Arima

Ritwik,Shubha,Manas,Arkajyoti,Arghamalya

14/11/2021

###Forecasting

```
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 4.1.1
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.1.1
```

```
library(foreign)
library(TSA)
```

```
## Warning: package 'TSA' was built under R version 4.1.1
```

```
## Registered S3 methods overwritten by 'TSA':
##   method       from
##   fitted.Arima forecast
##   plot.Arima   forecast
```

```
##
## Attaching package: 'TSA'
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
library(astsa)
```

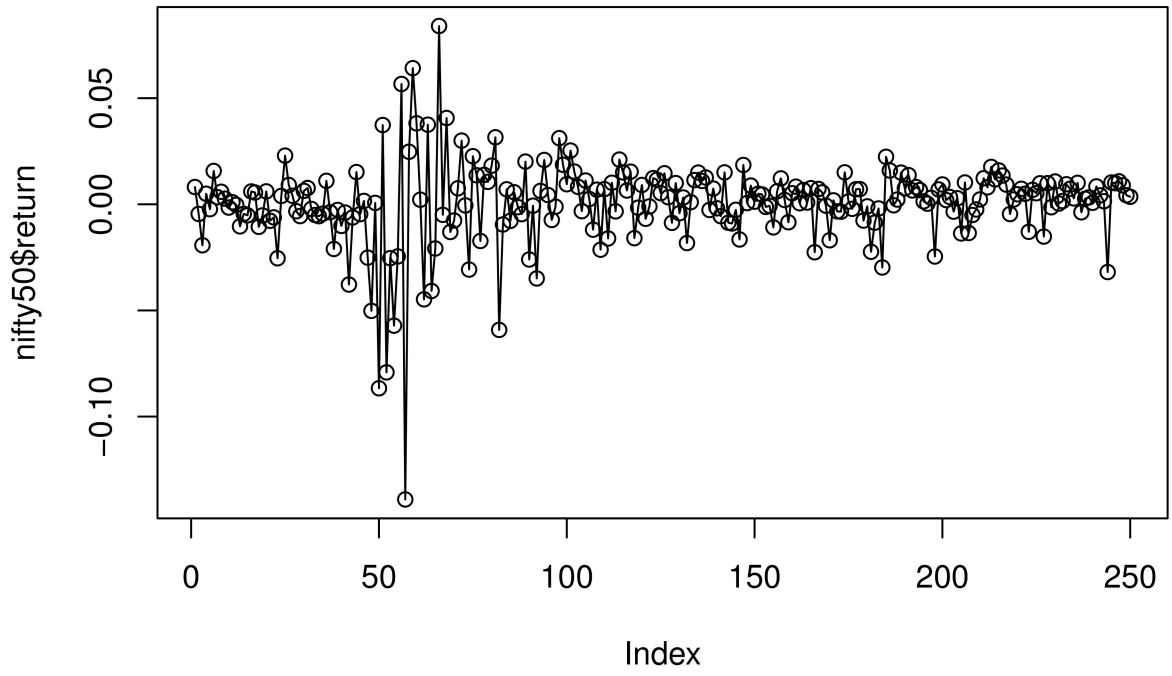
```
## Warning: package 'astsa' was built under R version 4.1.1
```

```
##
## Attaching package: 'astsa'
```

```
## The following object is masked from 'package:forecast':
##
##     gas
```

###For Nifty50

```
price=nifty50$Close
log_return=diff(log(price),lag=1)
nifty50=nifty50[-1,]
nifty50['return']=log_return
plot(nifty50$return,type="o")
```



###Test for Stationarity

```
adf.test(nifty50$return, alternative = "stationary")
```

```
## Warning in adf.test(nifty50$return, alternative = "stationary"): p-value smaller
## than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
```

```
## data:  nifty50$return
## Dickey-Fuller = -4.8436, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

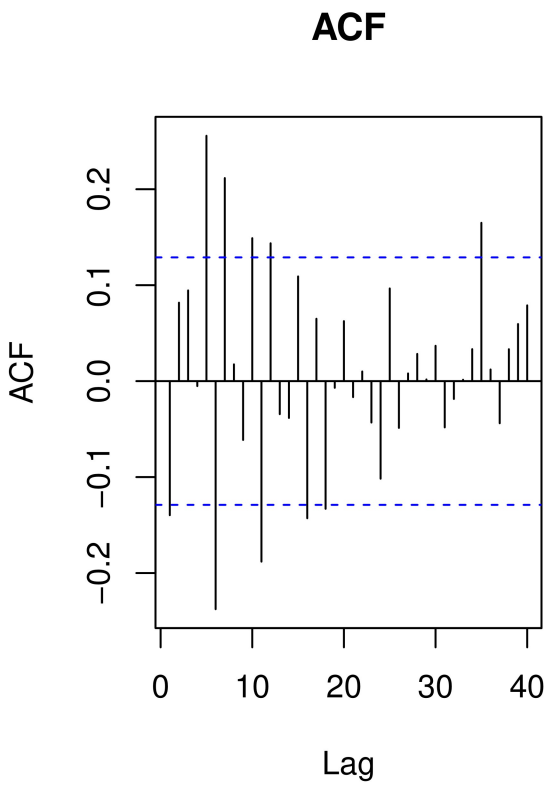
```
kpss=kpss.test(nifty50$return)
kpss
```

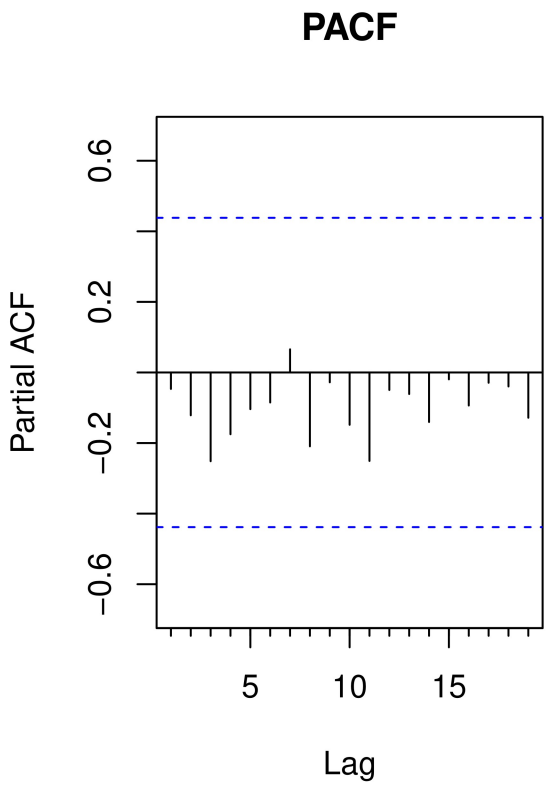
```
##
##  KPSS Test for Level Stationarity
##
## data:  nifty50$return
## KPSS Level = 0.39357, Truncation lag parameter = 5, p-value = 0.07993
```

