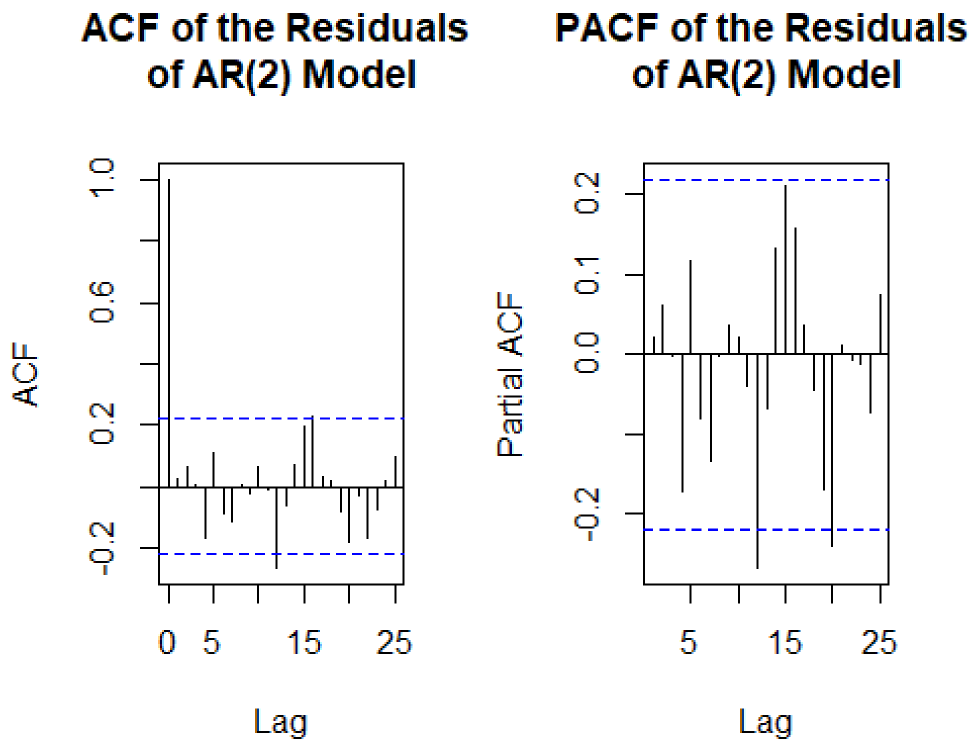
Following four figures illustrates outcome of the residual analysis. From visual inspection of residual plot, it seems that model is quite adequate. From the normal probability plot (at the top left), we see that points are closely following a straight line. From histogram (at bottom left) it is cleared that data distribution looks like some what a bell shaped (single peak). So that it can be said that residuals are normally distributed. Residual vs Fitted used to detect non-linearity between the variables and from the figure at the top right corner, we can say that residuals are randomly distributed which is non-linear. From the residuals vs observations order, it is observed that residuals are randomly varying and they are not following any patterns.

```
res.viscosity.ar1<-as.vector(residuals(viscosity.model.ar1))
#to obtain the fitted values we use the function fitted() from the forecast
package
library(forecast)

## Warning: package 'forecast' was built under R version 3.6.3

fit.viscosity.ar1<-as.vector(fitted(viscosity.model.ar1))

#4-in-1 plot of the residuals
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.viscosity.ar1,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(res.viscosity.ar1,datax=TRUE)
plot(fit.viscosity.ar1,res.viscosity.ar1,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(res.viscosity.ar1,col="gray",xlab='Residual',main='')
plot(res.viscosity.ar1,type="l",xlab='Observation Order', ylab='Residual')
points(res.viscosity.ar1,pch=16,cex=.5)
abline(h=0)
```



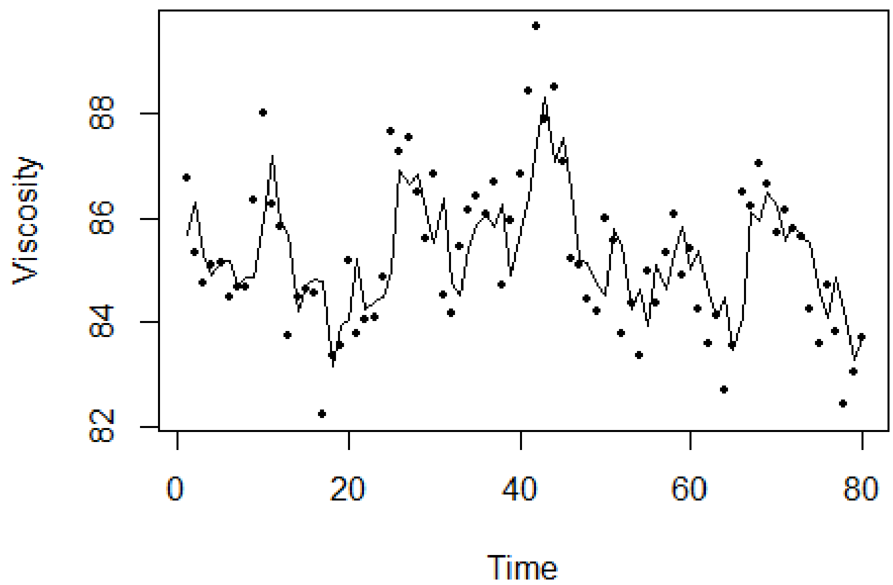
The sample ACF and PACF plots of the residuals also gives good information that there is no autocorrelation in the residuals.

```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(res.viscosity.ar1,lag.max=25,type="correlation",main="ACF of the
Residuals \nof AR(2) Model")
acf(res.viscosity.ar1, lag.max=25,type="partial",main="PACF of the Residuals
\nof AR(2) Model")
```



Following plot shows the model’s fit line over observation point, it seems that model closely flow the trend, although, it has some limitaiton to capture the high and low values. However, it can be acceptable model.

```
# Plot fitted value
plot(my.data1[,2],type="p",pch=16,cex=.5,xlab='Time',ylab='Viscosity')
lines(fit.viscosity.ar1)
legend(95,88,c("y(t)","yhat(t)"), pch=c(16, NA),lwd=c(NA,.5), cex=.55)
```



