FINAL PROJECT REPORT

Vallab Datta
Millin Desai
Syed Qavi
Shadhana Palaniswami
Dharun Selvan

Table of Contents

NON-TECHNICAL SUMMARY

Introduction

We planned to work on nifty 50 stock market data. It contains the stock market in India and has about 50 top companies in that market since we are a group of 5 students, we plan to work on two data sets. One is TCS data set and the other one is HDFC data set. The data sets consist of 16 years of data from 2004 to 2020. TCS (Tata Consultancy Services limited) is a Indian multinational information technology services as well as consultancy company that has its quarters in Mumbai and it operates about 46 countries and 149 locations. HDFC Bank is also an Indian private sector banking the top rated headquarter in Mumbai Maharashtra it is the largest bank in India by market capitalization as of March 2020. The main aim of this project is to find whether it is good to invest in both of these stocks and to find when to sell those stocks we invested in order to get the profit and how the company's projection will be in the future using multiple time series analysis concepts.

HDFC VWAP:

The volume weighted average price (VWAP) is a trading benchmark used by traders that gives the average price a security has traded at throughout the day, based on both volume and price. It is important because it provides traders with insight into both the trend and value of a security.VWAP is calculated by adding up the dollars traded for every transaction (price multiplied by the number of shares traded) and then dividing by the total shares traded.Large institutional buyers and mutual funds use the VWAP ratio to help move into or out of stocks with as small of a market impact as possible. Therefore, when possible, institutions will try to buy below the VWAP, or sell above it.Traders may use VWAP as a trend confirmation tool, and build trading rules around it. For example, when the price is above VWAP they may prefer to initiate long positions. When the price is below VWAP they may prefer to initiate short positions.

So in our analysis of the VWAP we have tried to predict the VWAP for the next twenty months and the VWAP maintains the same trend as it is now .Hence it is a good sign and the traders can go ahead and buy the stock or the companies planning to become a short period partner can go ahead and become one ,it will only result in profit.In this we have not considered any unforeseen changes like a recession or stock market crash .

HDFC CLOSE:

HDFC Close means the points at the end of the day when the stock closes on the day .This plays a major role in determining the weather we can buy the stock or not .When we developed the model and analyzed we could see the close variable showed a more stable trend .Which means that the price is stable for another year or so ,and it is a good to buy the stock at this interval .We have not considered any circumstances like a stock market crash ,pandemic etc .This might affect the prices in the future and we cannot consider those circumstances and their consequences unfortunately while modeling ,since they are unforeseeable .

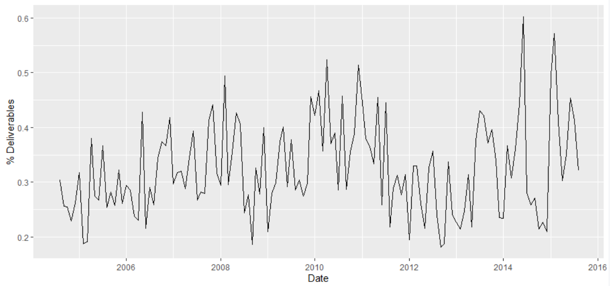
HDFC HIGH:

HDFC 'High' variable was analysed using various time series analysis models. Variable 'high' in the data set actually denotes the peak value attained by the stock on each day. This was analysed in the view of increasing the profit by finding the best time to buy and sell the stocks. Throughout the analysis, the variables have some good positive trend i.e the value of the stock is rising overall. The variable showed some nice trend before predictions and it continued after the analysis as well. This shows that it's good to buy those stocks.

Analyzing the Percentage deliverables of TCS stock market:

Exploratory Data Analysis:

Created time series of percentage deliverables ranging from 2004 to 2015 using monthly frequency. The time series plot looks stationary with a trend. The below adfTests table indicates that we can reject a unit root if we include either trend or a constant. The kpss test also supports that if the trend is included then the series does not have unit root.




```
> adfTest(TS_TCS, type='ct')

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: -6.1581
  P VALUE:
    0.01

Description:
 Mon Nov 23 19:49:51 2020 by user: vallab

Warning message:
In adfTest(TS_TCS, type = "ct") : p-value smaller than printed p-value
> kpss.test(TS_TCS, null='Trend')

        KPSS Test for Trend Stationarity

data:  TS_TCS
KPSS Trend = 0.13864, Truncation lag parameter = 4, p-value = 0.06363
```

ADF type='ct'	ADF type='c'	ADF type='nc'
P-value is <0.05	P-value is <0.05	P-value is >0.05

The normal plot (Appendix VD 1) and jarque.bera test (Appendix VD 2) show that the series is not normal. The acf and pacf plot (Appendix VD 3) has quick decay and has one significant spike indicating that the series exhibits stationarity and has autocorrelation at lag1. Since the series is trend stationary the first differencing and log returns leads to overdifferencing.

Modeling:

Firstly, performed regression on the trend. The summary of the regression model (Appendix VD 4) shows that time coefficient is significant and has reasonable slope. To check how good the fit residual analysis is performed. The residuals of the regression model have autocorrelation. Including the time for regression on the ARIMA model leads to insignificant coefficients, excluding the regression term.

The eacf plot shows ARIMA(1,0,1) or ARIMA(1,0,2). It was observed that MA(1) coefficient is insignificant in ARIMA(1,0,1) hence, fit model2 with ARIMA(1,0,2). The acf plot of squared residuals (Appendix VD 5) shows autocorrelation at lag7. The plot of squared residuals (Appendix VD 5) show volatility indicating arch effect. Then fitted model 3 using GARCH(1,1) with ARIMA(1,1). Residual analysis was performed on all the models to select the final model.

Diagnostics/Residual Analysis:

The residuals of model 1 (Appendix VD 6) show slight normality but have outliers. The residuals also show noticeable bias and significant autocorrelation. From the residuals of model 2 (Appendix VD 7) does not have any significant autocorrelation. The Box-pierce test (Appendix VD 7) also supports that we cannot reject white noise. The residuals of the GARCH model (Appendix VD 8) exhibit white noise behaviour. From squared residuals plot, It is observed that volatility was significantly reduced.

Model	AIC	BIC	Likelihood Measure
ARIMA(1,0,2)	-284.55	-270.1	147.28
ARIMA(1,1) GARCH(1,1)	-2.147	-2.016	148.7

The AIC and BIC values are less and log likelihood is more for the GARCH model when compared to ARIMA model. From the above residual analysis I would select the GARCH model as the final model.

Forecast Analysis:

The forecast of the ARIMA(1,0,2) (Appendix VD 9) model shows the expected behaviour of trend stationary series and it falls quickly to the mean. The 95% confidence interval in the GARCH model shows the size of the average volatility of the series. From the forecast of the GARCH model (Appendix VD 9) we can interpret that trend falls slowly to the mean and the lower confidence interval is increasing over the period of time.

Results and Discussion:

The alpha value of the GARCH model is 0.24 which indicates that the volatility will react to the shocks by adding 24% to the current volatility. The beta value is too low, this indicates that the volatility takes very less time to get back to the average variance. From the alpha and beta values we can conclude that the volatility of the percentage deliverables will persist for a short duration after reaching its peak.

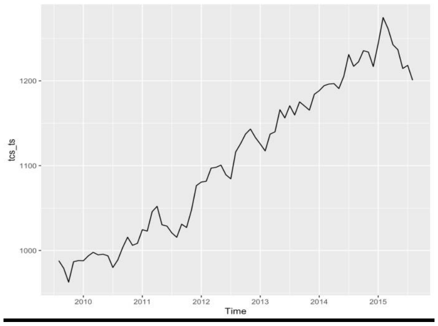
Analyzing the Close of TCS stock market:

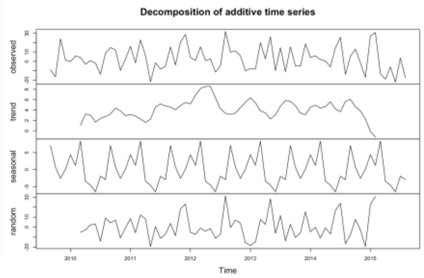
Exploratory Data Analysis:

Distribution of the close, you can see that the graph is not normally distributed; it has two peaks and it also has notable outliers in the data. In addition, from the histogram you can say there is high stock closing and low stock closing. With two different peaks it shows hypothetical bi-model distribution. You can easily identify global maximum and local maximum. So this data is split in two modes one distincts closing price which is around 1000 and another local minimum which is around 2500. We separate both distributions; it is not almost normally distributed with little left skewed. Now we will try time series plots.

Time series plot:

Here, I have considered frequency as 12 and created a time series for close. The time series plot is shown below. In addition, we are considering data from 2009 to 2015 only. From the time plot you can see that the closing price clearly shows upward trends with additive behavior. Also, you can observe seasonal components in the time series plot of the TCS stock data. From the graph below of the difference of series, You can observe seasonal patterns.

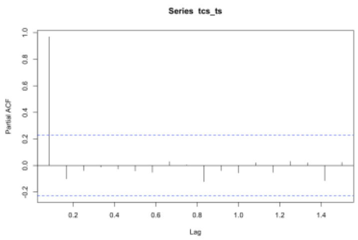


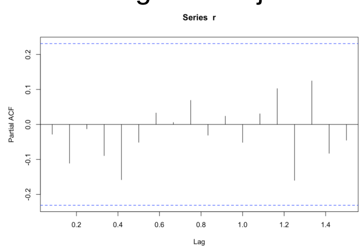


From the above decomposition graph, you can see that the graph follows a seasonality and it has an upward trend with a downfall. This time series has a random factor which shows it has moderately randomness.

Modeling:

I plotted the acf graph as above from the graph. You can say that there is no autocorrelation. Because graphs have a seasonal effect and trend also so you can say that it is non-stationary and so you can see that random walk. This model can be an AR1 model. Straight line shows drift in the model. Box-Jung test I rejected null hypothesis





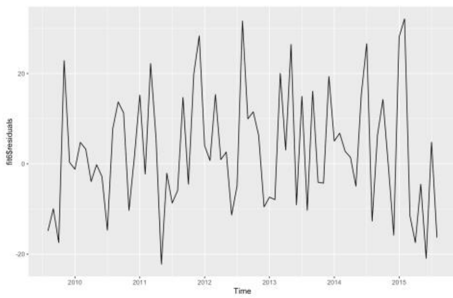
From the above eacf table you can select the order of the AR1 and MA0 and then can move forward. I've also tried the attached dicy-fuller test from which we can say p value is greater than so we can reject null hypothesis.