```
pp=PP.test(nifty50$return)
pp
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  nifty50$return
## Dickey-Fuller = -18.1, Truncation lag parameter = 5, p-value = 0.01
```

```
par(mfrow=c(1,2))
y_train=nifty50[1:231,]
y_test=nifty50[231:250,]
Acf(y_train$return, lag.max = 40, main='ACF')
Pacf(y_test$return, lag.max = 40, main = 'PACF')
```






```
tail(y_test$return)
```

```
## [1] 0.010293171 0.009960404 0.010833605 0.008974644 0.004272496 0.003535794
```

```
aic_model=auto.arima(y_train$return,trace=TRUE,d=0,max.p=10,max.q=10,ic="aic",approximation=FALSE)
```

```
##
##  ARIMA(2,0,2) with non-zero mean : -1136.634
##  ARIMA(0,0,0) with non-zero mean : -1133.306
##  ARIMA(1,0,0) with non-zero mean : -1135.844
##  ARIMA(0,0,1) with non-zero mean : -1135.181
##  ARIMA(0,0,0) with zero mean     : -1135.251
##  ARIMA(1,0,2) with non-zero mean : -1138.487
##  ARIMA(0,0,2) with non-zero mean : -1135.586
##  ARIMA(1,0,1) with non-zero mean : -1134.183
##  ARIMA(1,0,3) with non-zero mean : -1141.652
##  ARIMA(0,0,3) with non-zero mean : -1136.117
##  ARIMA(2,0,3) with non-zero mean : -1142.896
##  ARIMA(3,0,3) with non-zero mean : -1151.297
##  ARIMA(3,0,2) with non-zero mean : -1144.083
##  ARIMA(4,0,3) with non-zero mean : Inf
##  ARIMA(3,0,4) with non-zero mean : -1149.346
##  ARIMA(2,0,4) with non-zero mean : -1141.211
##  ARIMA(4,0,2) with non-zero mean : -1142.203
```

```
##  ARIMA(4,0,4) with non-zero mean : Inf
##  ARIMA(3,0,3) with zero mean     : -1153.257
##  ARIMA(2,0,3) with zero mean     : -1144.865
##  ARIMA(3,0,2) with zero mean     : -1146.052
##  ARIMA(4,0,3) with zero mean     : Inf
##  ARIMA(3,0,4) with zero mean     : -1151.301
##  ARIMA(2,0,2) with zero mean     : -1138.6
##  ARIMA(2,0,4) with zero mean     : -1143.18
##  ARIMA(4,0,2) with zero mean     : -1144.171
##  ARIMA(4,0,4) with zero mean     : Inf
##
##  Best model: ARIMA(3,0,3) with zero mean
```

```
aic_model
```

```
## Series: y_train$return
## ARIMA(3,0,3) with zero mean
##
## Coefficients:
##           ar1     ar2     ar3     ma1      ma2      ma3
##       -0.6529  0.4189  0.8167  0.5909  -0.2907  -0.7253
## s.e.   0.1662  0.1706  0.1047  0.1452   0.2060   0.2069
##
## sigma^2 estimated as 0.0003827:  log likelihood=583.63
## AIC=-1153.26   AICc=-1152.75   BIC=-1129.16
```

```
bic_model=auto.arima(y_train$return,trace=TRUE,d=0,max.p=10,max.q=10,ic="bic",approximation=FALSE)
```

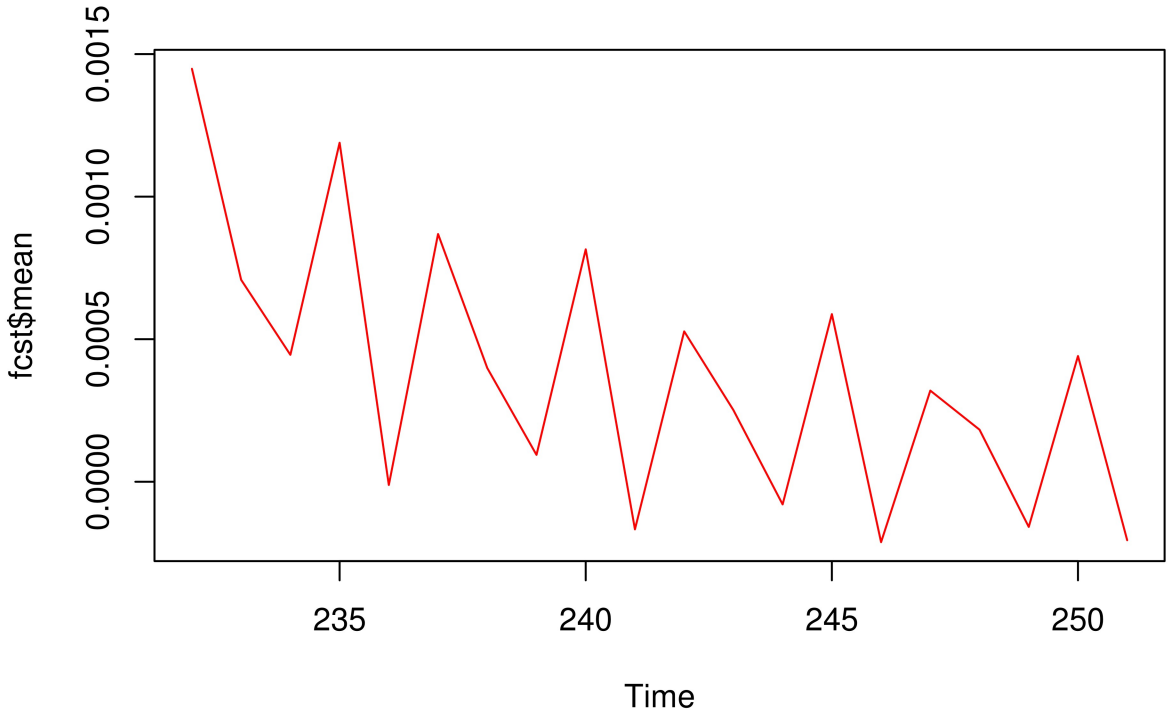
```
##
##  ARIMA(2,0,2) with non-zero mean : -1115.979
##  ARIMA(0,0,0) with non-zero mean : -1126.421
##  ARIMA(1,0,0) with non-zero mean : -1125.517
##  ARIMA(0,0,1) with non-zero mean : -1124.854
##  ARIMA(0,0,0) with zero mean     : -1131.809
##  ARIMA(1,0,1) with non-zero mean : -1120.413
##
##  Best model: ARIMA(0,0,0) with zero mean
```