Forecast model

For the AR(1) model the product of the required polynomials is $(\psi 0 + \psi 1\ B + \psi 2\ B^2 + \cdots)(1 - \phi 1\ B) = 1$

Equating like power of B, we find that

B^0: $\psi 0 = 1$

B^1: $\psi 1 - \phi 1\ \psi 0 = 0$, or $\psi 1 = \phi 1 \times 1 = \phi 1$

B^2: $\psi 2 - \phi 1\ \psi 1 = 0$, or $\psi 2 = \phi 1 \psi 1 = \phi 1 \times \phi 1 = \phi 1^2$

B^3: $\psi 3 - \phi 1\ \psi 2 = 0$, or $\psi 3 = \phi 1 \psi 2 = \phi 1 \times \phi 1^2 = \phi 1^3$

B^4: $\psi 4 - \phi 1\ \psi 3 = 0$, or $\psi 4 = \phi 1 \psi 3 = \phi 1 \times \phi 1^3 = \phi 1^4$

B^5: $\psi 5 - \phi 1\ \psi 4 = 0$, or $\psi 5 = \phi 1 \psi 4 = \phi 1 \times \phi 1^4 = \phi 1^5$

In general, we can show for the AR(1) model that $\psi j = \phi 1^j$

Here, the value of coefficients of forecast model are

$\psi 0 = 1$

$\psi 1 = \phi 1 = 0.6934$

$\psi 2 = \phi 1^2 = (0.6934)^2 = 0.4808$

$\psi3 = \phi1\text{^}3$ = (0.6934)^3 = 0.3334

$\psi4 = \phi1\text{^}4$ = (0.6934)^4 = 0.2312

$\psi5 = \phi1\text{^}5$ = (0.6934)^5 = 0.1603

………

Now we know the general forecast equation (Book page 379, eq 5.88) is

$$\hat{y}_{\{t+\tau\}} = \mu + \sum_{\{j=\tau\}}^{\{infty\}} \psi_j \, y_{\{t+\tau-j\}}$$

For our equation, μ= 85.2721,

$\psi0$ = 1

$\psi1$ = 0.6934,

$\psi2$= 0.4808,

$\psi3$ = 0.3334,

$\psi4$ = 0.2312,

$\psi5$ = 0.1603

……

So the forecast equation of τ step ahead is

$$\hat{y}_{\{t+\tau\}} = 85.2721 \; + \sum_{\{j=\tau\}}^{\{infty\}} \psi_j \, y_{\{t+\tau-j\}}$$

b. Forecast the last 20 observations.

Here, following code for the automatic arima model.

```
# Auto arima model generation for double check.
auto.arima(my.data1[,2])

## Series: my.data1[, 2]
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1     mean
##       0.6934  85.2721
## s.e.  0.0802   0.3756
##
```

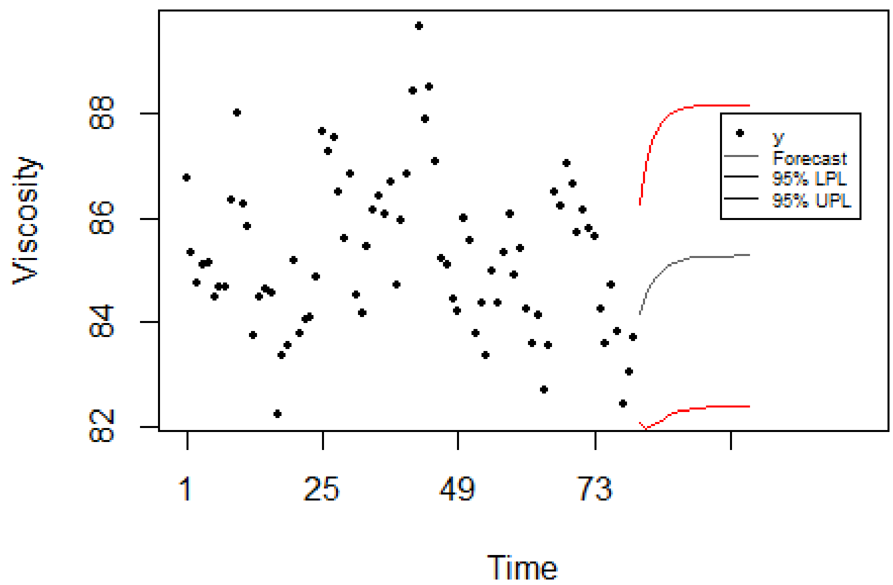
```
## sigma^2 estimated as 1.15:  log likelihood=-118.42
## AIC=242.84   AICc=243.16   BIC=249.99
```

Using auto.arima() function, We get that AR(1) is the most appropriate model. This AR(1) model is used to forcast last 20 observations. The calculated last 20 observations are as below:

```
# to obtain the 1- to 12-step ahead forecasts, we use the function forecast()
from the forecast package
library(forecast)
viscosity.ar1.forecast<-as.array(forecast(viscosity.model.ar1,h=20))
viscosity.ar1.forecast
```

```
##     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
##  81       84.18154 82.82452 85.53856 82.10616 86.25693
##  82       84.51592 82.86460 86.16724 81.99045 87.04139
##  83       84.74777 82.97224 86.52330 82.03233 87.46321
##  84       84.90853 83.07628 86.74078 82.10634 87.71072
##  85       85.02000 83.16109 86.87890 82.17704 87.86295
##  86       85.09729 83.22570 86.96887 82.23494 87.95963
##  87       85.15088 83.27322 87.02853 82.27926 88.02250
##  88       85.18803 83.30747 87.06860 82.31196 88.06410
##  89       85.21380 83.33184 87.09576 82.33559 88.09201
##  90       85.23166 83.34903 87.11429 82.35243 88.11090
##  91       85.24405 83.36110 87.12700 82.36432 88.12378
##  92       85.25264 83.36953 87.13575 82.37268 88.13260
##  93       85.25860 83.37541 87.14178 82.37852 88.13867
##  94       85.26272 83.37951 87.14594 82.38259 88.14286
##  95       85.26559 83.38235 87.14882 82.38543 88.14575
##  96       85.26757 83.38433 87.15082 82.38740 88.14775
##  97       85.26895 83.38570 87.15220 82.38877 88.14913
##  98       85.26990 83.38665 87.15315 82.38972 88.15009
##  99       85.27057 83.38732 87.15382 82.39038 88.15075
## 100       85.27102 83.38777 87.15428 82.39084 88.15121
```

```
plot(my.data1[,2],type="p",pch=16,cex=.5,xlab='Time',ylab='Viscosity',xaxt='n
',xlim=c(1,120))
axis(1, seq(1,120,24), my.data1[seq(1,120,24),1])
lines(81:100,viscosity.ar1.forecast$mean,col="grey40")
lines(81:100,viscosity.ar1.forecast$lower[,2], col="red")
lines(81:100,viscosity.ar1.forecast$upper[,2], col="red")
legend(95,88,c("y","Forecast","95% LPL","95% UPL"), pch=c(16, NA, NA,NA),
lwd=c(NA,.5,.5,.5),cex=.55,col=c("black","grey40","black","black"))
```