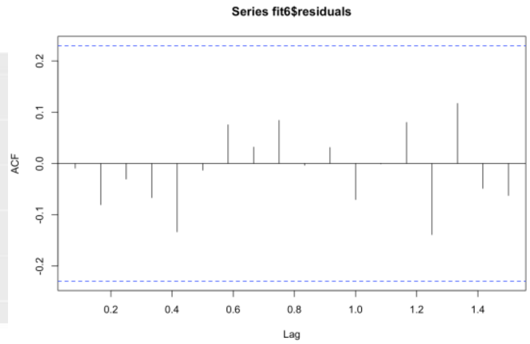
The autocorrelation function is decaying slowly which supports that the series is not moving average. From the acf plot it is difficult to tell the order of the model hence, let us check the pacf plot. From the, Box-Ljung test 0.6759 so we can not reject the null hypothesis.

```
> fit6 = Arima(tcs_ts, order=c(1, 0, 0))
> coeftest(fit6)

z test of coefficients:

           Estimate Std. Error z value  Pr(>|z|)
ar1       9.9082e-01 1.0525e-02  94.141 < 2.2e-16 ***
intercept 1.0978e+03 8.7217e+01  12.587 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual Analysis and Model Fitting:

Among all models, the best model was with the AR and MA 0 order (0,1,1). SSkewness is fairly symmetrical, as skewness is between -0.5 to 0.5. So we can say data is not skewed. Also from kurtosis you can see that it is negative - -0.69 so we can say that the normal graph is not too peaked or too flattened. From the jarque bera test you can see the mean of 2 and p-value 0.3484 so we can say this doesn't come from the normal distribution, so we can reject the null hypothesis.



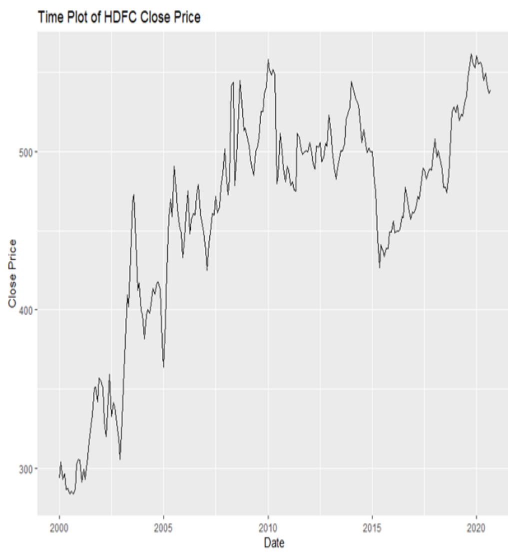


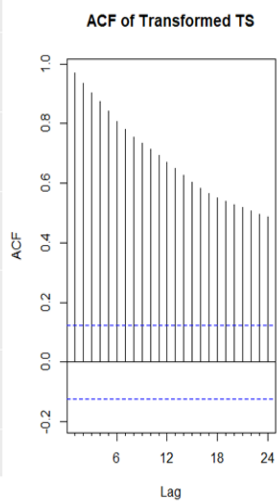
However, The forecast of the ARIMA(0,1,1) was following the mean with increasing confidence intervals. Also you can see that the graph is in a positive trend. You can also v verify that by pulling up the validation set after 2015 until 2020 and easily compare your prediction.

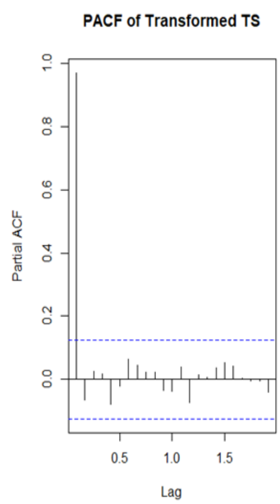
Analysing the Close variable of HDFC Stock market

Exploratory Analysis

The time plot of HDFC Close price vs. time is seen below ranging from 2000 to 2020, frequency is monthly. It looks like a random walk with drift and non-stationary. KPSS and adf test agree that this is a random walk with drift. It looks like the data needs to be transformed. The Box Cox Transformation method was used to transform the data and the best lambda was 2.0.







ADF TEST (type = “constant, trend”)	P-value: 0.1337
KPSS Test for Level Stationary	P-value: 0.01

The adf test and KPSS test both agree that it is a random walk with drift. The ACF shows exponential fall off so time series has non stationarity and PACF shows AR(1) behavior. The eacf (Appendix SQ 1) also showed unit root and showed an ARIMA(1,1,2) or (2,1,2).

Modeling Fitting

ARIMA model was fitted for the time series using the results from the exploratory analysis and it was modeled with a drift as seen in the time plot. Then a GARCH(1,1) model was fitted for the data, but first a log return of the HDFC Close price was taken to remove non stationarity and model the volatility and fit with GARCH model. Two GARCH models were performed on the returns, the standard GARCH fit and ugarchfit model.

Residual Analysis and Model Diagnostics

ARIMA model produced two different models that are very close (Appendix SQ 2), ARIMA(1,1,2) and ARIMA(0,1,2). Both have all significant coefficients and residual analysis are similar. AIC are similar but BIC is lower for (0,1,2) model but likelihood measure is higher for (1,1,2) model and the Ljung Box test rejects white noise with higher p value.

	AIC/BIC	Likelihood Measure	Ljung Box Test(lag 10)
ARIMA (1,1,2)	550/567	-2520	P-value: 0.169
ARIMA(0,1,2)	550/560	-2522	P-value: 0.1024

Residual Analysis was performed on the log returns to see if GARCH model would be useful. The squared residual plot (Appendix SQ 7) shows the change in volatility and plot of absolute residual (Appendix SQ 8) helped show the heteroscedasticity and showed that performing a GARCH model will be helpful in understanding the volatility of the HDFC Close price.So, a GARCH(1,1) model was fit with the ARMA(1,2) model (Appendix SQ 9), which gave the model equation : sigma^2 = 0.04_a^2_t-1 + 0.956_sigma^2_t-1. A ugarchfit model was performed as well to see if the GARCH model would improve (Appendix SQ 12), which gave the model equation of sigma^2 = 0.039_a^2_t-1 + 0.957_sigma^2_t-1. Both had same number of significant coefficients and white noise residuals. But the standard GARCH model had lower AIC and BIC and better likelihood measure. Plot of GARCH fit residuals(Appendix SQ 10)

	AIC/BIC	Likelihood Measure
Standard GARCH	-4.26/-4.16	534.8
uGARCH	-4.25/-4.15	533.9

Forecast Analysis

The forecast of the ARIMA model (Appendix SQ 3) shows expected behavior of a random walk with drift, where the forecast is increasing at the rate of the trend. Recall, there were two similar models, the 80/20 backtest showed that the ARIMA(1,1,2) model had a lower MAPE at 2.637% compared to the ARIMA(0,1,2) which had a MAPE of 2.698%. The forecast of the log return of the ARIMA model (Appendix SQ 5) shows the forecast fall quickly to the mean and this is expected since the returns are not correlated. The range of the confidence intervals are similar to the returns the last few years. Forecast for the GARCH model was also performed (Appendix SQ 11). The forecast shows 95% confidence interval are the size of the average volatility. Backtest was also performed for GARCH and ugarch (Appendix SQ 13). The MAE is a little lower for ugarch which was 0.012 compared to GARCH which was 0.0124.

Results and Discussion

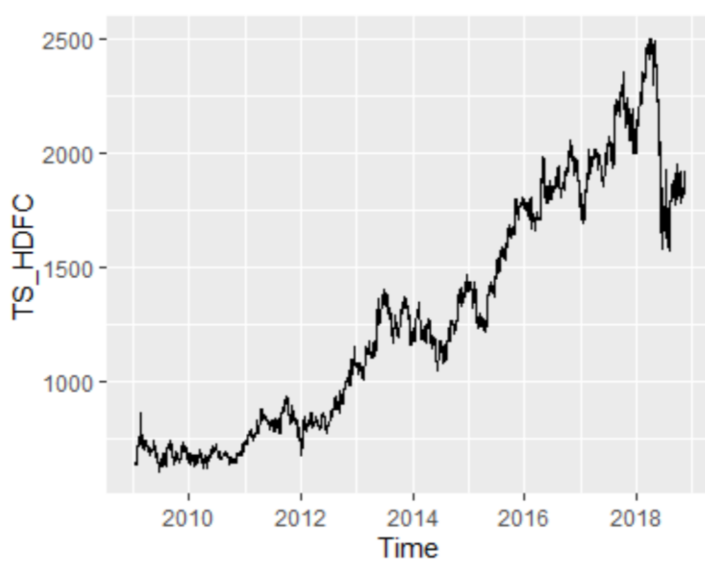
GARCH model results showed that the alpha or the lagged squared residuals is the reaction of the volatility to the shocks or how the lagged volatility adds to the current volatility which is around 4% which is not a lot. The beta is around 0.956 and is close to 1 and tells how long the volatility takes to get back to the long run average variance. The alpha and beta variables are close to one so we have long persistence and after volatility reaches its peak, it will continue for a longer duration of time. The volatility of the close stock price has been decreasing for the last few years and the forecast predicted that the volatility will continue to be low which means that it is a good time to buy stocks since the risk is low.

Analysing the High variable of HDFC stock market:

Exploratory Analysis:

The variable ‘High’ has only a slight difference from the close variable. For analysis of ‘high’ variables, we used only 11 years of the data. We found that there are no missing values or null values in the column. We learned that log return is required to get some good results. The transformed data looks stationary in the qq plot and that was confirmed using box test. When we did decompose the plot, we could see some seasonality with a good upward trend. We went on to start analysing ACF, PACF and EACF of the log return time series. The ACF had shown some Autoregressive characteristics. Pacf showed a similar trend. Eacf showed the possible models I can try with.

ADF TEST	P-value: 0.01
KPSS Test for Level Stationary	P-value: 0.1



Model Fitting:

We started with arima models for variable ‘High’. Because of the seasonality, all 3 Arima model based on the eacf i.e. Arima (1,0,0) , Arima (1,0,1) doesn’t produce any good results. Then we started modelling with seasonality. Sarima (0,0,1)(2,0,0)[21] with zero mean was the best model when compared to the Sarima(1,0,1)(2,0,0)[21] with zero mean. We tried the Garch model too. The variable ‘Close’ showed some good results. But the Garch model didn’t fit nicely to the ‘High’ variable. The ACF and ACF^2 also proved that.

Residual Analysis and Model Fitting:

Arima Models:

Arima models for variable ‘High’ didn’t fit nicely for variable ‘High’. This was confirmed by the residual analysis. Though P values of ADF test and KPSS test produce good results, both ACF and ACF^2 show some peaks above and below the lines.

	AIC/BIC	Likelihood Measure	Ljung Box Test/KPSS
ARIMA (1,0,0)	-13445.37 /-13422.11	6726.69	P-value: 0.01/0.1
ARIMA(1,0,1)	-13480.47 /-13652.43	6729.43	P-value: 0.01/0.1

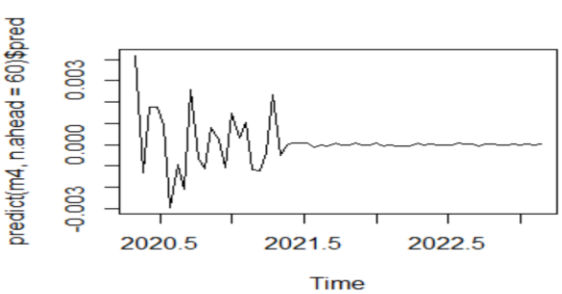
Sarima Models:

I did residual analysis to all those Sarima models. ACF, ACF^2 , adf test, KPSS test and back testing showed some good results and that suggests to me the best model.

	AIC/BIC	Likelihood Measure	Ljung Box Test
SARIMA (0,0,1)(2,0,0) [21]	-1119.23 /-1105.48	563.62	P-value: 0.01
SARIMA(1,0,1)(2,0,0) [21]	--1134.45 /-1124.68	566.38	P-value: 0.01

FORECAST ANALYSIS :

On plotting the forecast for next 3 years, it showed as a good result as we expected. The highest value is archived 03/2021 and falls to mean soon after that. All these were predicted in a 95% confidence interval.



Results and Discussion :

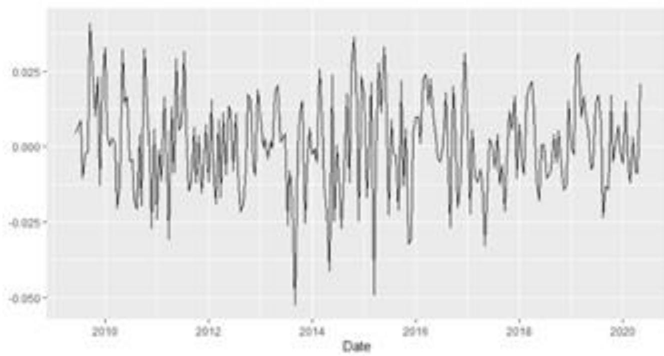
On analysing the variable ‘High’ of HDFC data, we can conclude that it is safe to invest in HDFC stock, very low risk the stock will have in late 2020’s but 2021 will be good. In order to get more profit, the person should sell his stock in the march,2021.

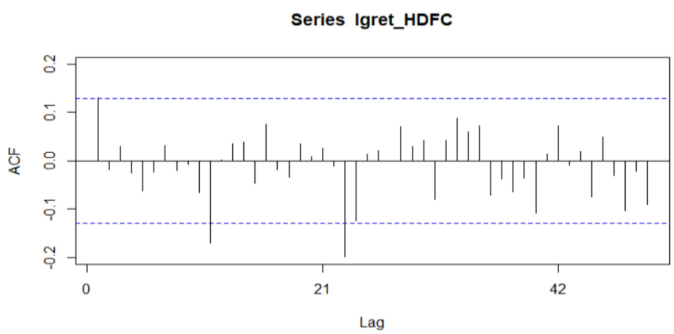
Analyzing The Volume Weighted Average Of Hdfc:

Exploratory Analysis:

The time series plot of HDFC VWAP vs time was taken and it's log return was calculated to transform the data . The frequency was monthly. We could observe that the time series was stationary and there was a weak seasonality .The adf test and KPSS test also shows that the time series is stationary .When we look at the ACF,PACF and the EACF plots we can see very weak seasonality in it .Since the seasonality is weak ,we can say that it is not going to affect the trading decisions much in the future .We could observe a seasonality at the lag 11 when we looked at the ACF/PACF plots . Below are the Autoplot and the ACF plot for the log returns of HDFC VWAP :

The PACF and the EACF plots are the Appendix SP 2 and SP 3





ADF TEST	**p-value :0.01**
KPSS TEST	**p-value:0.1**

Model Fitting:

Initially ARIMA models was built using the initial data exploration .ARIMA model with the order of (0,0,1) and seasonality order of (1,0,0) was the the best ARIMA model ,that could be built which could capture the weak seasonality as well.The auto- ARIMA gave a model with order (0,0,1) and seasonality order(2,0,0).But it failed to capture the seasonality since it is very weak.Then the volatility and the change in mean was plotted and a significant change in volatility and mean was noticed.Hence,a GARCH model with arma(0,1) was built .It turned out to be the best model also when the back testing was performed .

Residual Analysis and Model Fitting:

The best ARIMA model that was built has an order of (0,0,1) and seasonality order of (1,0,0).Residual analysis was performed on this model and it was noted that the residuals of the model ,form a white noise.The ACF plot(Appendix SP 6) and the autoplot suggest that there is not auto correlation .The Dickey fuller Test has a p value is 0.01 ,so it rejects the null hypothesis that there is non-stationarity ,and the KPSS test also has a p-value 0.1 failing to reject the null hypothesis that there is stationarity in the series. The Ljung-Box test also failed to reject the null hypothesis that there is white noise .The volatility and the change mean plot(Appendix SP7) suggested that there is significant change in the volatility and change of mean ,hence GARCH modeling was performed .Here only the ma1 and beta 1 values were significant.The AIC ,BIC and the log likelihood values of the GARCH model looked good and the Ljung-Box test also failed to reject the null hypothesis that there is white noise .We then performed the back testing for the best ARIMA model built and the GARCH model .The MAPE value for the best ARIMA model was 1.3 and that of the GARCH was 1.09 .So we can say that GARCH model is the best of the two ,and hence GARCH was determined to be the best model .The model equation is sigma^2= 0.9206*sigma^2_t-1.The plots of the GARCH variance and the

Model	AIC/BIC	Log-Likelihood	Ljung-Box Test	MAPE
ARIMA	-1244.56/-1230.8	626.28	0.9942	1.3
GARCH	-5.387/-5.297	625.5	0.99608	1.09

Forecast Analysis :

When we look at the forecast(Appendix SP7) of the ARIMA model build we can say the forecast fall in the mean and we can say it follows the same trend as that of the original series .When we look at the forecast plot(Appendix SP10) of the next twenty months for the GARCH model we can see that they fall in the mean.We can see the forecast shows that 90% confidence interval are the size of the average volatility .We can see that the forecast follows the same trend as that of the original time series and hence,it is a good idea to buy the stocks of HDFC .

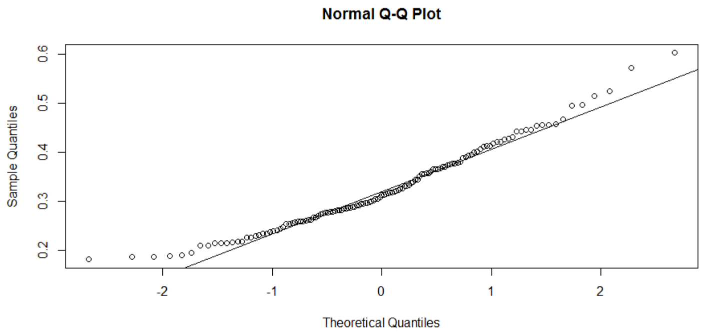
Results and Discussion :

When we look at the final GARCH model we can see that the Beta value is 0.9206 which is nearly 1.Beta represents s long run persistence .The beta value is almost 1 which me means we have a long persistence .we can see a reduce in volatility in the past few years and we can expect the same, for the next few months according to the forecast..Hence we can say that it is a good time to buy HDFC stocks.

APPENDIX

VD (Vallabha Datta)

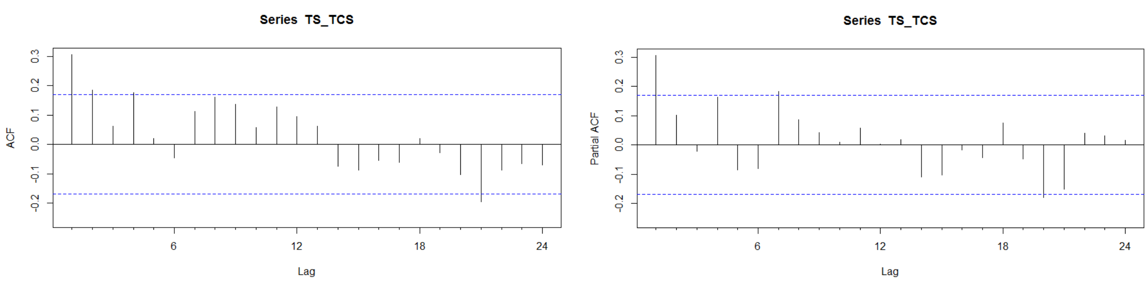
VD 1:



VD 2:

```
> jarque.bera.test(TS_TCS)

        Jarque Bera Test

data:  TS_TCS
X-squared = 9.2588, df = 2, p-value = 0.009761
```

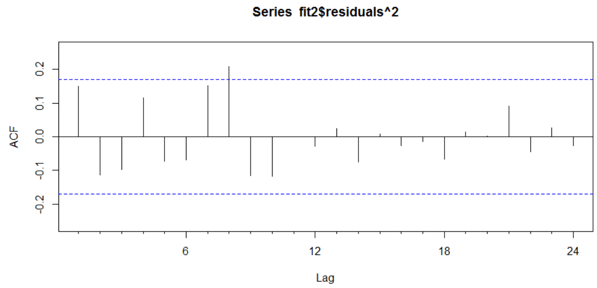
VD 3:

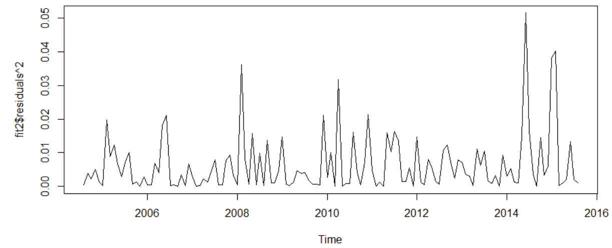


VD 4:

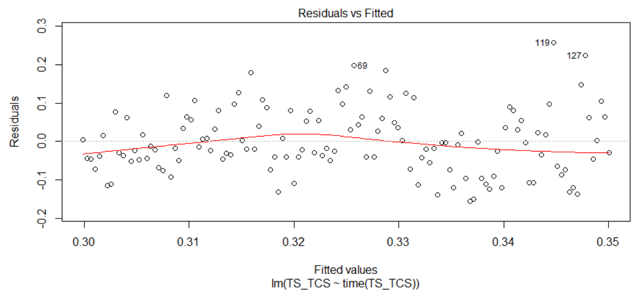
```
> fit1 = lm(TS_TCS ~ time(TS_TCS))
> summary(fit1)

Call:
lm(formula = TS_TCS ~ time(TS_TCS))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.15516 -0.04936 -0.01744  0.05614  0.25736 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -8.841437   4.596180  -1.924   0.0566 .
time(TS_TCS)  0.004560   0.002287   1.994   0.0482 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08437 on 131 degrees of freedom
Multiple R-squared:  0.02947,	Adjusted R-squared:  0.02206 
F-statistic: 3.977 on 1 and 131 DF,  p-value: 0.04819
```

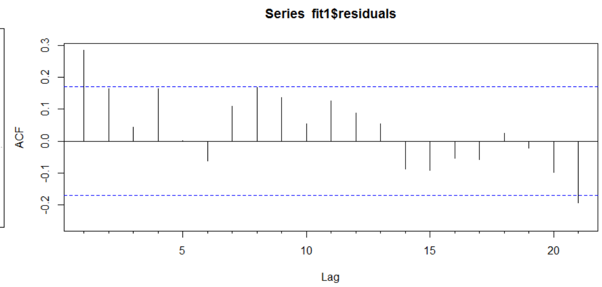
VD5:

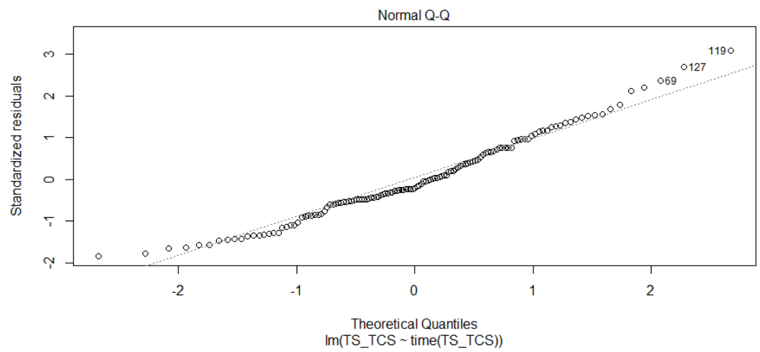




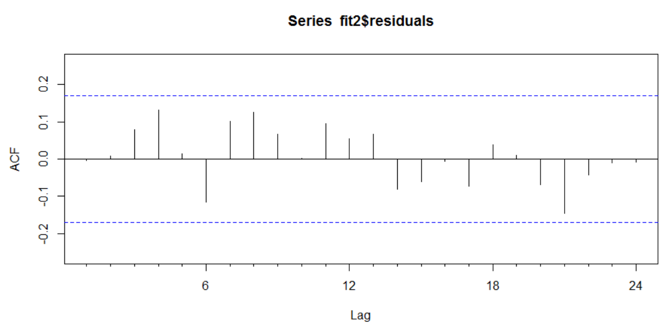
VD 6:





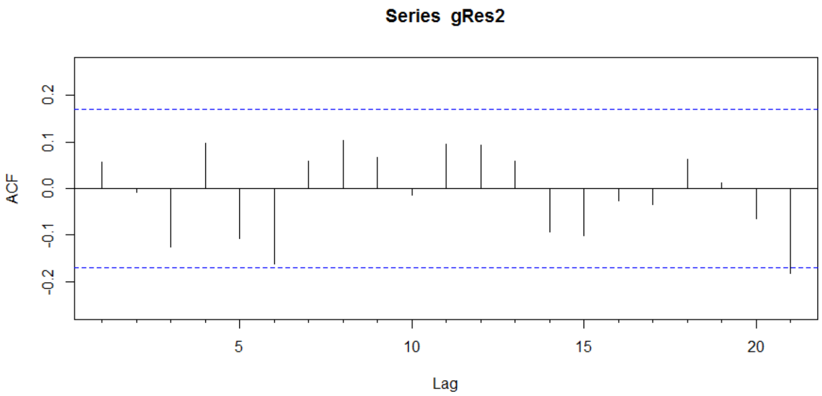


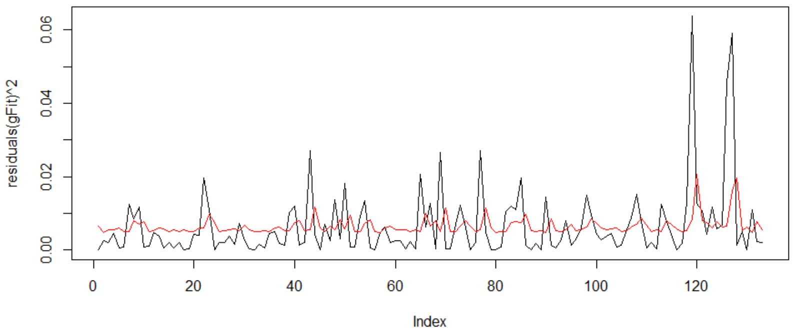
VD 7:




```
> Acf(fit2$residuals)
> Box.test(fit2$residuals, lag=20)    # Cannot reject white noise.

        Box-Pierce test

data:  fit2$residuals
X-squared = 14.032, df = 20, p-value = 0.8289
```

VD 8:



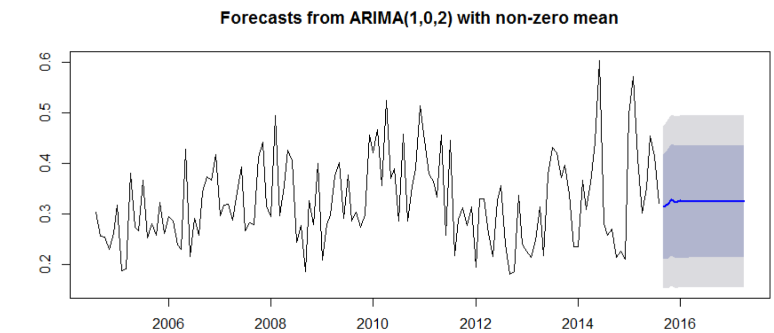


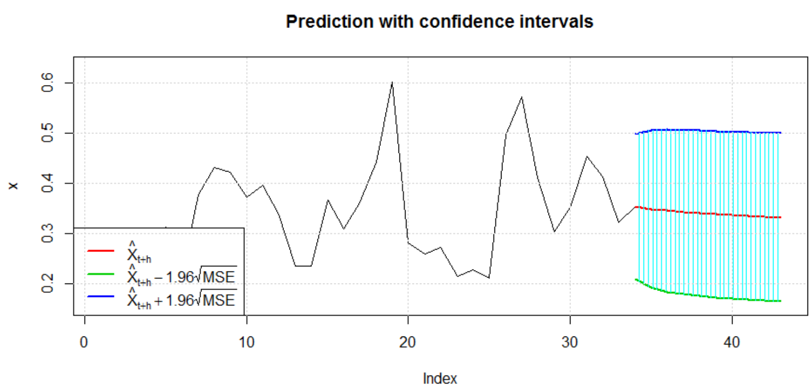
GARCH Model

```
Call:
 garchFit(formula = ~arma(1, 1) + garch(1, 1), data = TS_TCS,
    trace = F)

Mean and Variance Equation:
 data ~ arma(1, 1) + garch(1, 1)
<environment: 0x00000063ba14f518>
 [data = TS_TCS]

Conditional Distribution:
 norm

Coefficient(s):
        mu         ar1         ma1       omega      alpha1       beta1
 0.05061385  0.84481390 -0.65490048  0.00490190  0.24823793  0.00000001

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      5.061e-02   4.072e-02    1.243 0.213862
ar1     8.448e-01   1.255e-01    6.732 1.68e-11 ***
ma1    -6.549e-01   1.888e-01   -3.468 0.000524 ***
omega   4.902e-03   1.951e-03    2.512 0.011991 *
alpha1  2.482e-01   1.831e-01    1.356 0.175211
beta1   1.000e-08   3.457e-01    0.000 1.000000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 148.7807    normalized:  1.118652

Description:
 Mon Nov 23 20:58:22 2020 by user: vallab


Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  7.383408  0.02492949
 Shapiro-Wilk Test  R    W      0.9664433 0.002276031
 Ljung-Box Test     R    Q(10)  11.925    0.2901076
 Ljung-Box Test     R    Q(15)  17.96292  0.2646202
 Ljung-Box Test     R    Q(20)  19.55673  0.4859472
 Ljung-Box Test     R^2  Q(10)  18.33321  0.04959637
 Ljung-Box Test     R^2  Q(15)  20.06312  0.1695246
 Ljung-Box Test     R^2  Q(20)  20.5534   0.423825
 LM Arch Test       R    TR^2   19.53879  0.07632904

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-2.147079 -2.016687 -2.150920 -2.094092
```

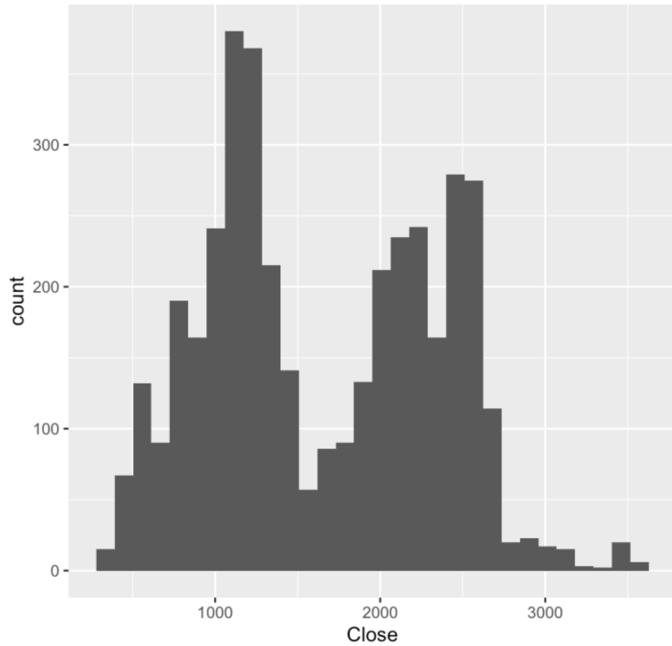
```
> gFit

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(1, 1) + garch(1, 1), data = TS_TCS,
    trace = F)

Mean and Variance Equation:
 data ~ arma(1, 1) + garch(1, 1)
<environment: 0x00000063ba14f518>
 [data = TS_TCS]

Conditional Distribution:
 norm

Coefficient(s):
        mu         ar1         ma1       omega      alpha1       beta1
 0.05061385  0.84481390 -0.65490048  0.00490190  0.24823793  0.00000001