```
bic_model
```

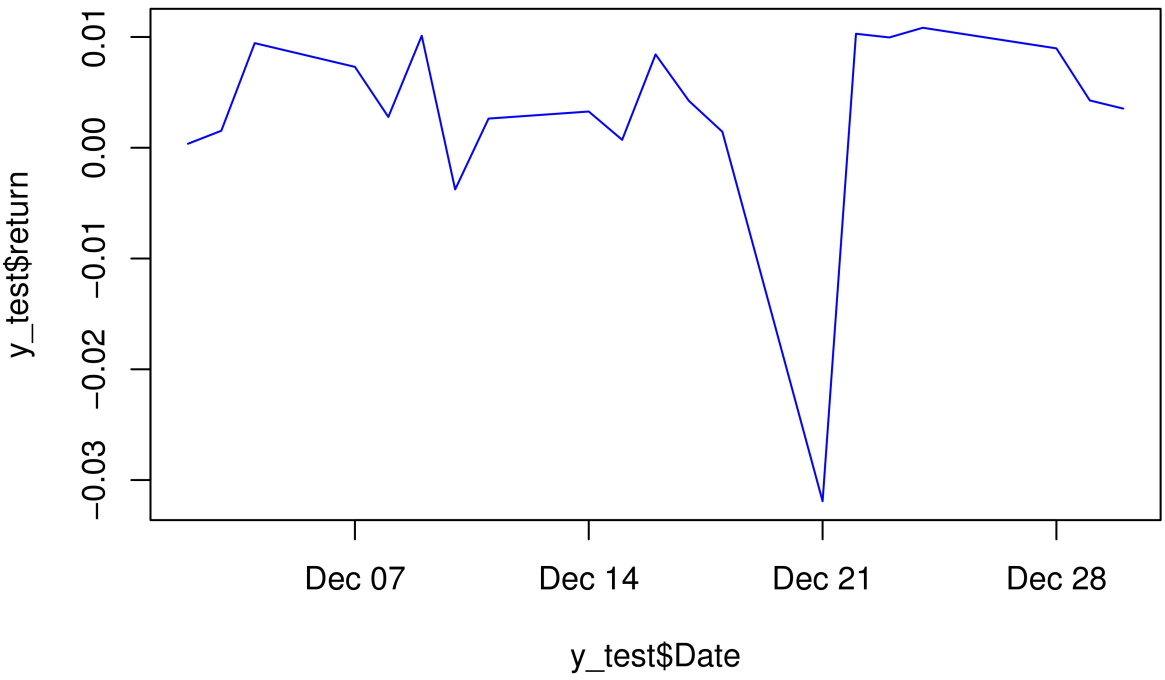
```
## Series: y_train$return
## ARIMA(0,0,0) with zero mean
##
## sigma^2 estimated as 0.000426:  log likelihood=568.63
## AIC=-1135.25   AICc=-1135.23   BIC=-1131.81
```

##Forecasting using AIC_model

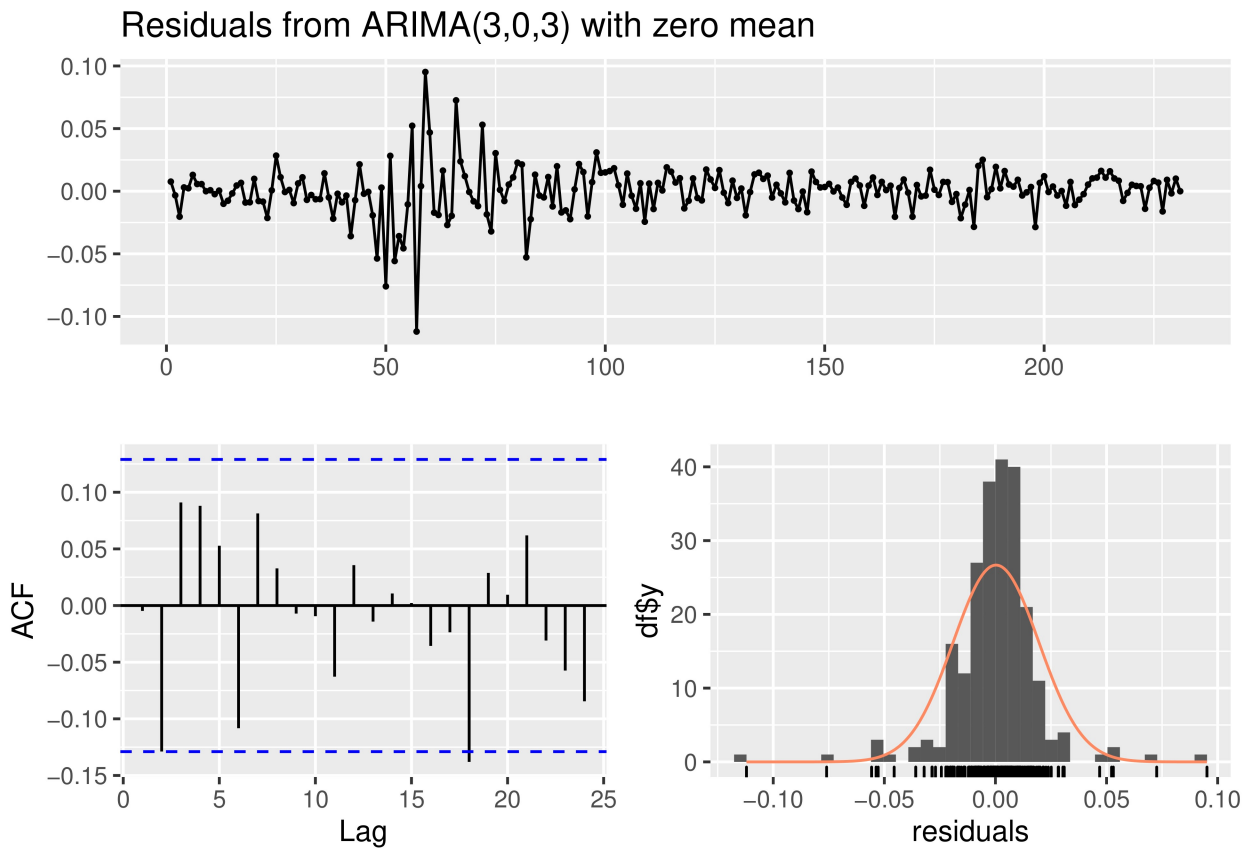
```
fcst=forecast(aic_model,h=20)
plot(fcst$mean,col='red')
```