c. Show how to obtain prediction intervals for the forecasts in part b above.

The prediction interval formula for the forecsts value for our AR(1) model is

$$\hat{y}_{t+\tau}(T) + / - Z_{\alpha/2} * \sqrt{Var[e_T(\tau)]}$$

$$\hat{y}_{t+\tau}(T) + / - Z_{\alpha/2} * \sigma\sqrt{(1-\phi^{2\tau})/(1-\phi^2)]}$$

where, ϕ = 0.693, Z = 1.96 for 95% and Z = 1.28 for 80%

Exercise 5.33

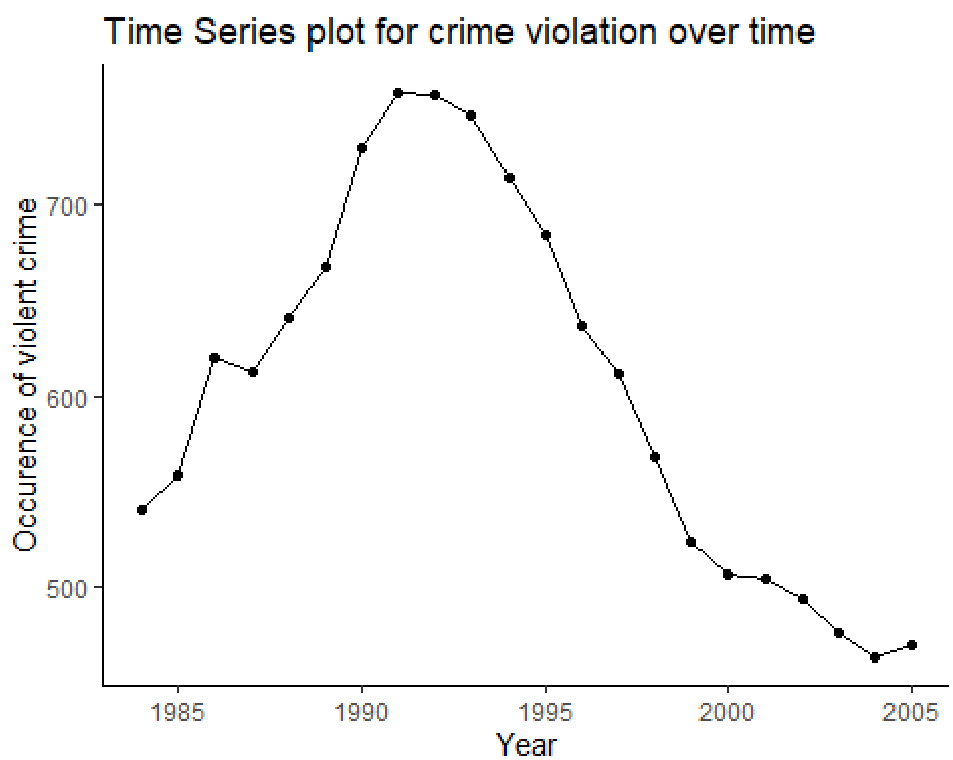
Table B.15 presents data on the occurrence of violent crimes. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed Violent Crime Rate, per 100,000 inhabitants

Develop an appropriate ARIMA model and a procedure for forecasting for these data.

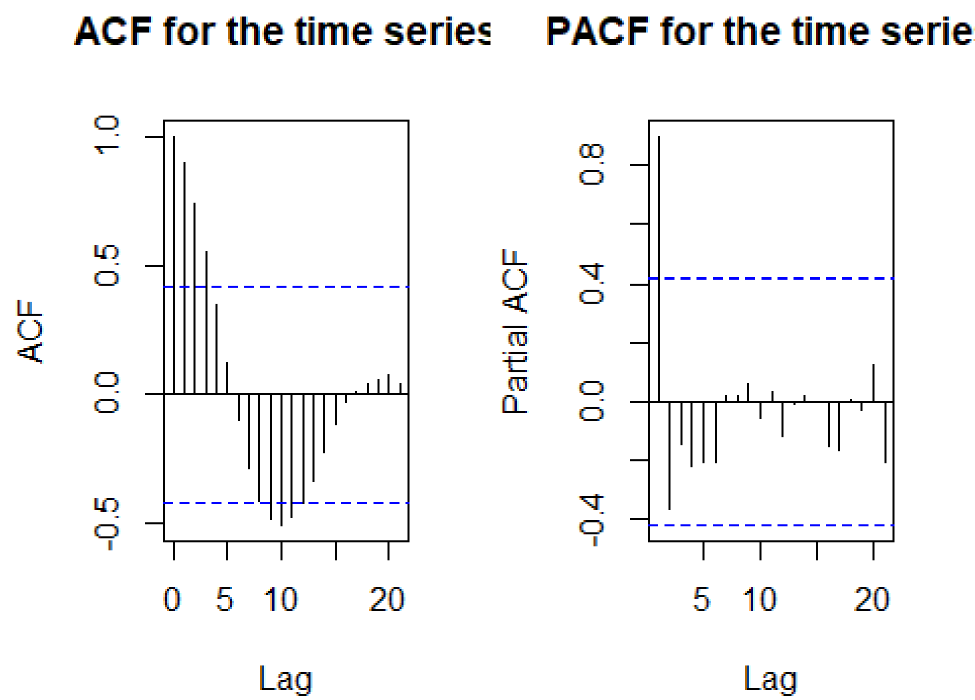
Answer:

The time series plot for the occurrence of violent crimes are shown in following figure. From time serires plot, it seems that this is a nonstationary time series having varing mean and variance.

```
ggplot(data=my.data2)+
  geom_line (aes(x=Year, y = Crime_rate))+
  geom_point (aes(x=Year, y = Crime_rate))+
  theme_classic()+
labs(x = "Year",
    y = "Occurence of violent crime",
    title = "Time Series plot for crime violation over time")
```




```
par(mfrow=c(1,2))
my.data2.acf <- acf(my.data2[,2],lag.max=25,type="correlation",main="ACF for
the time series")
my.data2.pacf <- acf(my.data2[,2],lag.max=25,type="partial",main="PACF for
the time series")
```

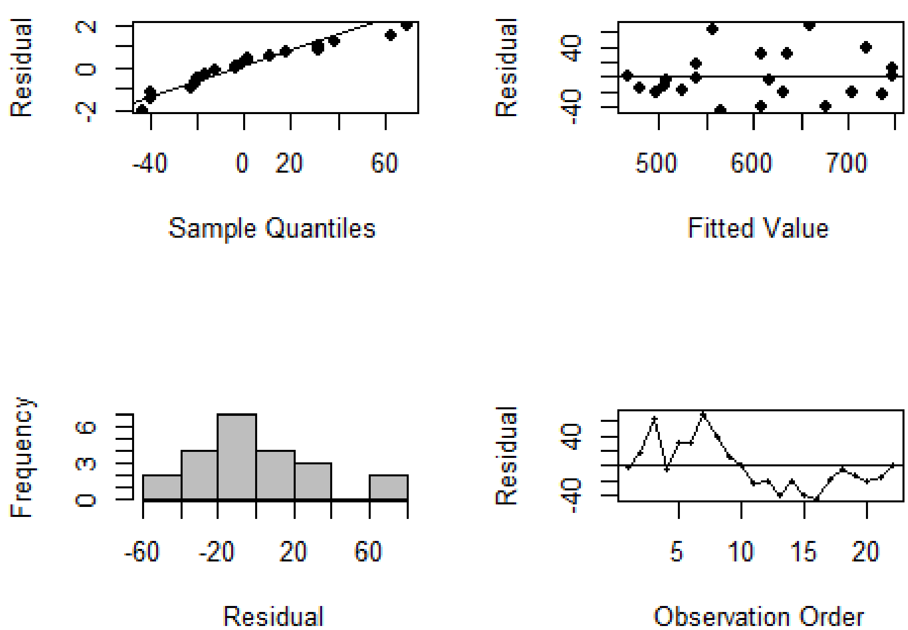


The sample ACF and PSCF shows that values are decreasing slowly with a significant value at lag 1. So it would be good to take differences to get the constant process time series. ACF shows expnential decay and sin pattern. PACF shows good shape for AR(1), because cut off at 1. We can go for AR(1) process model, but it seems that it would be good to try with difference (to find stationaly. Howerver check the residual plot and ACF and PACF of residual.

```
# AR(1) model create with code ARIMA (1,0,0)
crime.model.ar1<-arima(my.data2[,2],order=c(1, 0, 0))
crime.model.ar1

##
## Call:
## arima(x = my.data2[, 2], order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.9451   543.1955
## s.e.  0.0483    76.9095
##
## sigma^2 estimated as 924.1:  log likelihood = -107.45,  aic = 220.9
```

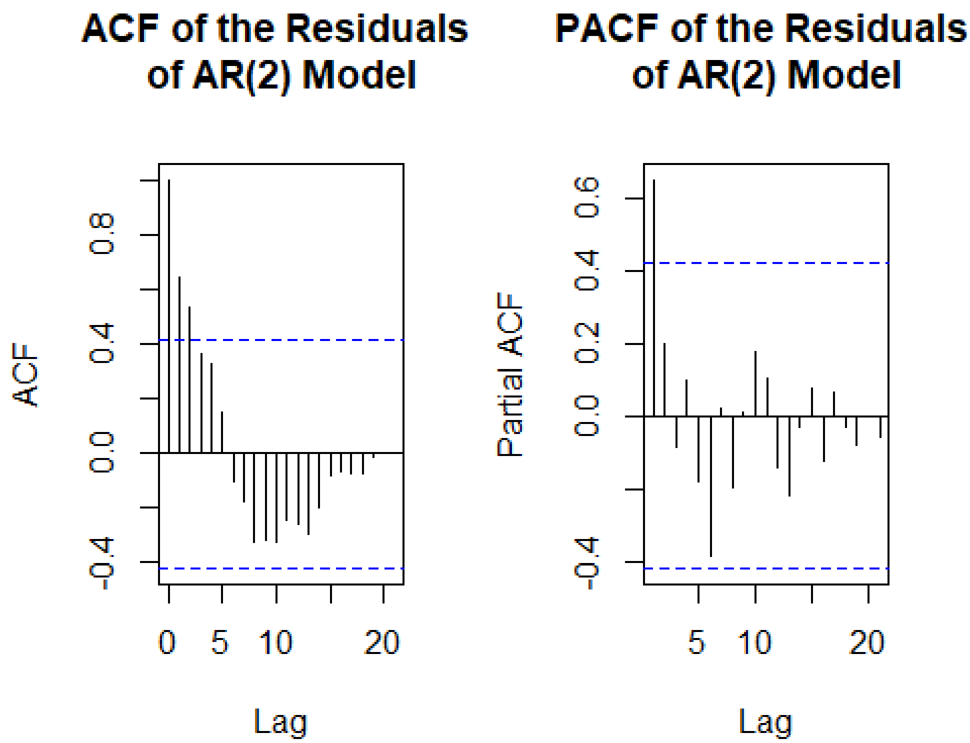
It seems that residual are not good as expected for normality plot expecailly for higher value of residuals. Histrogram is good. Residual vs fit almost good but residual vs observation shows that there is a pattern of residual. They are not randomly distributed.

```r
res.crime.ar1<-as.vector(residuals(crime.model.ar1))
#to obtain the fitted values we use the function fitted() from the forecast package
fit.crime.ar1 <-as.vector(fitted(crime.model.ar1))

#4-in-1 plot of the residuals
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.crime.ar1,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(res.crime.ar1,datax=TRUE)
plot(fit.crime.ar1,res.crime.ar1,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(res.crime.ar1,col="gray",xlab='Residual',main='')
plot(res.crime.ar1,type="l",xlab='Observation Order', ylab='Residual')
points(res.crime.ar1,pch=16,cex=.5)
abline(h=0)
```