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      5.061e-02   4.072e-02    1.243 0.213862
ar1     8.448e-01   1.255e-01    6.732 1.68e-11 ***
ma1    -6.549e-01   1.888e-01   -3.468 0.000524 ***
omega   4.902e-03   1.951e-03    2.512 0.011991 *
alpha1  2.482e-01   1.831e-01    1.356 0.175211
beta1   1.000e-08   3.457e-01    0.000 1.000000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 148.7807    normalized:  1.118652
```

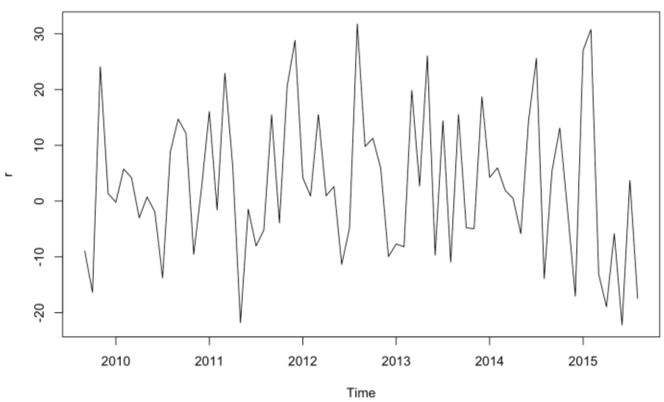
VD 9:

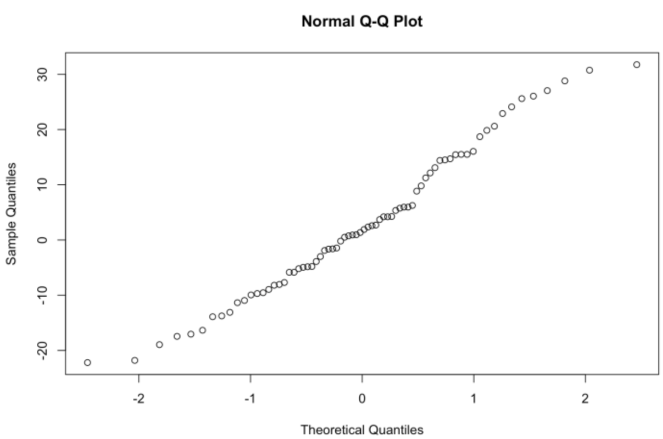


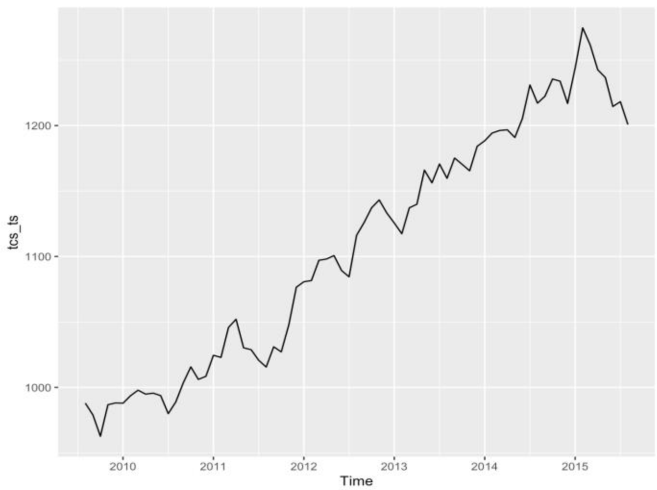


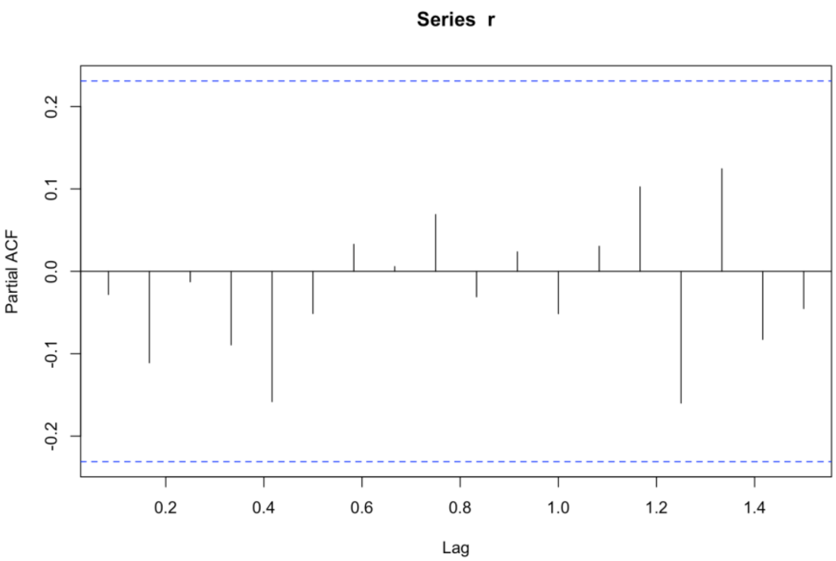
Milin (MD)


count
Close
0
100
200
300
1000
2000
3000


r
Time
2010
2011
2012
2013
2014
2015


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles





AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	o	o	o	o	o	o	o	o	o	o	o	o	o
1	o	o	o	o	o	o	o	o	o	o	o	o	o	o
2	o	x	o	o	o	o	o	o	o	o	o	o	o	o
3	o	o	x	o	o	o	o	o	o	o	o	o	o	o
4	x	o	o	o	o	o	o	o	o	o	o	o	o	o
5	x	o	o	o	o	o	o	o	o	o	o	o	o	o
6	o	o	o	o	o	o	o	o	o	o	o	o	o	o
7	x	o	o	o	x	o	o	o	o	o	o	o	o	o

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o o o o o  o  o  o
1 o o o o o o o o o o o  o  o  o
2 o x o o o o o o o o o  o  o  o
3 o o x o o o o o o o o  o  o  o
4 x o o o o o o o o o o  o  o  o
5 x o o o o o o o o o o  o  o  o
6 o o o o o o o o o o o  o  o  o
7 x o o o x o o o o o o  o  o  o
```

```
> skewness(r)
[1] 0.2736113
attr(,"method")
[1] "moment"
> kurtosis(r)
[1] -0.69054
attr(,"method")
[1] "excess"
> jarque.bera.test(r)

        Jarque Bera Test

data:  r
X-squared = 2.1088, df = 2, p-value = 0.3484
```

SQ (Syed Qavi)

SQ 1:

EACF of HDFC Close TIme Series

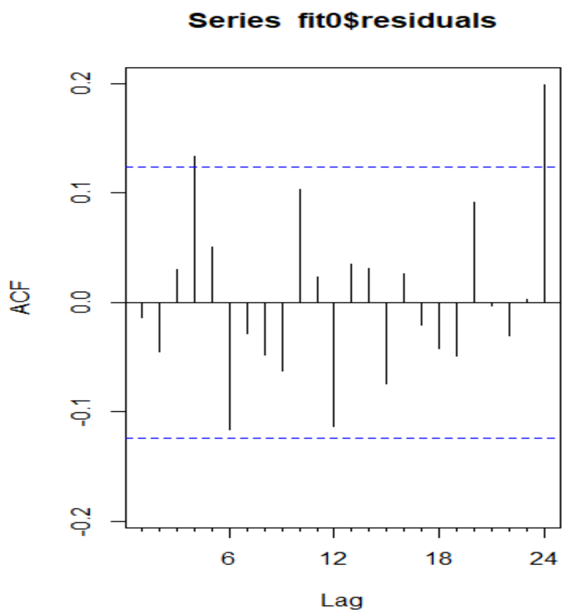
```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x x x x  x  x  x
1 x x o o o x o o o o o  o  o  o
2 x x o x o x o o o x o  o  o  o
3 o x o o o x o o o o o  o  o  o
4 x x o o o x o o x o o  o  o  o
5 o x o x o x o o o o o  o  o  o
6 o x o x o x o o o o o  o  o  o
7 x x o x o x x o o o o  o  o  o
```

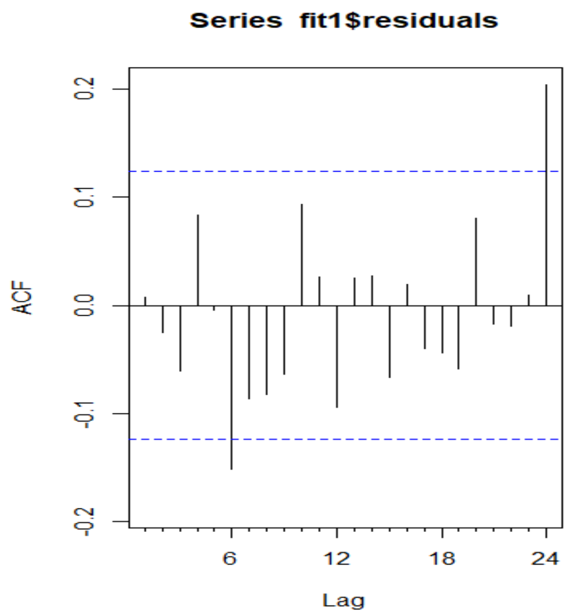
SQ 2:

```
Series: yBox
ARIMA(1,1,2) with drift

Coefficients:
         ar1      ma1      ma2     drift
      0.6793  -0.5338  -0.2863  431.9226
s.e.  0.1413   0.1449   0.0659  226.8521

sigma^2 estimated as 39882575:  log likelihood=-2520.18
AIC=5050.36   AICc=5050.61   BIC=5067.93
> coeftest(fit0)

z test of coefficients:

        Estimate Std. Error z value  Pr(>|z|)    
ar1     0.679304   0.141282  4.8081 1.523e-06 ***
ma1    -0.533806   0.144942 -3.6829 0.0002306 ***
ma2    -0.286311   0.065929 -4.3427 1.407e-05 ***
drift 431.922627 226.852054  1.9040 0.0569123 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Series: yBox
ARIMA(0,1,2)

Coefficients:
         ma1      ma2
      0.1322  -0.1881
s.e.  0.0631   0.0620

sigma^2 estimated as 40180149:  log likelihood=-2522.04
AIC=5050.09   AICc=5050.19   BIC=5060.63
> coeftest(fit1)

z test of coefficients:

     Estimate Std. Error z value Pr(>|z|)   
ma1  0.132207   0.063075  2.0960 0.036080 * 
ma2 -0.188093   0.061991 -3.0342 0.002412 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

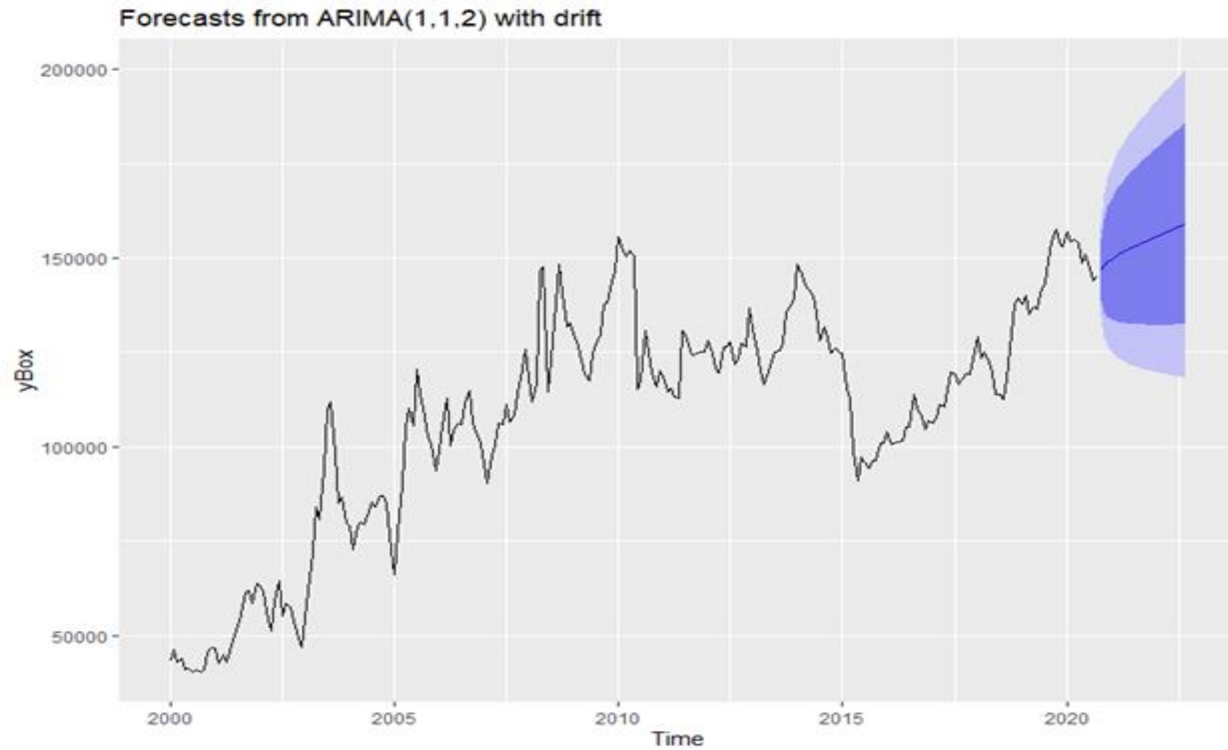
```
        Box-Ljung test

data:  fit1$residuals
X-squared = 15.903, df = 10, p-value = 0.1024

> Box.test(fit0$residuals, lag=10, type = "Ljung-Box")