```
plot(y_test$Date,y_test$return,type='l',col='blue')
```




```
checkresiduals(aic_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,0,3) with zero mean
## Q* = 13.049, df = 4, p-value = 0.01104
##
## Model df: 6.   Total lags used: 10
```

###Calculating MAPE

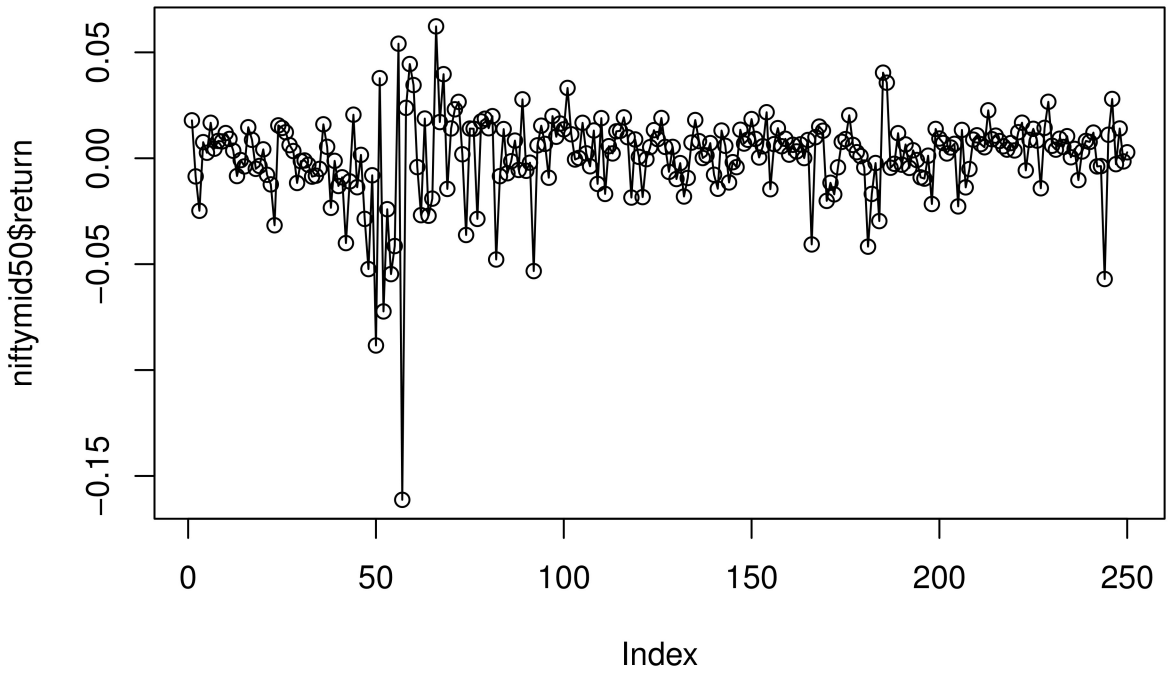
```
p1 = fcst$mean
obs1 = y_test$return
MSE_foreast1 = mean(abs(p1 - obs1))
MSE_foreast1
```

```
## [1] 0.006625704
```

####ARIMA for Nifty midcap50

```
rm(list=ls())
niftymid50=read.csv('nifty midcap50.csv')
niftymid50$Date <- as.Date(niftymid50$Date)
niftymid50$Close=as.numeric(as.character(niftymid50$Close))
price1=niftymid50$Close
```

```
log_return1=diff(log(price1),lag=1)
niftymid50=niftymid50[-1,]
niftymid50['return']=log_return1
plot(niftymid50$return,type="o")
```

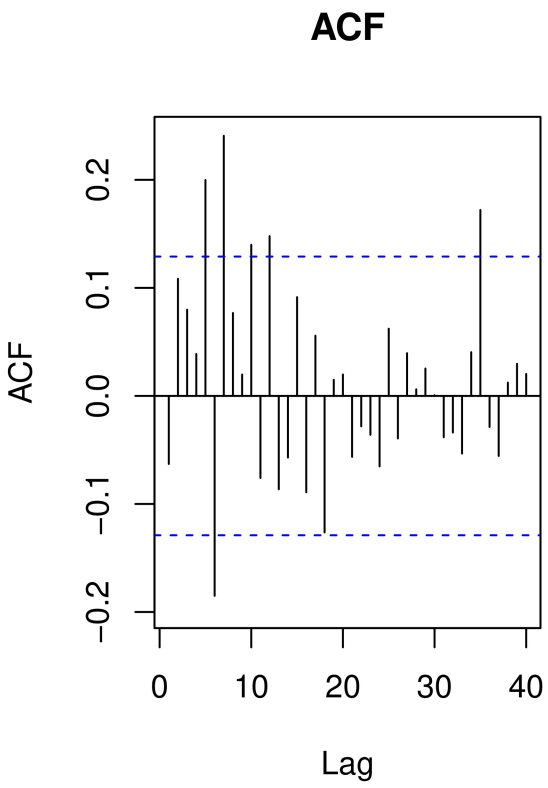


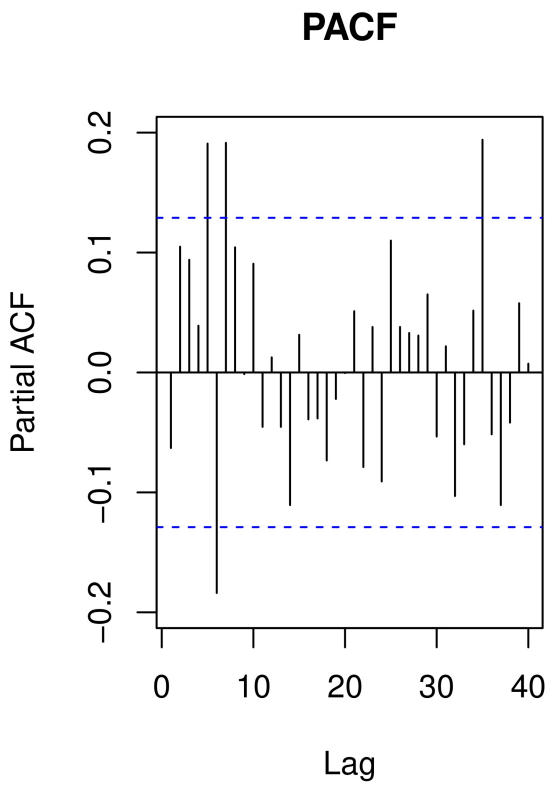