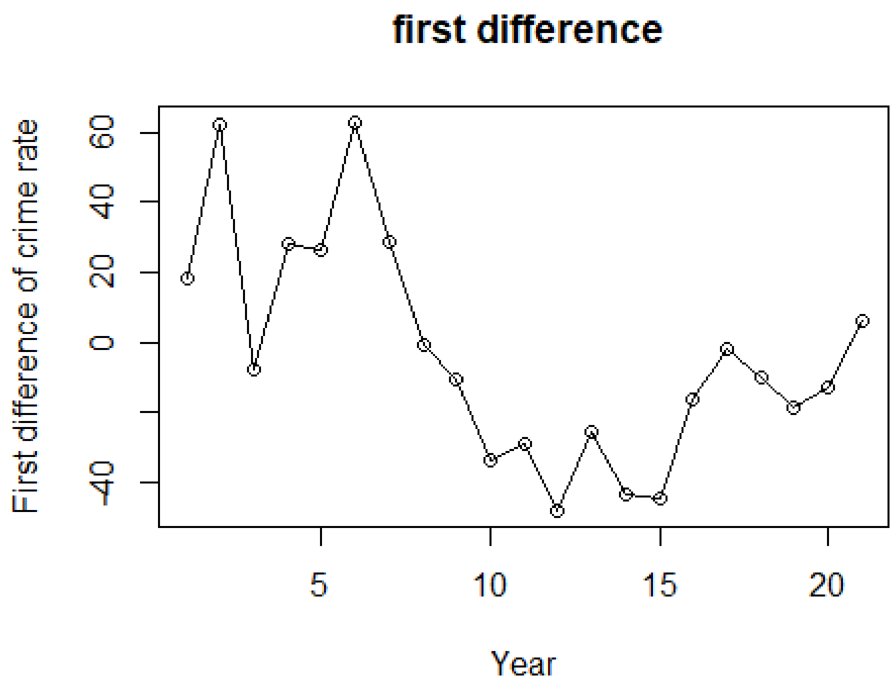
It seems that for AR(1), saple ACF and PACF plots of residuals shows there might have autocorrelation in residuals.

```r
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(res.crime.ar1,lag.max=25,type="correlation",main="ACF of the Residuals
\nof AR(2) Model")
acf(res.crime.ar1, lag.max=25,type="partial",main="PACF of the Residuals \nof
AR(2) Model")
```

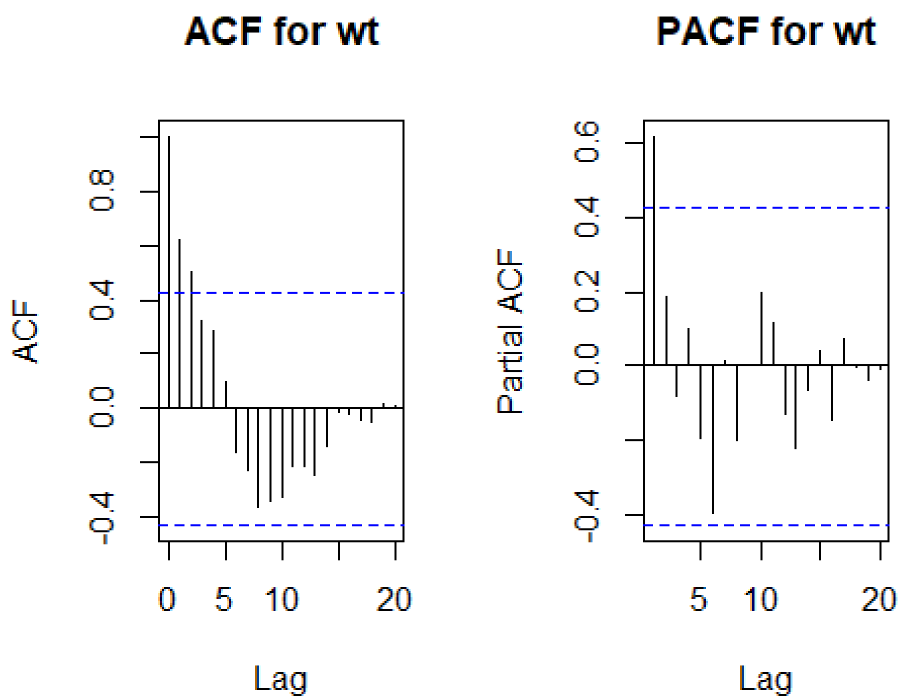


Lets take difference and see if any possible improvement in our model. Following time series plot, still shows nonstationary pattern as the means are changing over time. ACF shows expnential decay and sin pattern. Again, PACF shows good shape for AR(1), because cut off at 1.

```
my.data2.wt1 <- diff(my.data2[,2])
plot(my.data2.wt1, type = "o", main=" first difference", xlab= "Year",
ylab="First difference of crime rate")
```