        Box-Ljung test

data:  fit0$residuals
X-squared = 14.087, df = 10, p-value = 0.169
```



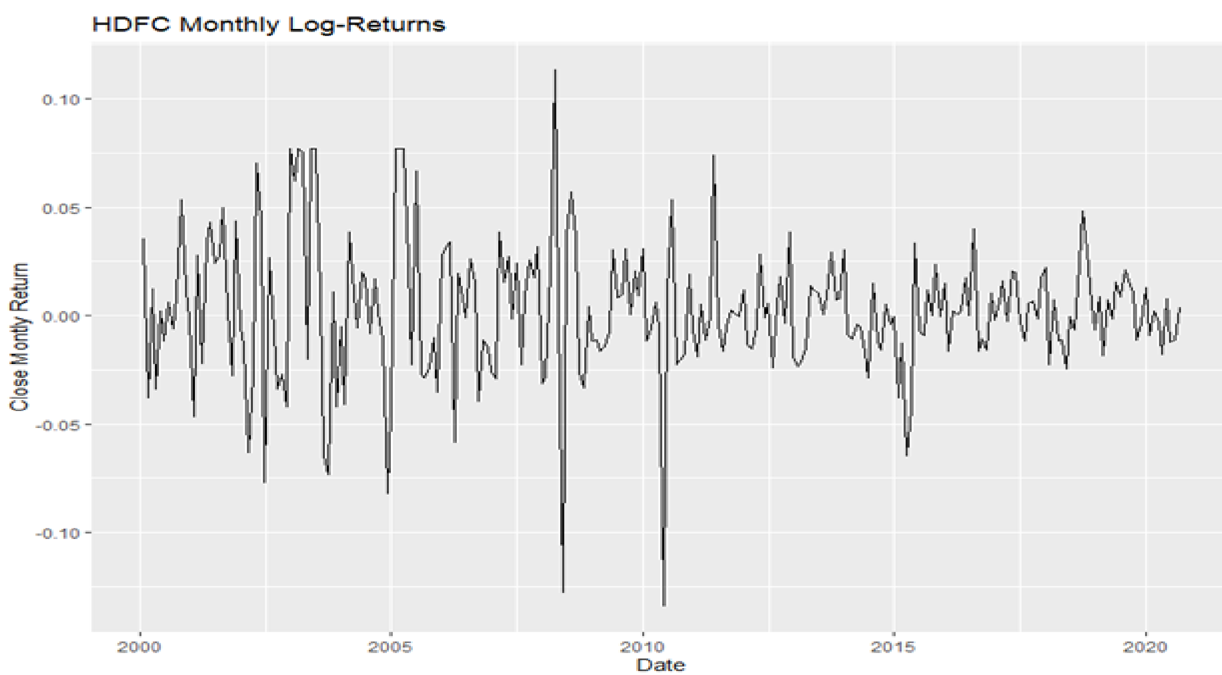


SQ 3:

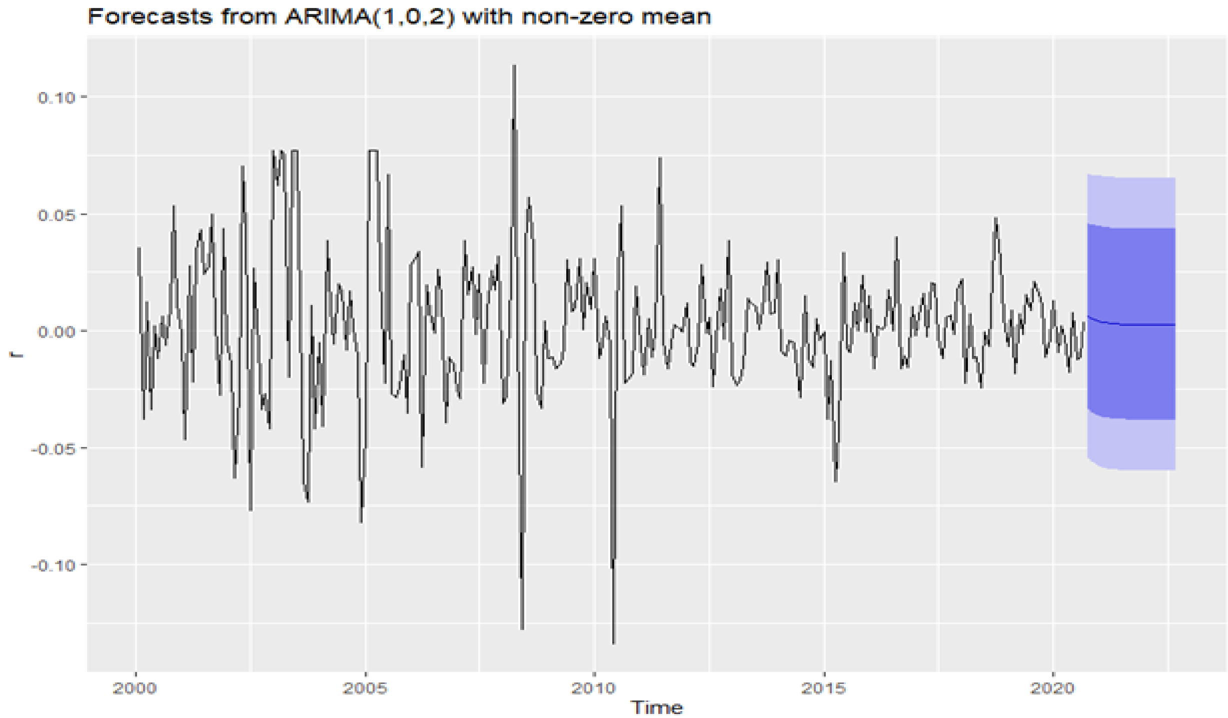
```
> backtest(fit0, yBox, orig=.8*length(yBox), h=1)
[1] "RMSE of out-of-sample forecasts"
[1] 4048.008
[1] "Mean absolute error of out-of-sample forecasts"
[1] 3407.781
[1] "Mean Absolute Percentage error"
[1] 0.02637116
[1] "Symmetric Mean Absolute Percentage error"
[1] 0.02654871
> backtest(fit1, yBox, orig=.8*length(yBox), h=1)
[1] "RMSE of out-of-sample forecasts"
[1] 4170.911
[1] "Mean absolute error of out-of-sample forecasts"
[1] 3440.767
[1] "Mean Absolute Percentage error"
[1] 0.02676895
[1] "Symmetric Mean Absolute Percentage error"
[1] 0.0269761
```


Forecasts from ARIMA(1,1,2) with drift
200000
150000
100000
50000
yBox
2000
2005
2010
2015
2020
Time

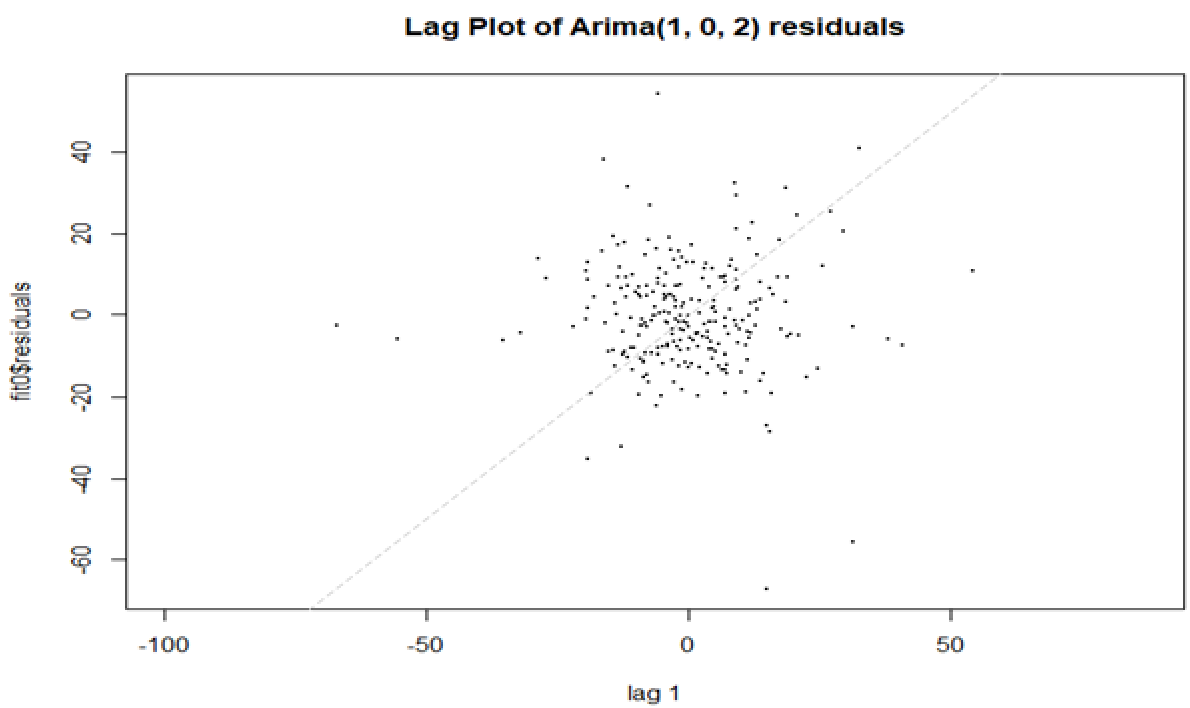
SQ 4:


HDFC Monthly Log-Returns
0.10
0.05
0.00
-0.05
-0.10
Close Montly Return
2000
2005
2010
2015
2020
Date

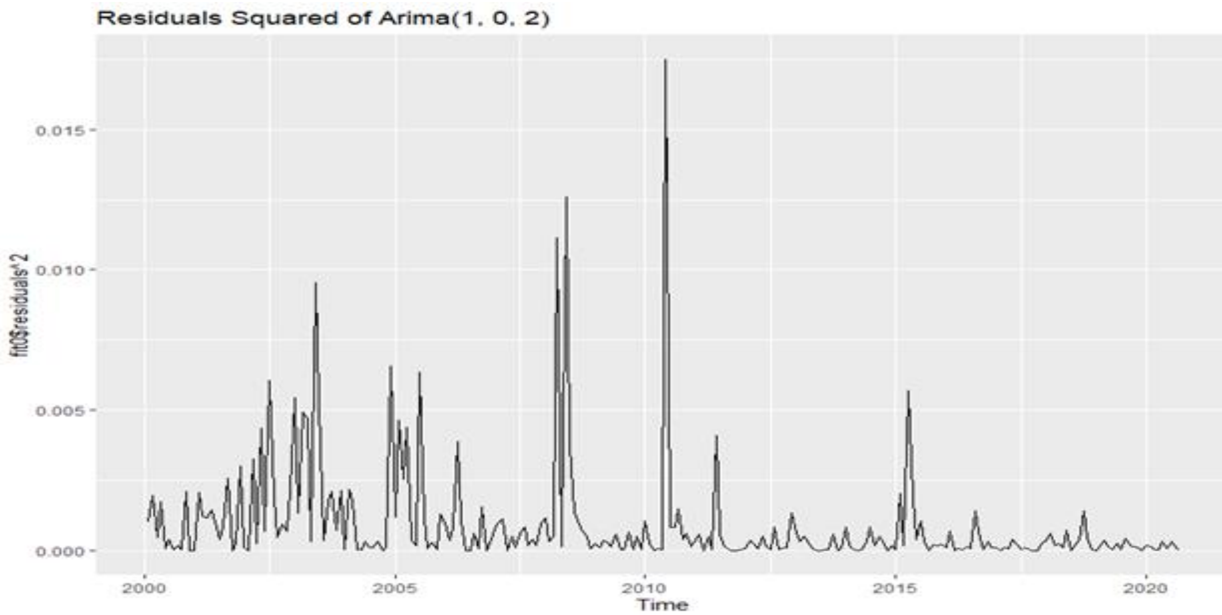
SQ 5:


Forecasts from ARIMA(1,0,2) with non-zero mean
0.10
0.05
0.00
-0.05
-0.10
r
2000
2005
2010
2015
2020
Time

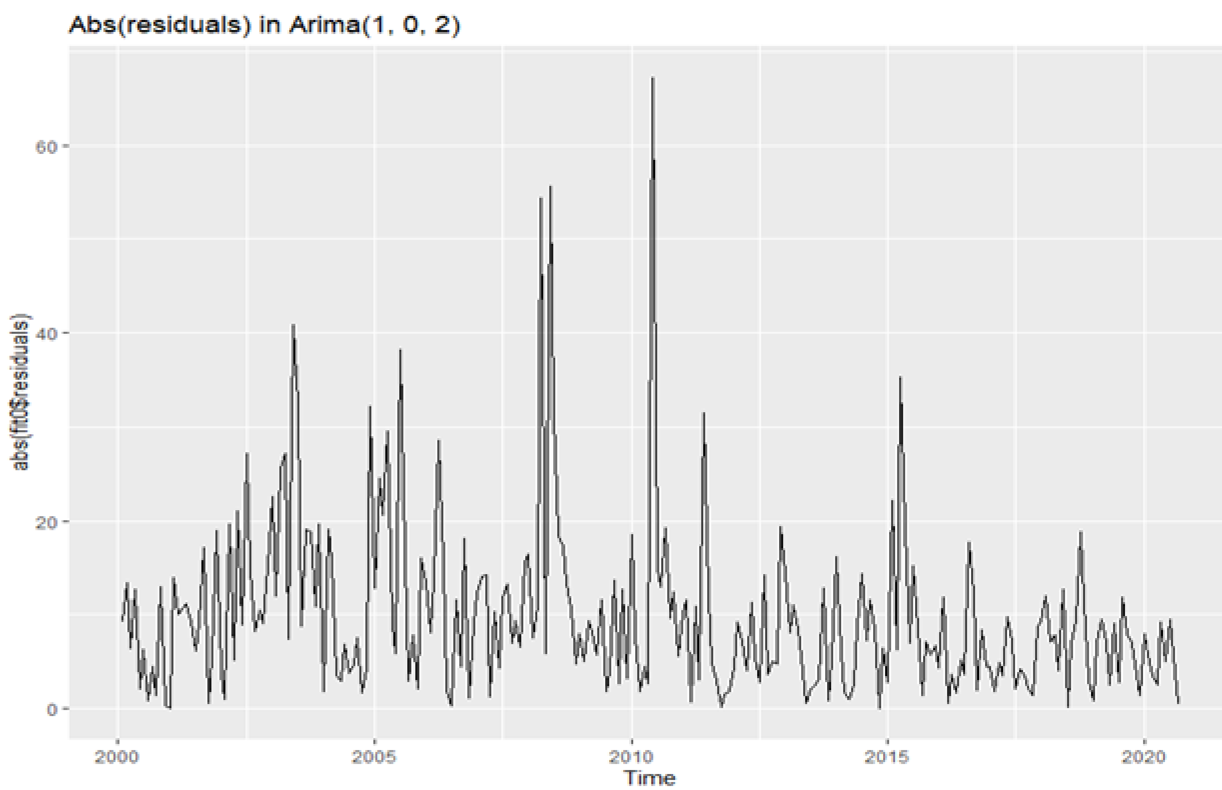
SQ 6:


Lag Plot of Arima(1, 0, 2) residuals
40
20
0
-20
-40
-60
fit0$residuals
-100
-50
0
50
lag 1

SQ 7:
Plot of Volatility of ARIMA(1,0,2) Squared Residuals



SQ 8:
Plot of change in Mean of ARIMA(1,0,2) Absolute Residuals



SQ 9:

```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(1, 2) + garch(1, 1), data = r, trace = F)

Mean and Variance Equation:
 data ~ arma(1, 2) + garch(1, 1)
<environment: 0x0000020751 8d0da8>
 [data = r]

Conditional Distribution:
 norm

Coefficient(s):
         mu          ar1          ma1          ma2        omega       alpha1        beta1
 6.2048e-04   6.4841e-01  -4.8848e-01  -2.0855e-01   1.0035e-09   4.0046e-02   9.5603e-01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      6.205e-04   5.570e-04    1.114 0.265272
ar1     6.484e-01   1.953e-01    3.321 0.000898 ***
ma1    -4.885e-01   2.006e-01   -2.435 0.014890 *
ma2    -2.085e-01   6.975e-02   -2.990 0.002791 **
omega   1.003e-09   8.711e-06    0.000 0.999908
alpha1  4.005e-02   1.953e-02    2.050 0.040336 *
beta1   9.560e-01   2.488e-02   38.421  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 534.7686    normalized:  2.156325
```

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  59.89769  9.847678e-14
 Shapiro-Wilk Test  R    W      0.9704405 5.038003e-05
 Ljung-Box Test     R    Q(10)  7.569952  0.6707625
 Ljung-Box Test     R    Q(15)  12.28995  0.6569625
 Ljung-Box Test     R    Q(20)  14.54304  0.8019308
 Ljung-Box Test     R^2  Q(10)  8.925977  0.5391428
 Ljung-Box Test     R^2  Q(15)  11.30157  0.730938
 Ljung-Box Test     R^2  Q(20)  13.06494  0.8745806
 LM Arch Test       R    TR^2   9.903434  0.6244321

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-4.256198 -4.157029 -4.257734 -4.216277
```