```
adf.test(niftymid50$return, alternative = "stationary")
```

```
## Warning in adf.test(niftymid50$return, alternative = "stationary"): p-value
## smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  niftymid50$return
## Dickey-Fuller = -4.5885, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

```
par(mfrow=c(1,2))
y_trainmid=niftymid50[1:231,]
y_testmid=niftymid50[231:250,]
Acf(y_trainmid$return, lag.max = 40, main='ACF')
Pacf(y_trainmid$return, lag.max = 40, main = 'PACF')
```






```
best.order = c(0, 0, 0)
best.aic = Inf
for (q in 1:10) for (p in 1:10) {
  fit.model= arima(y_trainmid$return, order = c(p, 0, q), optim.control = list(maxit = 2000) ,method='M
  fit.aic = fit.model$aic
  if (fit.aic < best.aic) {
    best.order = c(p, 0, q)
    best.arima = fit.model
    best.aic = fit.aic
  }
}
```

```
## Warning in log(s2): NaNs produced
```

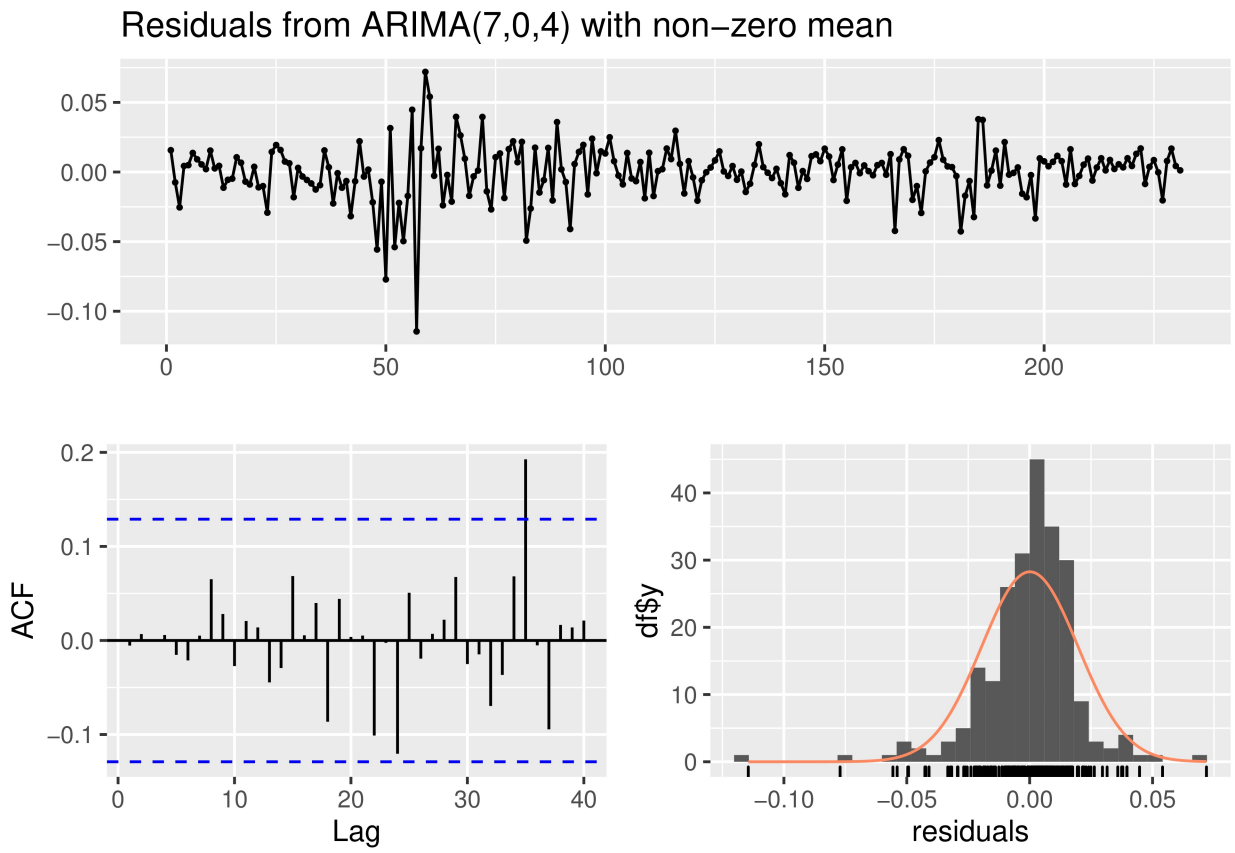
```
best.order; best.aic
```

```
## [1] 7 0 4
```

```
## [1] -1132.937
```

```
model.fit = best.arima
```

```
checkresiduals(model.fit, lag.max = 40)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(7,0,4) with non-zero mean
## Q* = 3.6149, df = 3, p-value = 0.3062
##
## Model df: 12.   Total lags used: 15
```