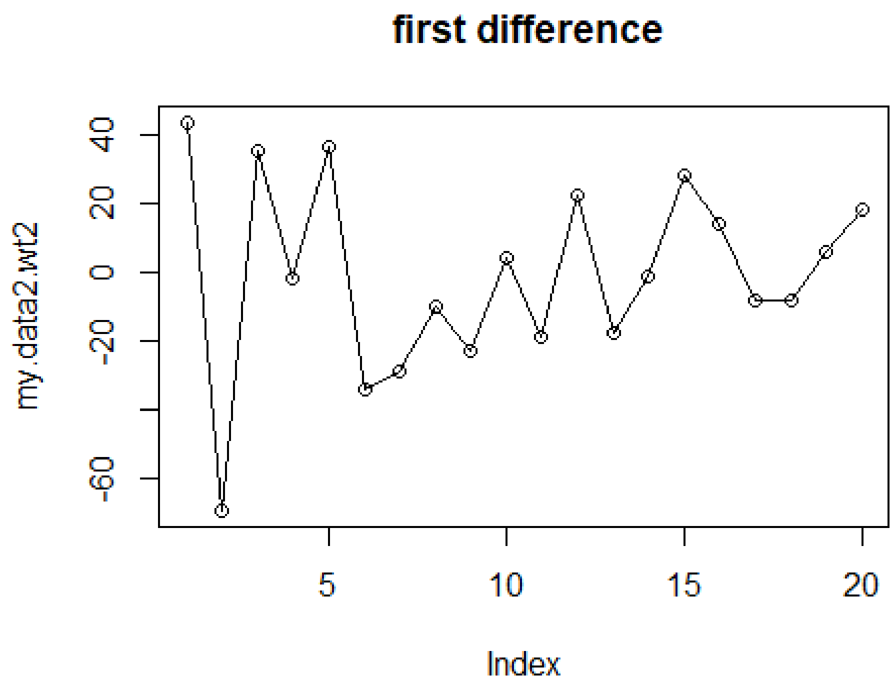



```
par(mfrow=c(1,2))
my.data2.wt1.acf <- acf(my.data2.wt1,lag.max=25,type="correlation",main="ACF for wt")
my.data2.wt1.pacf <- acf(my.data2.wt1,lag.max=25,type="partial",main="PACF for wt")
```



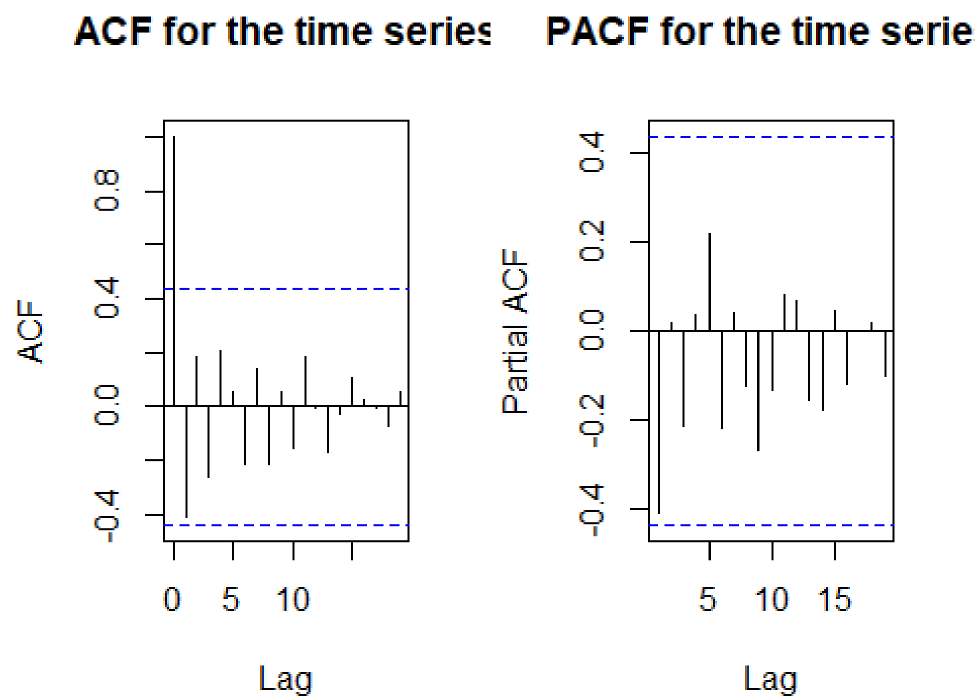
Lets take the 2nd difference and see if any possible improvement in our model to get stationary time series. Following time series plot (below) is for 2nd difference. It seems that it shows almost stationary as the means are not any more changing over time and varince are considerablly constant.

```
my.data2.wt2 <- diff(my.data2.wt1)
plot(my.data2.wt2, type = "o", main=" first difference")
```



ACF shows mixed pattern. Again, PACF shows good shape for AR(1), because cut off at 1 (this time PACF value is negative as we take 2nd difference. Lets fit model AR(1) for 2nd order difference. That means ARIMA (1, 2, 0).

```
par(mfrow=c(1,2))
my.data2.wt2.acf <- acf(my.data2.wt2,lag.max=25,type="correlation",main="ACF
for the time series")
my.data2.wt2.pacf <- acf(my.data2.wt2,lag.max=25,type="partial",main="PACF
for the time series")
```



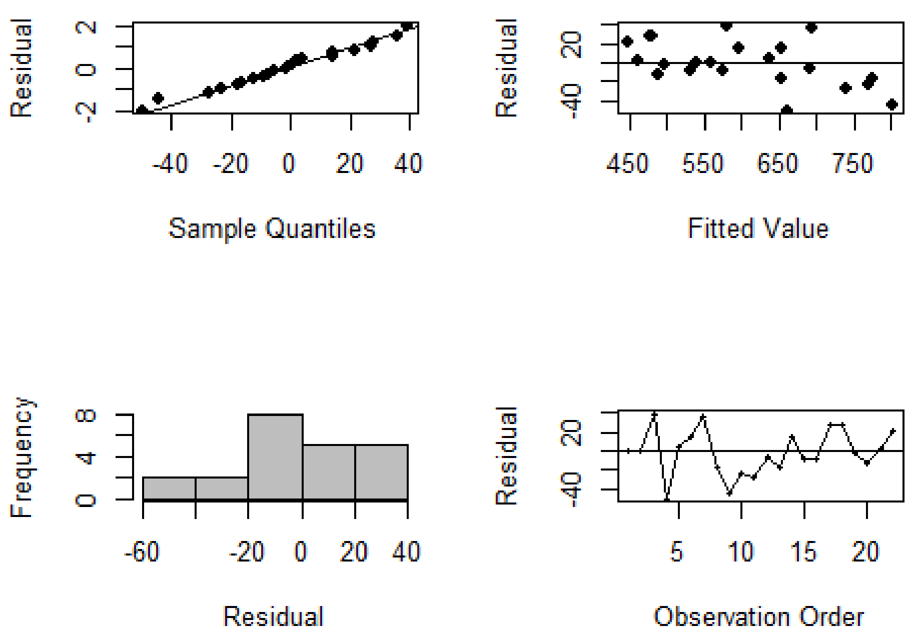

```
# model create with code ARIMA (1,2,0)
crime.model.arima120<-arima(my.data2[,2],order=c(1, 2, 0))
crime.model.arima120

##
## Call:
## arima(x = my.data2[, 2], order = c(1, 2, 0))
##
## Coefficients:
##           ar1
##       -0.4533
## s.e.   0.2091
##
## sigma^2 estimated as 592.3:  log likelihood = -92.33,  aic = 188.67
```