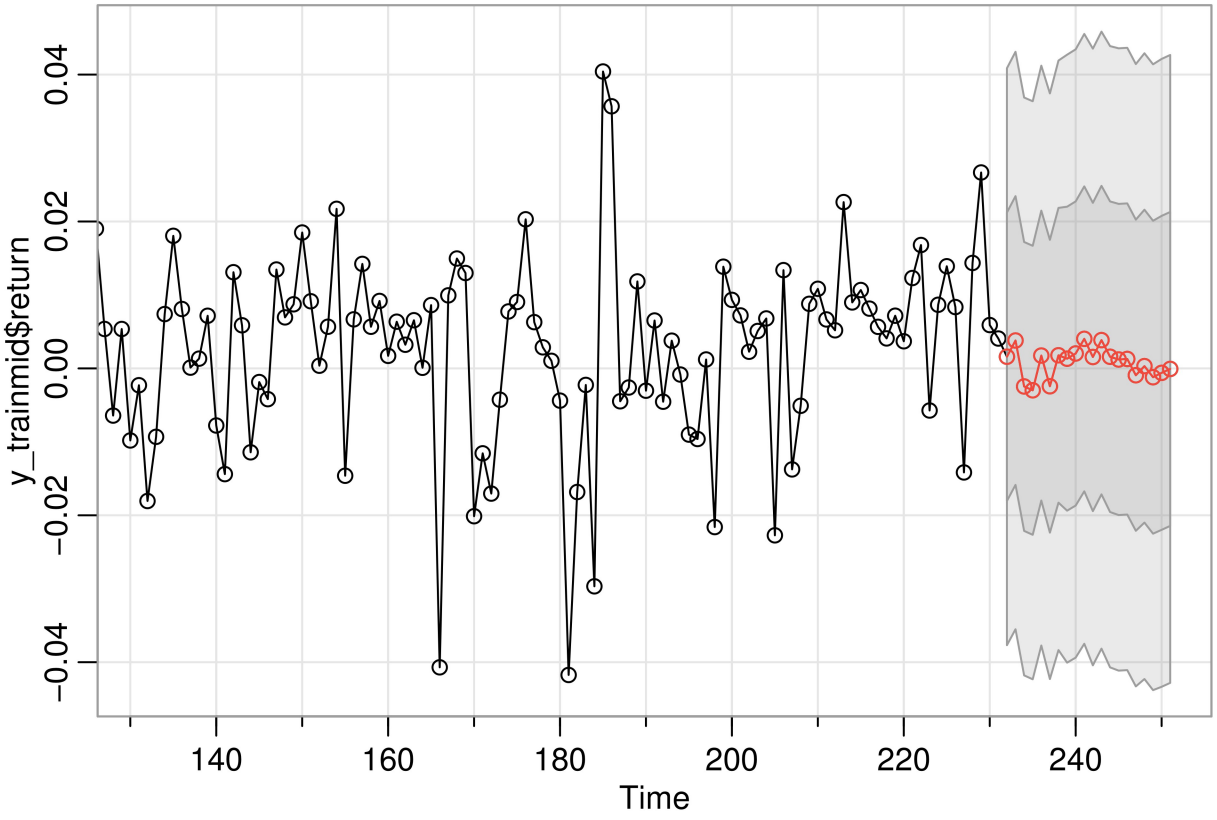
```
######Violations of assumption here so ARCH can be fir here
```

```
forecast2 = sarima.for(y_trainmid$return, n.ahead =length(y_testmid$return), p=7, d=0, q=4)
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```



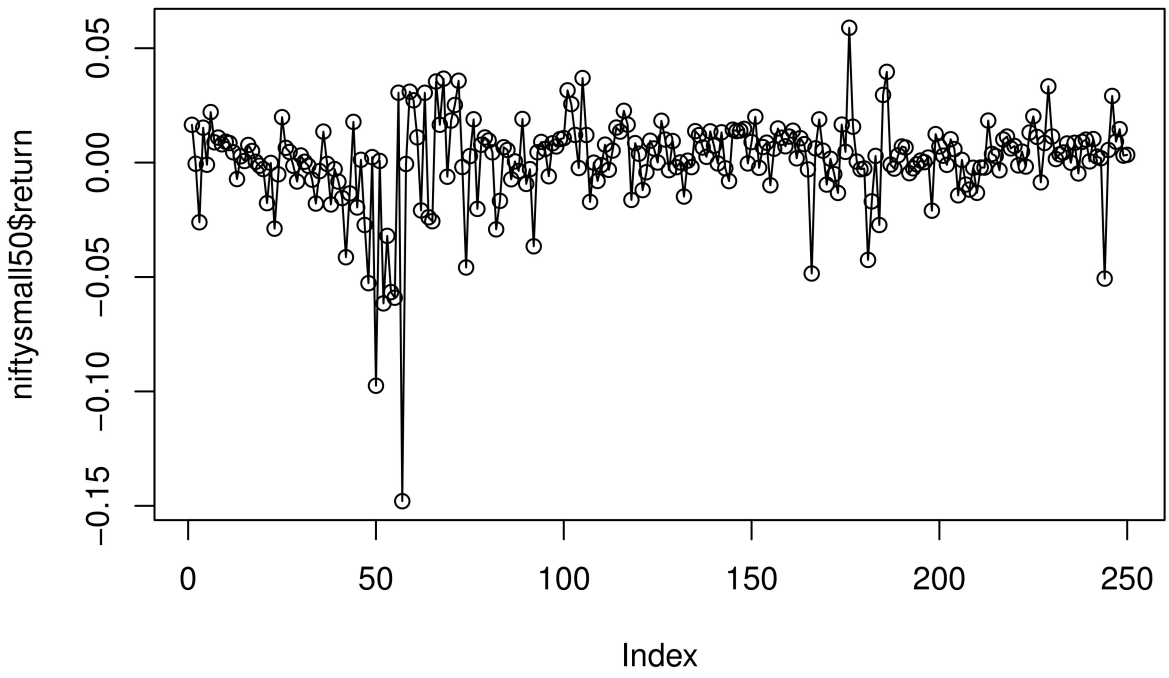

```
p2 = forecast2$pred
obs2 = y_testmid$return
MSE_foreast2 = mean(abs((p2 - obs2)))
MSE_foreast2
```

```
## [1] 0.01018547
```

```
####
```

###For nifty small cap

```
rm(list=ls())
niftysmall50=read.csv('nifty smallcap50.csv')
niftysmall50$Date <- as.Date(niftysmall50$Date)
niftysmall50$Close=as.numeric(as.character(niftysmall50$Close))
price1=niftysmall50$Close
log_return2=diff(log(price1),lag=1)
niftysmall50=niftysmall50[-1,]
niftysmall50['return']=log_return2
plot(niftysmall50$return,type="o")
```

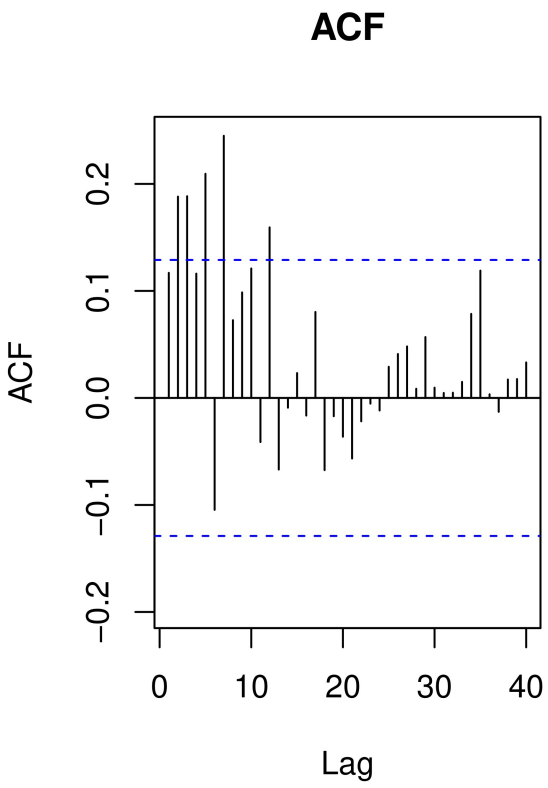


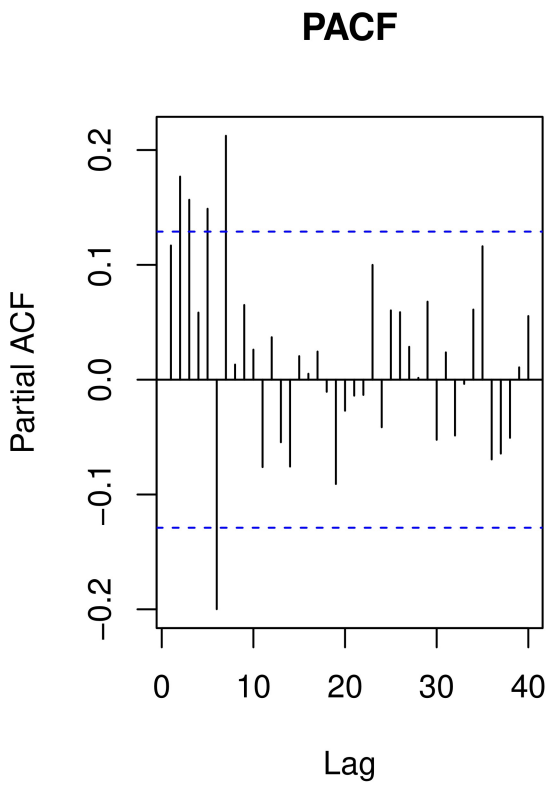

```
adf.test(niftysmall50$return, alternative = "stationary")
```

```
## Warning in adf.test(niftysmall50$return, alternative = "stationary"): p-value
## smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  niftysmall50$return
## Dickey-Fuller = -4.2713, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

```
par(mfrow=c(1,2))
y_trainsmall=niftysmall50[1:231,]
y_testsmall=niftysmall50[231:250,]
Acf(y_trainsmall$return, lag.max = 40, main='ACF')
Pacf(y_trainsmall$return, lag.max = 40, main = 'PACF')
```






```
aic_model1=auto.arima(y_trainsmall$return,trace=TRUE,d=0,max.p=10,max.q=10,ic="aic",approximation=FALSE
```

```
##
##  ARIMA(2,0,2) with non-zero mean : -1144.999
##  ARIMA(0,0,0) with non-zero mean : -1134.417
##  ARIMA(1,0,0) with non-zero mean : -1135.591
##  ARIMA(0,0,1) with non-zero mean : -1134.792
##  ARIMA(0,0,0) with zero mean     : -1136.184
##  ARIMA(1,0,2) with non-zero mean : -1146.998
##  ARIMA(0,0,2) with non-zero mean : -1138.085
##  ARIMA(1,0,1) with non-zero mean : -1146.928
##  ARIMA(1,0,3) with non-zero mean : -1145.002
##  ARIMA(0,0,3) with non-zero mean : -1143.255
##  ARIMA(2,0,1) with non-zero mean : -1146.915
##  ARIMA(2,0,3) with non-zero mean : -1149.802
##  ARIMA(3,0,3) with non-zero mean : -1156.435
##  ARIMA(3,0,2) with non-zero mean : -1150.845
##  ARIMA(4,0,3) with non-zero mean : -1154.838
##  ARIMA(3,0,4) with non-zero mean : -1155.178
##  ARIMA(2,0,4) with non-zero mean : -1148.629
##  ARIMA(4,0,2) with non-zero mean : -1149.722
##  ARIMA(4,0,4) with non-zero mean : -1144.885
##  ARIMA(3,0,3) with zero mean     : -1158.332
##  ARIMA(2,0,3) with zero mean     : -1151.702
##  ARIMA(3,0,2) with zero mean     : -1152.745
```

```
##  ARIMA(4,0,3) with zero mean     : Inf
##  ARIMA(3,0,4) with zero mean     : -1157.07
##  ARIMA(2,0,2) with zero mean     : -1146.893
##  ARIMA(2,0,4) with zero mean     : -1150.526
##  ARIMA(4,0,2) with zero mean     : -1151.62
##  ARIMA(4,0,4) with zero mean     : -1146.776
##
##  Best model: ARIMA(3,0,3) with zero mean
```

```
aic_model1
```

```
## Series: y_trainsmall$return
## ARIMA(3,0,3) with zero mean
##
## Coefficients:
##           ar1     ar2     ar3     ma1      ma2      ma3
##       -0.7801  0.5784  0.7519  0.8740  -0.3704  -0.5178
## s.e.   0.0985  0.1324  0.0894  0.1268   0.1840   0.1137
##
## sigma^2 estimated as 0.0003748:  log likelihood=586.17
## AIC=-1158.33   AICc=-1157.83   BIC=-1134.23
```

```
bic_model1=auto.arima(y_trainsmall$return,trace=TRUE,d=0,max.p=10,max.q=10,ic="bic",approximation=FALSE
```

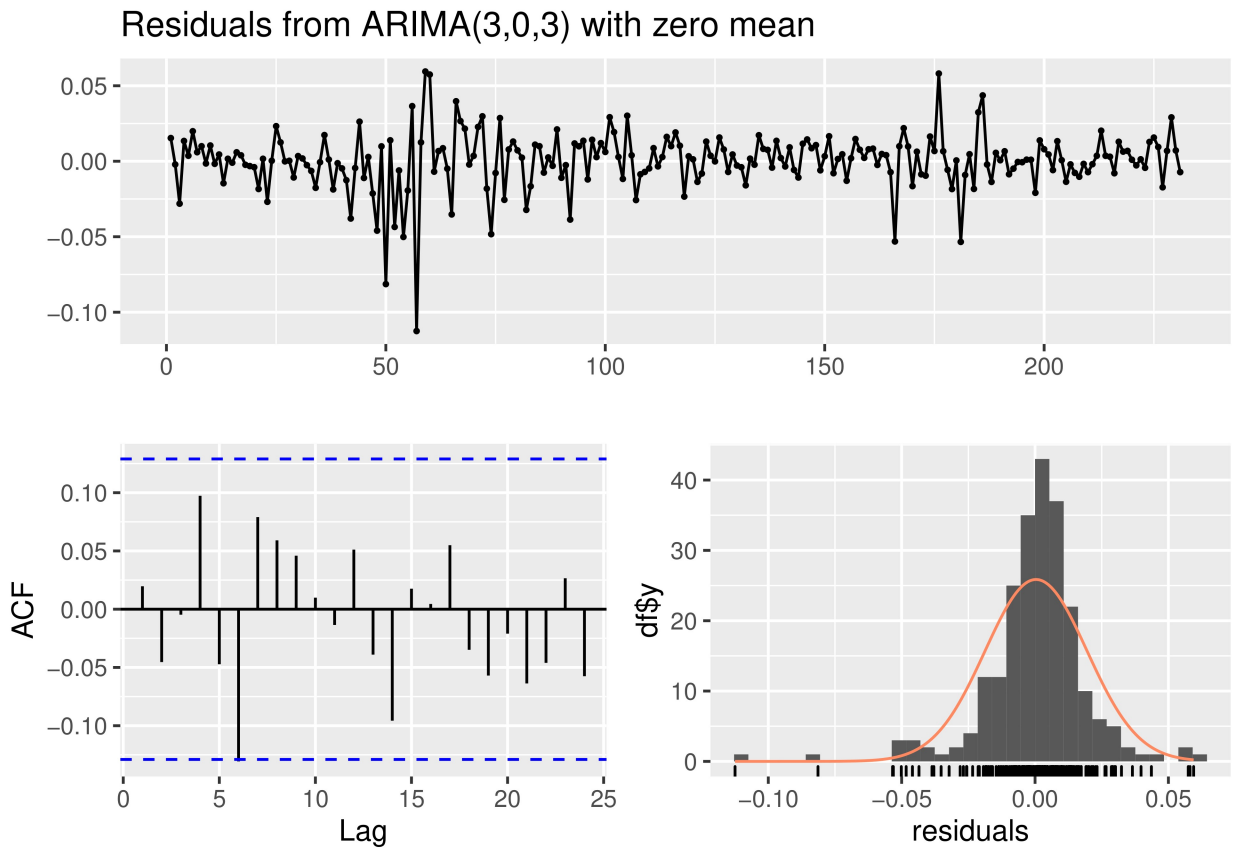
```
##
##  ARIMA(2,0,2) with non-zero mean : -1124.345
##  ARIMA(0,0,0) with non-zero mean : -1127.532
##  ARIMA(1,0,0) with non-zero mean : -1125.264
##  ARIMA(0,0,1) with non-zero mean : -1124.465
##  ARIMA(0,0,0) with zero mean     : -1132.741
##  ARIMA(1,0,1) with non-zero mean : -1133.158
##  ARIMA(2,0,1) with non-zero mean : -1129.703
##  ARIMA(1,0,2) with non-zero mean : -1129.786
##  ARIMA(0,0,2) with non-zero mean : -1124.316
##  ARIMA(2,0,0) with non-zero mean : -1127.129
##  ARIMA(1,0,1) with zero mean     : -1138.487
##  ARIMA(0,0,1) with zero mean     : -1129.705
##  ARIMA(1,0,0) with zero mean     : -1130.516
##  ARIMA(2,0,1) with zero mean     : -1135.038
##  ARIMA(1,0,2) with zero mean     : -1135.122
##  ARIMA(0,0,2) with zero mean     : -1129.583
##  ARIMA(2,0,0) with zero mean     : -1132.426
##  ARIMA(2,0,2) with zero mean     : -1129.681
##
##  Best model: ARIMA(1,0,1) with zero mean
```

```
bic_model1
```

```
## Series: y_trainsmall$return
## ARIMA(1,0,1) with zero mean
##
## Coefficients:
```