The genral form of our model is

$$(1-\phi B)(1-B)^2 y_t = \delta + \epsilon_t;\ where, \phi = -0.4533$$

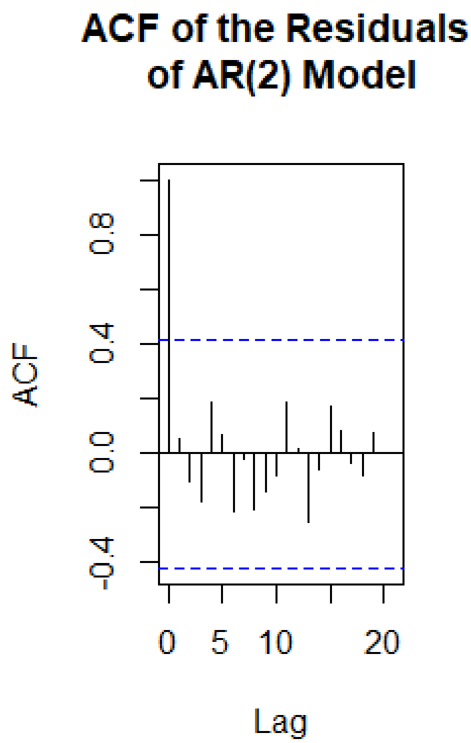
Now check the residual for the ARIMA(1,2,0). Residual 4-1 plot is shown below. It seems that qq plot looks good and closely follow the straight line. Residual vs fit plot also shows random plot of the residual. Histogram plot is not good shape, however it might be acceptable as they have single peak. Residual vs observation order shows good shape compare to previous and shows that there is no unic pattern. Values are varives around the zero line.

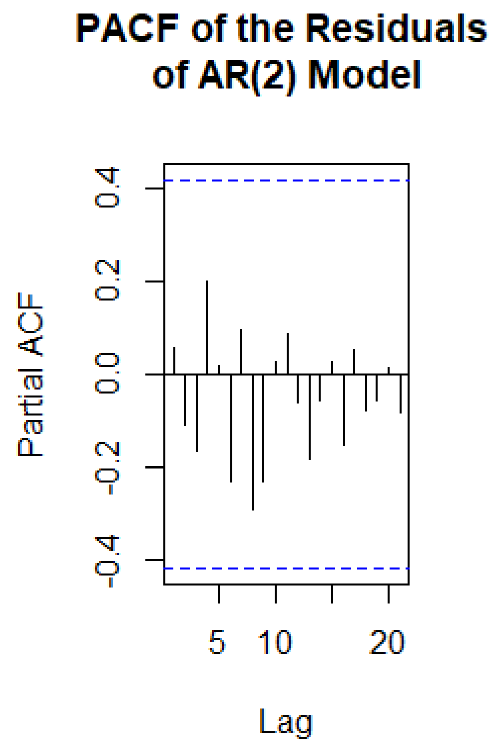
```r
res.crime.arima12<-as.vector(residuals(crime.model.arima120))
#to obtain the fitted values we use the function fitted() from the forecast package
library(forecast)
fit.crime.arima120 <-as.vector(fitted(crime.model.arima120))

#4-in-1 plot of the residuals
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.crime.arima12,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(res.crime.arima12,datax=TRUE)
plot(fit.crime.arima120,res.crime.arima12,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(res.crime.arima12,col="gray",xlab='Residual',main='')
plot(res.crime.arima12,type="l",xlab='Observation Order', ylab='Residual')
points(res.crime.arima12,pch=16,cex=.5)
abline(h=0)
```



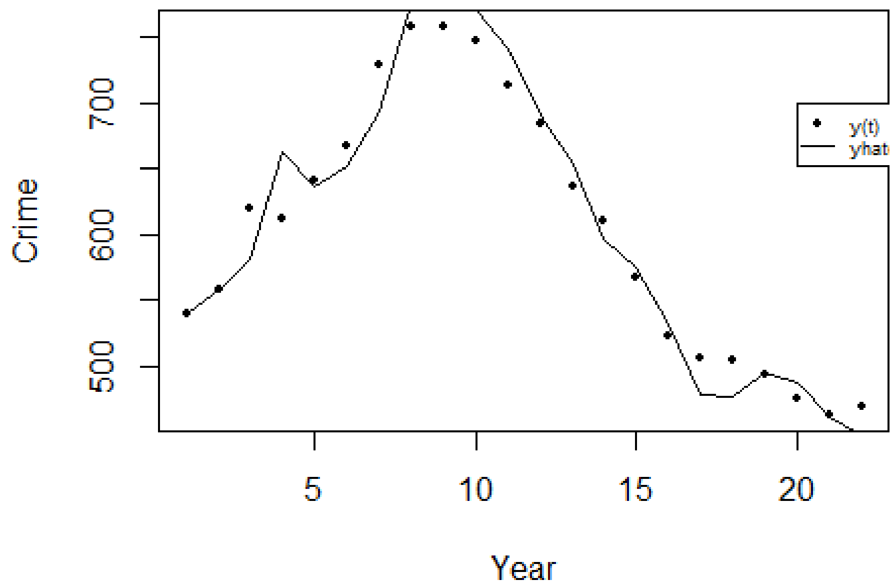
The sample ACF and PACF plots of the residuals also gives good information that there is no autocorrelation in the residuals after taking 2nd order differences.

```r
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(res.crime.arima12,lag.max=25,type="correlation",main="ACF of the
Residuals \nof AR(2) Model")
acf(res.crime.arima12, lag.max=25,type="partial",main="PACF of the Residuals
\nof AR(2) Model")
```






```r
# Plot fitted value
plot(my.data2[,2],type="p",pch=16,cex=.5,xlab='Year',ylab='Crime')
lines(fit.crime.arima120)
legend(20,700,c("y(t)","yhat(t)"), pch=c(16, NA),lwd=c(NA,.5), cex=.55)
```