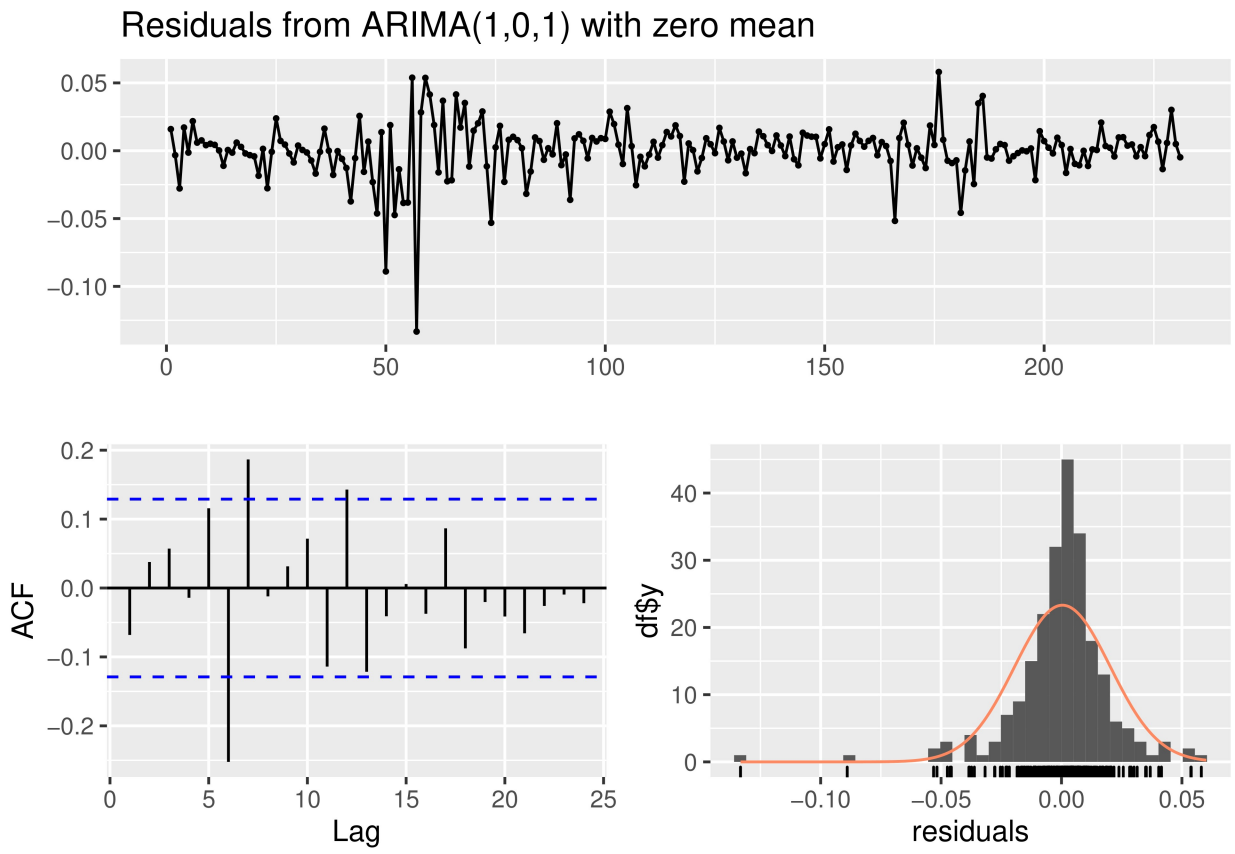
```
##          ar1      ma1
##       0.9019  -0.7846
## s.e.  0.0584   0.0806
##
## sigma^2 estimated as 0.000398:  log likelihood=577.41
## AIC=-1148.81   AICc=-1148.71   BIC=-1138.49
```

```
checkresiduals(aic_model1)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,0,3) with zero mean
## Q* = 10.32, df = 4, p-value = 0.03536
##
## Model df: 6.   Total lags used: 10
```

```
checkresiduals(bic_model1)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,1) with zero mean
## Q* = 30.543, df = 8, p-value = 0.0001694
##
## Model df: 2.   Total lags used: 10
```

Taking the aic model because bic model violates the assumptions

```
fcst1=forecast(aic_model1,h=20)
p3 = fcst1$mean
obs3 = y_testsmall$return
MSE_foreast3 = mean(abs(p3 - obs3))
MSE_foreast3
```

```
## [1] 0.008600537
```