Formula for the forecast model: To get forecast model, we need to calcualte the psi_i$'s.Thegeneralformulatogetphi_i$’s as bellow

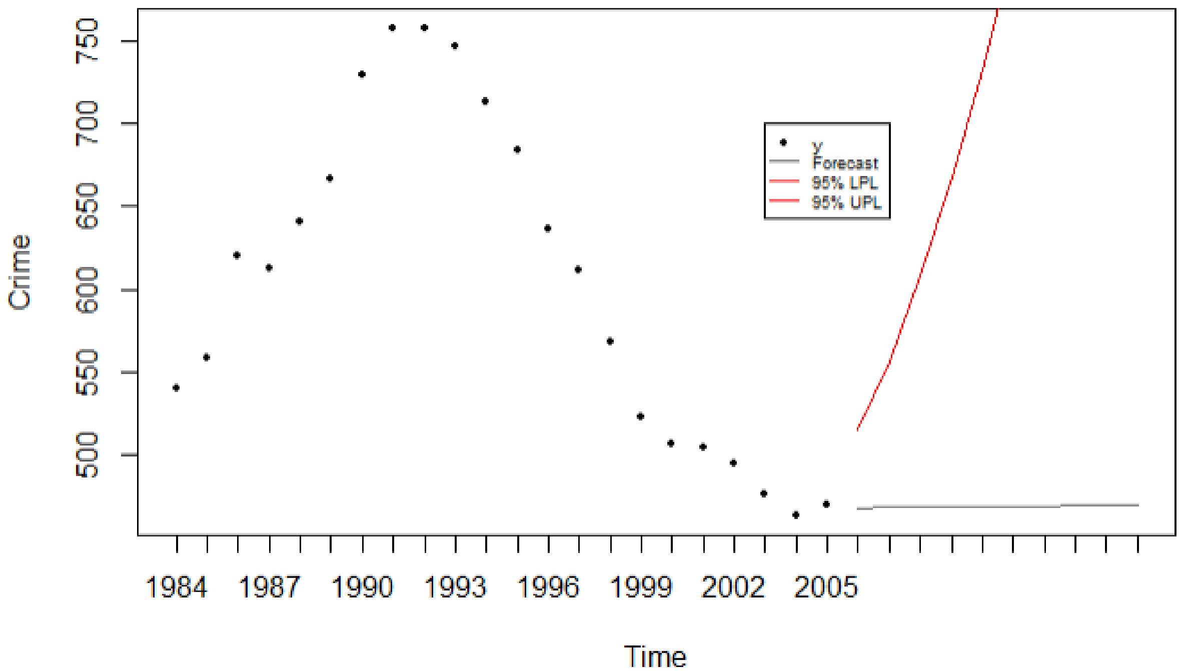
$$(\Psi_0 + \Psi_1 B + \Psi_2 B^2 + \ldots)(1 - \phi_1 B)(1 - B)^2 = 1$$

From this above equation, we will get all the psi_i$’s value by equating the B’s coefficient.

The following table shows the prediction interval for the 1 to 10 step ahead forcast.

```
# to obtain the 1- to 10-step ahead forecasts, we use the function forecast()
from the forecast package

crime.arima120.forecast<-as.array(forecast(crime.model.arima120,h=10))
crime.arima120.forecast

##    Point Forecast     Lo 80    Hi 80      Lo 95     Hi 95
## 23       466.7685 435.57949 497.9575  419.06902  514.4680
## 24       468.1591 410.71483 525.6033  380.30568  556.0124
## 25       467.8171 375.94389 559.6902  327.30918  608.3249
## 26       468.2604 338.06244 598.4584  269.13979  667.3811
## 27       468.3478 295.21144 641.4842  203.55859  733.1370
## 28       468.5965 248.80248 688.3906  132.45054  804.7425
## 29       468.7721 198.69568 738.8486   55.72585  881.8184
## 30       468.9809 145.31996 792.6418  -26.01578  963.9775
## 31       469.1746  88.78690 849.5623 -112.57818 1050.9274
## 32       469.3751  29.29789 909.4524 -203.66493 1142.4152
```

```r
plot(my.data2[,2],type="p",pch=16,cex=.5,xlab='Time',ylab='Crime',xaxt='n',xlim=c(1,32))
axis(1, seq(1,32,1), my.data2[seq(1,32,1),1])
lines(23:32,crime.arima120.forecast$mean,col="grey40")
lines(23:32,crime.arima120.forecast$lower[,2], col="red")
lines(23:32,crime.arima120.forecast$upper[,2], col="red")
legend(20,700,c("y","Forecast","95% LPL","95% UPL"), pch=c(16, NA, NA,NA),
lwd=c(NA,.5,.5,.5),cex=.55,col=c("black","grey40","red","red"))
```



The prediction interval formula for the forecsts value for our ARIMA(1,2,0) model is

$$\hat{y}_{t+\tau}(T) + / - Z_{\alpha/2} * \sqrt{Var[e_T(\tau)]}$$

where,

Z = 1.96 for 95%

Z = 1.28 for 80% and

$$Var[e_T(\tau)] = \sigma^2 \sum_{i=0}^{\tau-1} \psi_i^2$$

The psi_i$s values are need to get the variance, Var[e_T()].