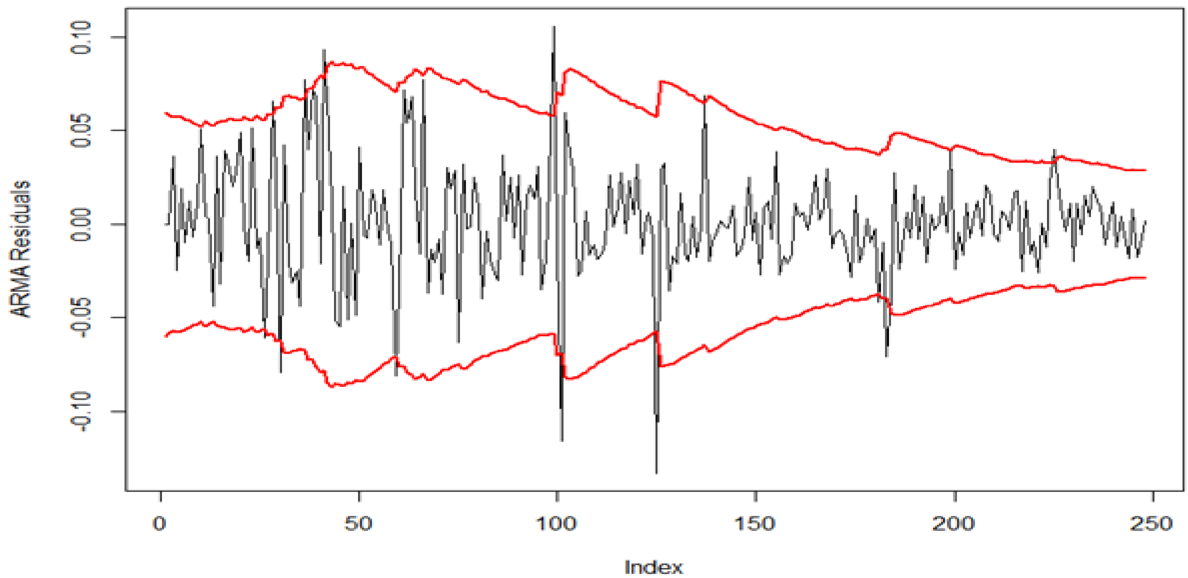
MODEL EQUATION:

return_t = 0.65_v_t-1 – 0.49_a_t-1 - 0.21_a_t-2 + a_t

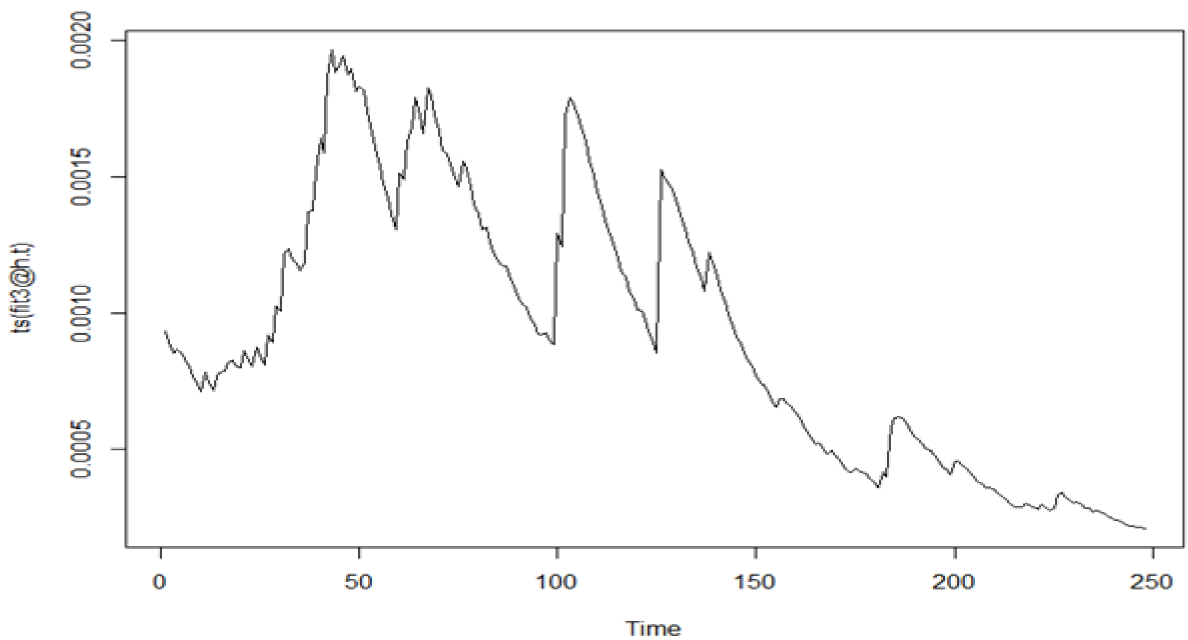
sigma^2 = 0.04_a^2_t-1 + 0.956_sigma^2_t-1

SQ 10:

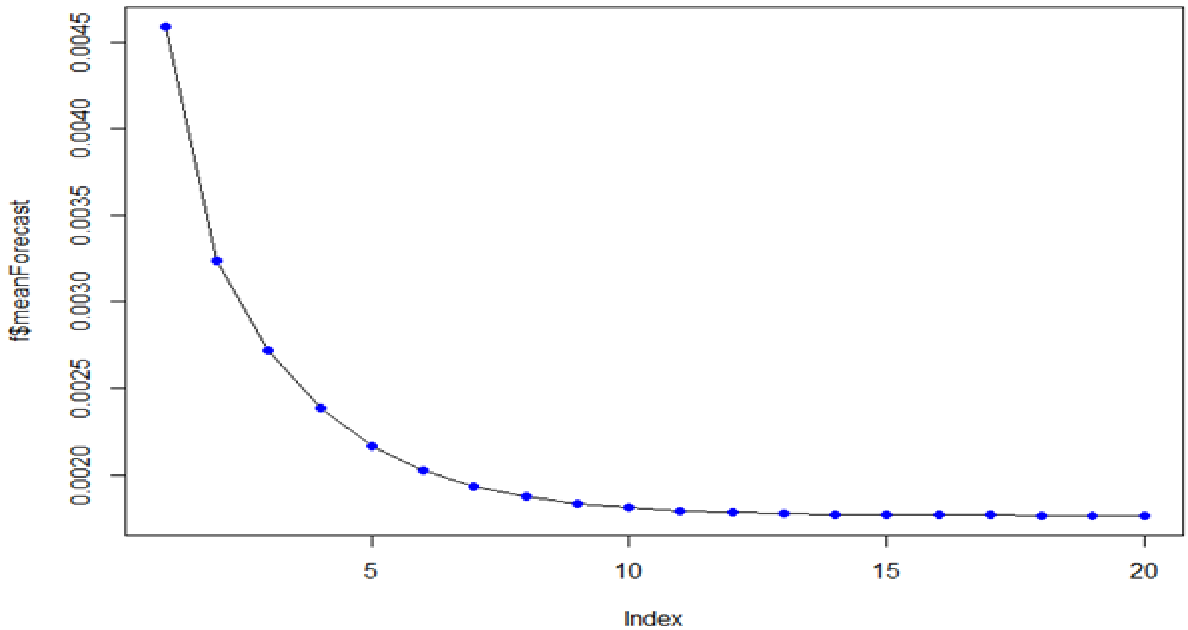
Plot of Garch Fit Residuals with 95% C.I.



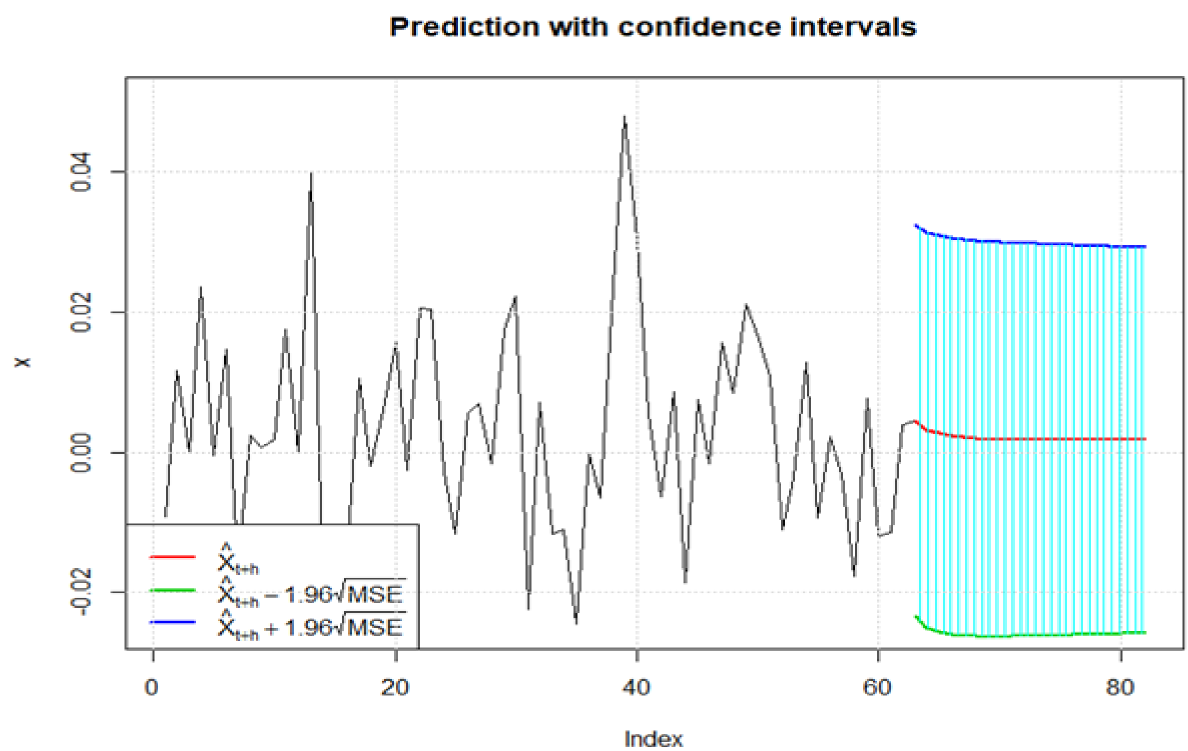
Plot of GARCH Fit Variance



SQ 11:



Forecast of Next Year



SQ 12:

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(1,0,2)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.001681    0.001078  1.559235 0.118941
ar1     0.820359    0.138961  5.903541 0.000000
ma1    -0.667750    0.146466 -4.559071 0.000005
ma2    -0.208640    0.065625 -3.179278 0.001476
omega   0.000000    0.000004  0.000002 0.999998
alpha1  0.038484    0.015794  2.436611 0.014826
beta1   0.956998    0.013902 68.837565 0.000000

Robust Standard Errors:
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.001681    0.001075  1.562842 0.118090
ar1     0.820359    0.126308  6.494918 0.000000
ma1    -0.667750    0.132418 -5.042747 0.000000
ma2    -0.208640    0.058382 -3.573675 0.000352
omega   0.000000    0.000004  0.000002 0.999998
alpha1  0.038484    0.024412  1.576437 0.114925
beta1   0.956998    0.020412 46.883382 0.000000

LogLikelihood : 533.9898

Information Criteria
------------------------------------

fAkaike       -4.2499
Bayes        -4.1507
Shibata      -4.2515
Hannan-Quinn -4.2100

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                  0.0001481  0.9903
Lag[2*(p+q)+(p+q)-1][8]  2.1888429  1.0000
Lag[4*(p+q)+(p+q)-1][14] 5.3767292  0.8503
d.o.f=3
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                       statistic p-value
Lag[1]                   0.03947 0.84252
Lag[2*(p+q)+(p+q)-1][5]  8.24114 0.02567
Lag[4*(p+q)+(p+q)-1][9]  9.49861 0.06406
d.o.f=2
```

```
Weighted ARCH LM Tests
------------------------------------
             Statistic Shape Scale P-Value
ARCH Lag[3]     0.3919 0.500 2.000  0.5313
ARCH Lag[5]     0.5123 1.440 1.667  0.8800
ARCH Lag[7]     0.7914 2.315 1.543  0.9449

Nyblom stability test
------------------------------------
Joint Statistic:  3.7088
Individual Statistics:
mu     0.1540
ar1    0.2366
ma1    0.1926
ma2    0.2249
omega  0.2422
alpha1 0.2472
beta1  0.1657

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.69 1.9 2.35
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias           1.5099 0.1324
Negative Sign Bias  0.4578 0.6475
Positive Sign Bias  1.0106 0.3132
Joint Effect        2.3665 0.4999


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     19.58       0.4202
2    30     34.58       0.2186
3    40     38.13       0.5094
4    50     46.76       0.5645


Elapsed time : 0.3882339
```

SQ 13:

```
> backtestGarch(fit3, r, testLen, 1) |
[1] "Testing 0 of 24\n"
[1] "Testing 10 of 24\n"
[1] "Testing 20 of 24\n"
[1] "RMSE of out-of-sample forecasts"
[1] 0.01526952
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.01239633
[1] "Mean Absolute Percentage error"
[1] 1.046861
[1] "Symmetric Mean Absolute Percentage error"
[1] 1.553998
```

```
GARCH Roll Mean Forecast Performance Measures
---------------------------------------------
Model          : sGARCH
No.Refits      : 10
No.Forecasts: 48

        Stats
MSE  0.0002105
MAE  0.0120600
DAC  0.5000000
```

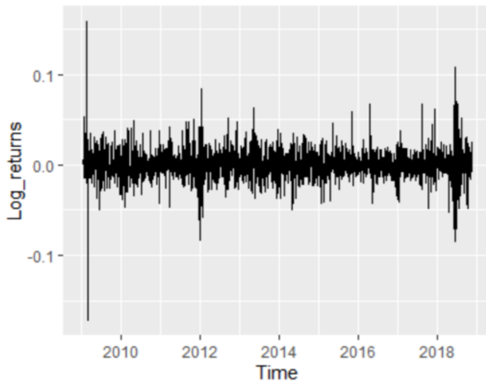
Model Equation:

return_t = 0.82_v_t-1 – 0.67_a_t-1 - 0.21_a_t-2 + a_t

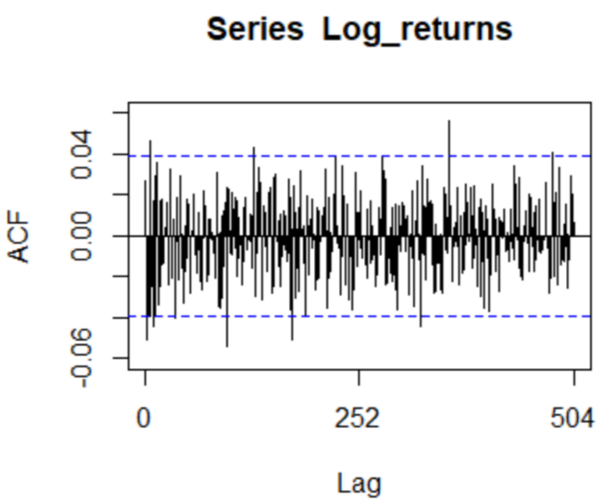
sigma^2 = 0.039_a^2_t-1 + 0.957_sigma^2_t-1

Dharun Selvan

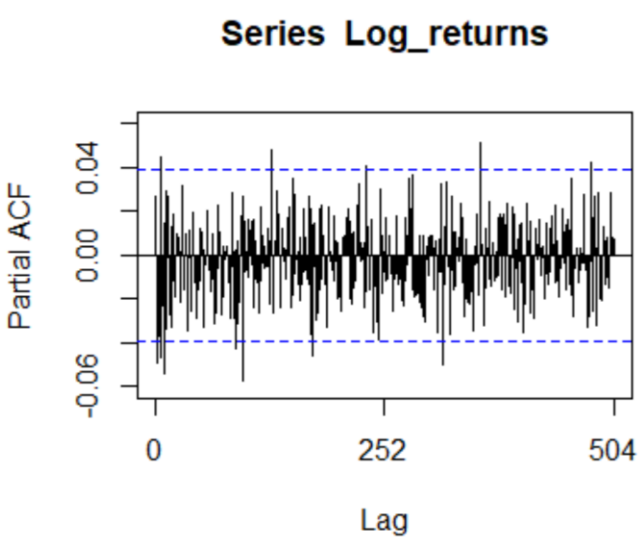
DS1:



DS2:



DS3:



DS4:

```
> eacf(Log_returns)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o x o x o o o o x o  o  o  o
1 x o o o o x o o o o x  o  o  o
2 x x o x o x o o o o o  o  o  o
3 x x x x x x o o o o o  o  o  o
4 x o x x o o o o o o o  o  o  o
5 x x x x x o o o o o o  o  o  o
6 x x x x x x o o o o o  o  o  o
7 x x o x x x x o o o o  o  o  o
```

DS5:

Series: Log_returns

ARIMA(1,0,1) with non-zero mean

Coefficients:

	ar1	ma1	mean
	-0.8549	0.8814	5e-04
s.e.	0.0919	0.0839	3e-04

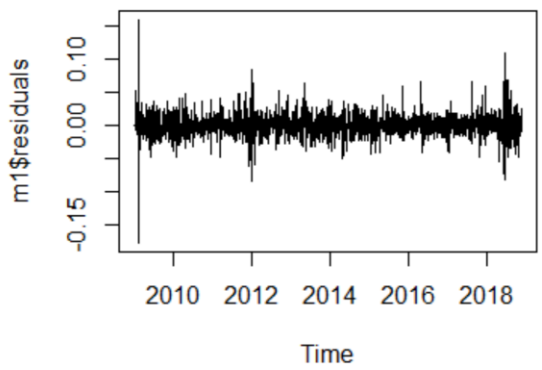
sigma^2 estimated as 0.0002566: log likelihood=6726.69

AIC=-13445.37   AICc=-13445.36   BIC=-13422.11

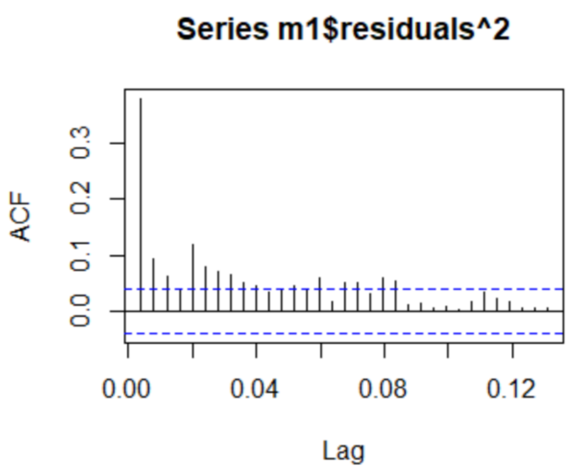
```
> coeftest(m1)

z test of coefficients:

             Estimate  Std. Error z value Pr(>|z|)    
ar1       -0.85486877  0.09194243 -9.2979   <2e-16 ***
ma1        0.88142075  0.08392576 10.5024   <2e-16 ***
intercept  0.00045064  0.00032689  1.3786    0.168    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

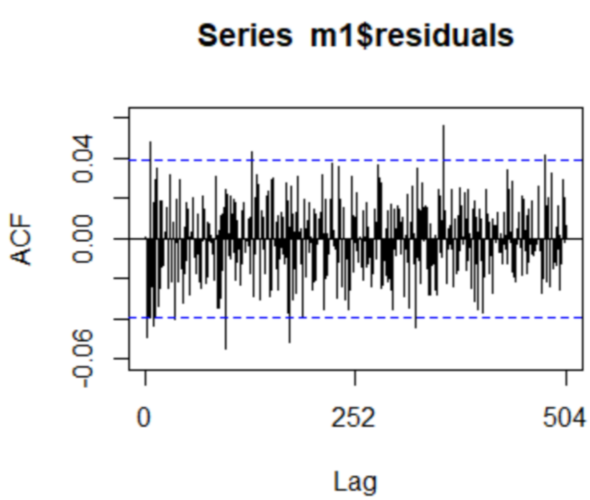
```
> adf.test(m1$residuals)

	Augmented Dickey-Fuller Test

data:  m1$residuals
Dickey-Fuller = -13.995, Lag order = 13, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(m1$residuals) : p-value smaller than printed p-value
> kpss.test(m1$residuals)

	KPSS Test for Level Stationarity

data:  m1$residuals
KPSS Level = 0.038829, Truncation lag parameter = 8, p-value = 0.1

Warning message:
In kpss.test(m1$residuals) : p-value greater than printed p-value
```

DS6:


m1$residuals
-0.15
0.00
0.10
2010
2012
2014
2016
2018
Time

DS7:

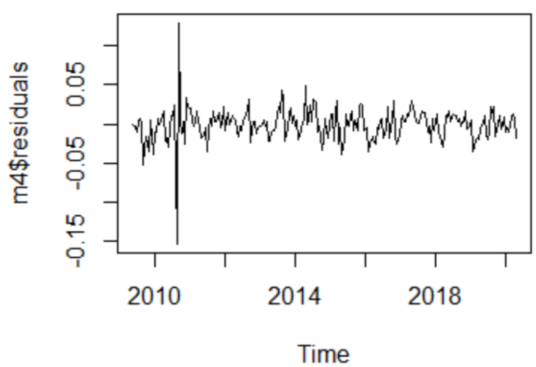

Series m1$residuals^2
ACF
0.0
0.1
0.2
0.3
0.00
0.04
0.08
0.12
Lag


Series m1$residuals
ACF
-0.06
0.00
0.04
0
252
504
Lag

DS8:
[1] "RMSE of out-of-sample forecasts" 0.02068562
[1] "Mean absolute error of out-of-sample forecasts" 0.01708682
[1] "Mean Absolute Percentage error" 1.396245
[1] "Symmetric Mean Absolute Percentage error" 1.462866
DS9:
Series: Log_returns
ARIMA(0,0,1)(2,0,0)[21] with zero mean
Coefficients:

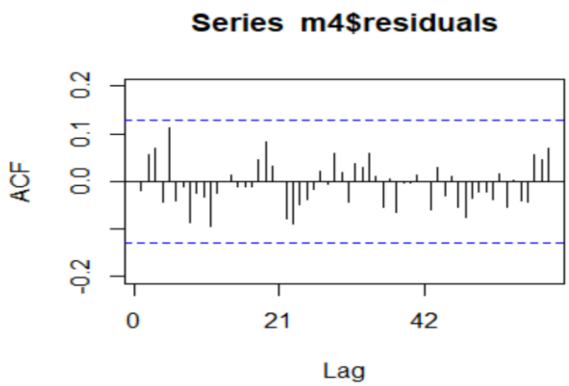
	ma1	sar1	sar2
	-0.2436	-0.1147	-0.0177
s.e.	0.0593	0.0667	0.0877

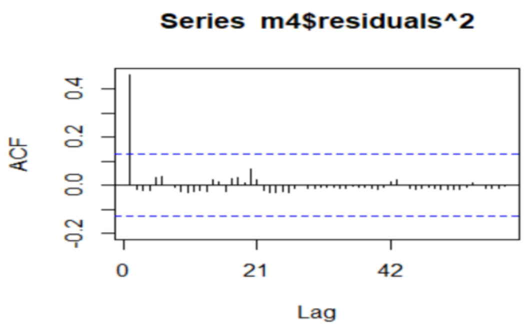
sigma^2 estimated as 0.0004407: log likelihood=563.62
AIC=-1119.23   AICc=-1119.05   BIC=-1105.48
DS10:




```
> adf.test(m4$residuals)

        Augmented Dickey-Fuller Test

data:  m4$residuals
Dickey-Fuller = -5.3321, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(m4$residuals) : p-value smaller than printed p-value
> kpss.test(m4$residuals)

        KPSS Test for Level Stationarity

data:  m4$residuals
KPSS Level = 0.095472, Truncation lag parameter = 4, p-value = 0.1

Warning message:
In kpss.test(m4$residuals) : p-value greater than printed p-value
```

DS11:





DS12:

[1] "RMSE of out-of-sample forecasts" 0.02134199

[1] "Mean absolute error of out-of-sample forecasts" 0.01781447

[1] "Mean Absolute Percentage error" 1.355788

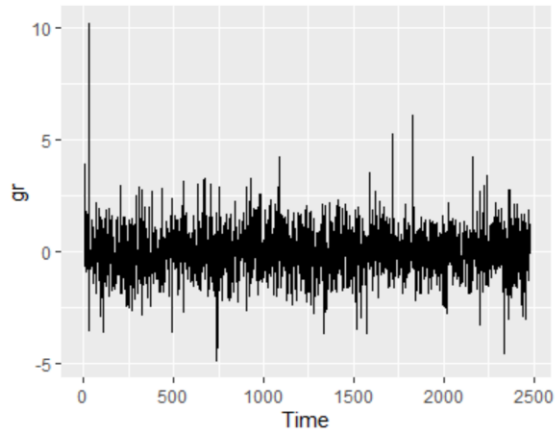
[1] "Symmetric Mean Absolute Percentage error" 1.538083

DS13:

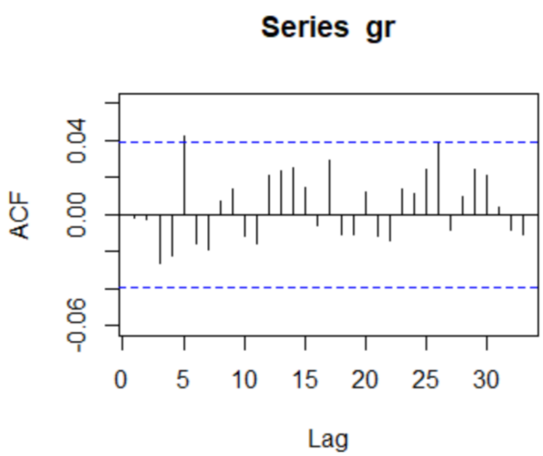
```
Call:
 garchFit(formula = ~arma(1, 2) + garch(1, 1), data = Log_returns,
    trace = F)

Mean and Variance Equation:
 data ~ arma(1, 2) + garch(1, 1)
<environment: 0x000002818f808990>
 [data = Log_returns]

Conditional Distribution:
 norm

Coefficient(s):
         mu          ar1          ma1          ma2        omega       alpha1        beta1
 3.5929e-05   9.4692e-01  -8.4884e-01  -1.1760e-01   1.3751e-05   1.0009e-01   8.4413e-01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      3.593e-05   3.122e-05    1.151     0.25
ar1     9.469e-01   4.005e-02   23.646  < 2e-16 ***
ma1    -8.488e-01   4.389e-02  -19.338  < 2e-16 ***
ma2    -1.176e-01   2.282e-02   -5.154 2.55e-07 ***
omega   1.375e-05   2.881e-06    4.774 1.81e-06 ***
alpha1  1.001e-01   1.419e-02    7.055 1.73e-12 ***
beta1   8.441e-01   2.180e-02   38.715  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

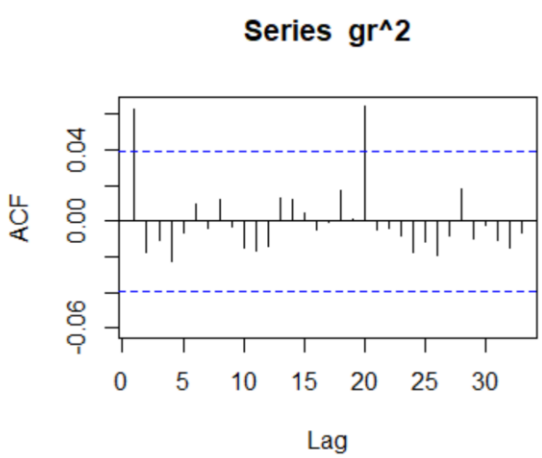
Log Likelihood: 6963.567   normalized: 2.811291

DS14:

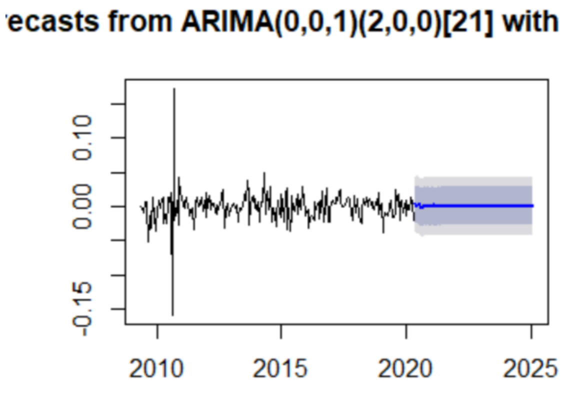

10
5
0
-5
gr
0
500
1000
1500
2000
2500
Time

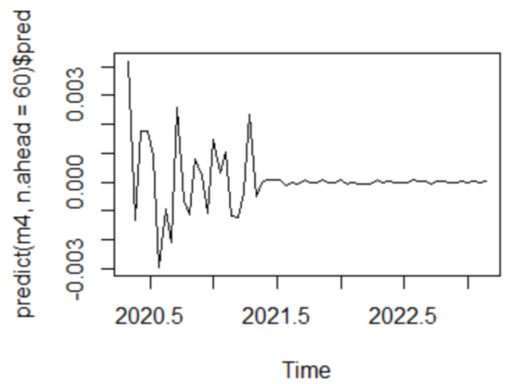
DS15:


Series gr
ACF
0.04
0.00
-0.06
0
5
10
15
20
25
30
Lag


Series gr^2
ACF
0.04
0.00
-0.06
0
5
10
15
20
25
30
Lag

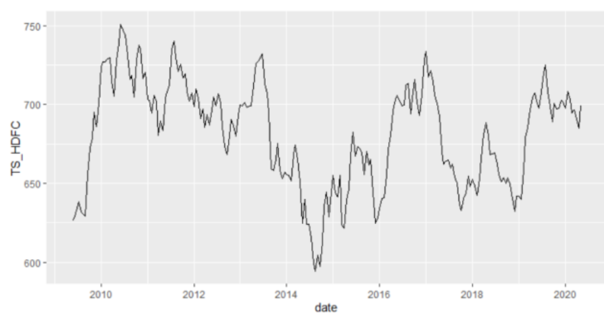
DS16:



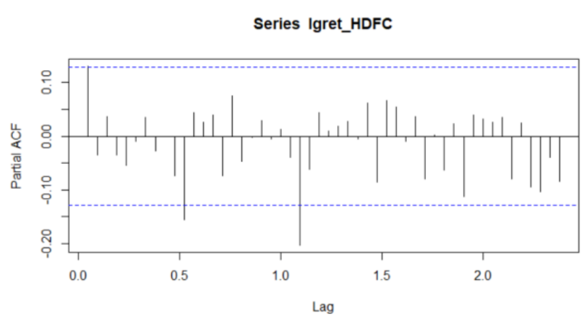


Shadhana Palaniswami(SP)

SP 1



SP 2



SP 3:

```
> eacf(lgret_HDFC)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o o o o x  o  o  o
1 x o o o o o o o o o x  o  o  o
2 x o o o o o o o o o x  o  o  o
3 x o x o o o o o o o x  o  o  o
4 x x o o o o o o o o x  o  o  o
5 x x o x o o o o o o x  o  o  o
6 x x o x o x o o o o x  o  o  o
7 x x o o o x o o o o x  o  o  o
```

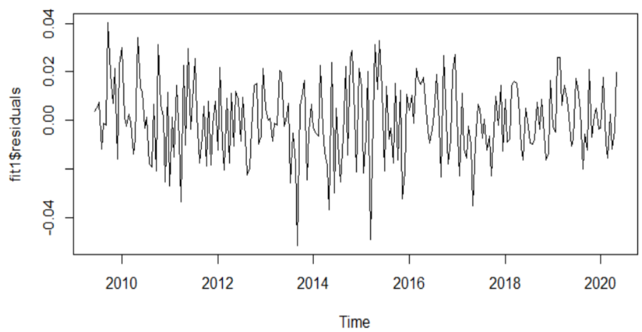
SP 4:

```
> fit1 = Arima(lgret_HDFC, order = c(0,0,1), seasonal = list(order = c(1,0,0), period=11))
> fit1
Series: lgret_HDFC
ARIMA(0,0,1)(1,0,0)[11] with non-zero mean

Coefficients:
         ma1      sar1    mean
      0.1352   -0.1720   4e-04
s.e.  0.0675    0.0657   1e-03

sigma^2 estimated as 0.0002555:  log likelihood=626.28
AIC=-1244.56   AICc=-1244.38   BIC=-1230.8
> coeftest(fit1)

z test of coefficients:

             Estimate   Std. Error z value Pr(>|z|)
ma1        0.13522664  0.06753706  2.0023  0.04526 *
sar1      -0.17196106  0.06569905 -2.6174  0.00886 **
intercept  0.00043921  0.00102330  0.4292  0.66777
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

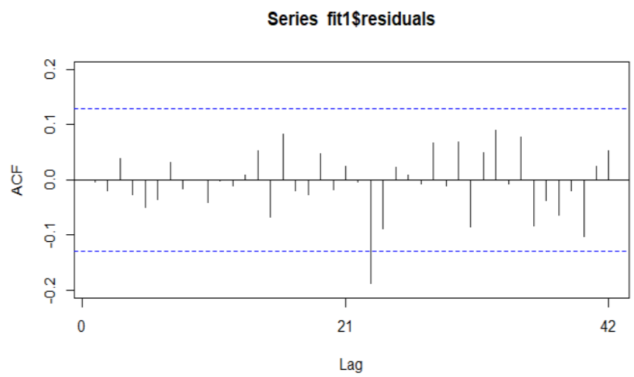
SP 5:

Residual Analysis

Autoplot



ACF:



Dickey Fuller Test :

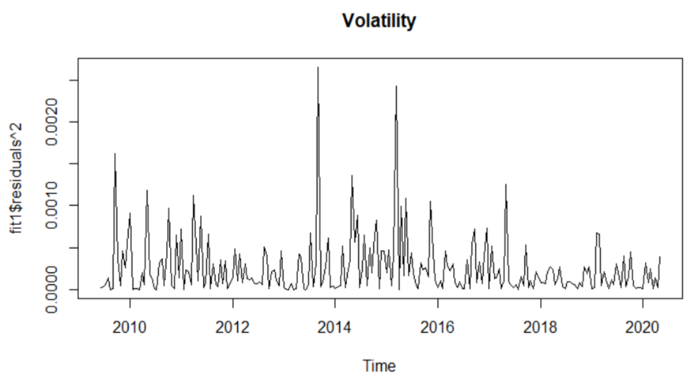
```
> Box.test(fit1$residuals, lag=10, type="Ljung")

        Box-Ljung test

data:  fit1$residuals
X-squared = 2.2385, df = 10, p-value = 0.9942
```

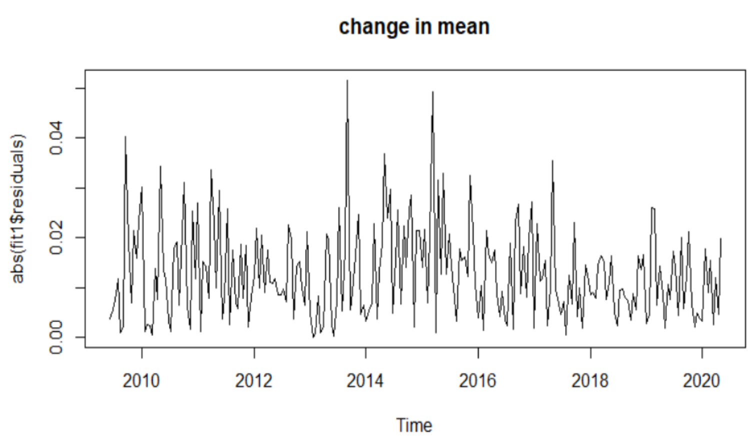
KPSS TEST :

```
> kpss.test(fit1$residuals)
KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
 lag   stat p.value
   3 0.0815     0.1
-----
 Type 2: with drift no trend
 lag   stat p.value
   3 0.0811     0.1
-----
 Type 1: with drift and trend
 lag   stat p.value
   3 0.0776     0.1
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
> 
```

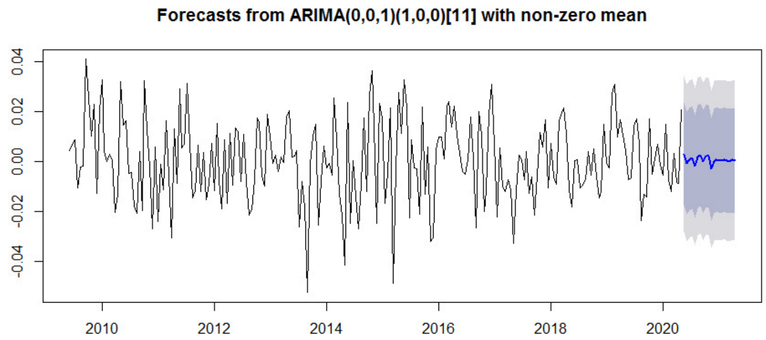
SP 6:

Volatility and Change of mean:





SP 7:
Forecast of ARIMA

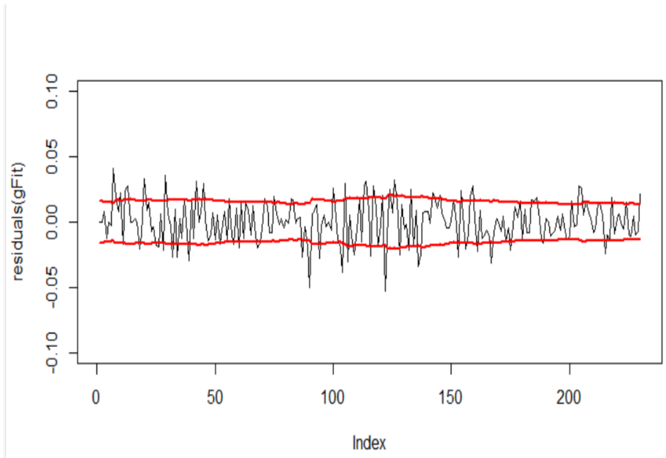


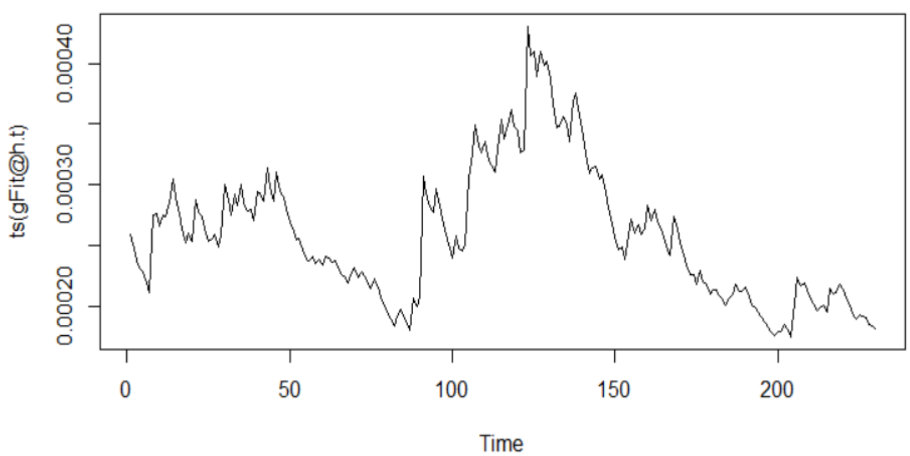
SP 8:
GARCH MODEL :

```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(0, 2) + garch(1, 1), data = lgret_HDFC,
    trace = F)

Mean and Variance Equation:
 data ~ arma(0, 2) + garch(1, 1)
<environment: 0x000001b571b9ae98>
 [data = lgret_HDFC]

Conditional Distribution:
 norm

Coefficient(s):
         mu          ma1          ma2        omega       alpha1        beta1
 4.2147e-04   1.6315e-01  -2.3033e-02   9.2184e-06   4.2332e-02   9.2064e-01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      4.215e-04   1.169e-03    0.360   0.7185
ma1     1.632e-01   6.841e-02    2.385   0.0171 *
ma2    -2.303e-02   6.839e-02   -0.337   0.7363
omega   9.218e-06   9.373e-06    0.984   0.3253
alpha1  4.233e-02   2.721e-02    1.556   0.1197
beta1   9.206e-01   4.936e-02   18.652   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 625.5152    normalized:  2.719631
```

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  0.6862835 0.7095376
 Shapiro-Wilk Test  R    W      0.994091  0.5037005
 Ljung-Box Test     R    Q(10)  2.033117  0.99608
 Ljung-Box Test     R    Q(15)  9.851242  0.8289914
 Ljung-Box Test     R    Q(20)  12.03206  0.914968
 Ljung-Box Test     R^2  Q(10)  6.400493  0.7805686
 Ljung-Box Test     R^2  Q(15)  12.25601  0.6595567
 Ljung-Box Test     R^2  Q(20)  15.84416  0.7262437
 LM Arch Test       R    TR^2   6.913978  0.8632455

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.387089 -5.297400 -5.388404 -5.350910
```

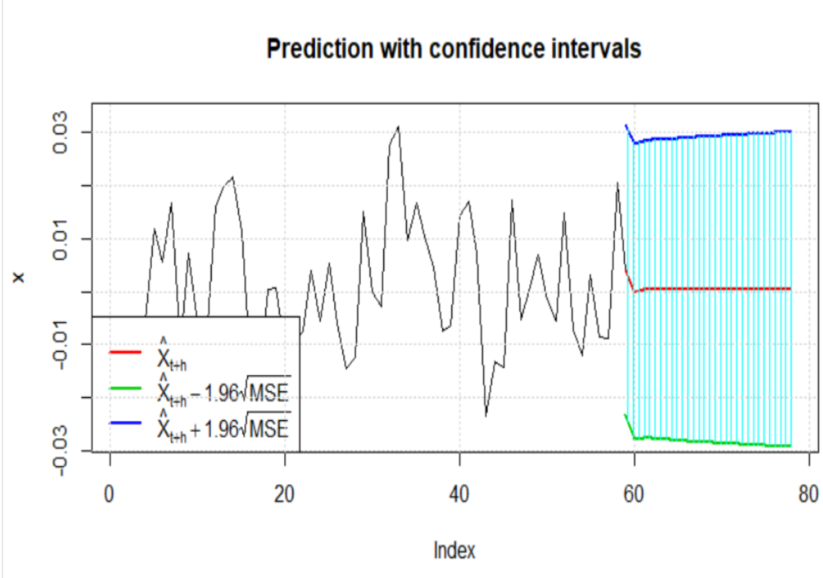
SP 9:
RESIDUALS OF GARCH AND VARIANCE PLOT





SP 10:

Forecast Plot :



R CODE

Syed Qavi:

```
library(ggplot2) #ggplot2 Functions
library(ggfortify)
library(lubridate)
library(tseries)
library(forecast)
library(lmtest)
library(fGarch)
source("eacf.R")
source("backtest.R")
source("backtestGarch.R")

#READ DATASET
hdfc <- read.csv(file="HDFC.csv", header=TRUE, sep=",")

dim(hdfc)
head(hdfc)
names(hdfc)
class(hdfc$Date)
hdfc$Date <- ymd(hdfc$Date)
hdfc
# qqplot
qqnorm(hdfc$Close)
qqline(hdfc$Close, col="red", lw=2)
#histogram
hist(hdfc$Close, xlab="HDFC Stock Close Price", prob=TRUE, main="Histogram")
xfit = seq(min(hdfc$Close),max(hdfc$Close),length=40)
yfit = dnorm(xfit,mean=mean(hdfc$Close),sd=sd(hdfc$Close))
lines(xfit, yfit, col="blue", lwd=2)

qplot(hdfc$Date, hdfc$Close, geom="line", main="HDFC", ylab="Price ($)", xlab="Date") #too
complicated adjust to 12 frequency

# CREATE Time Series
hdfcTS = ts(log(hdfc$Close), start = c(2000, 1, 3), end = c(2020, 9, 30), frequency=12)
hdfcTS
autoplot(hdfcTS, main="Time Plot of HDFC Close Price", ylab="Close Price", xlab="Date") #take
log since multiplicative
autoplot(log(hdfcTS), main="Time Plot of HDFC Close Log Price", ylab="ln(Close Price)",
xlab="Date") #for log return

r = diff(hdfcTS)  #this gets rid of unit root
autoplot(r, xlab="Date", ylab="Close Montly Return", main="HDFC Monthly Log-Returns")
```

```
#BOX COX Transformation
library(lmSupport)
library(car)
fit = lm(hdfcTS ~ time(hdfcTS))
plot(fit$residuals ~ fit$fitted.values)

fitLog = lm(log(hdfcTS) ~ time(hdfcTS))
plot(fitLog)

b = modelBoxCox(fit)
#library(car) needed with box cox

lambda = 2
yBox = (hdfcTS^lambda - 1) / lambda

autoplot(yBox)
fit = lm(yBox ~ time(hdfcTS))
plot(fit)
plot(fit$residuals ~ fit$fitted.values)

Acf(yBox, main = "ACF of Transformed TS")
pacf(yBox, main = "PACF of Transformed TS")
eacf(yBox)

adfTest(yBox)
adfTest(yBox, type="ct") #FAIL TO reject rw w/ drift

kpss.test(yBox) #reject trend stat

# exploring returns
qqnorm(r)
qqline(r, col="red", lw=2)

hist(r, xlab="HDFC Stock Close Price", prob=TRUE, main="Histogram")

jarque.bera.test(r)

Acf(r, main = "ACF of Returns", lag.max = 100)
Box.test(r, lag=10, type="Ljung")

#Checking returns acf, pacf, eacf
Box.test(r, lag = 5, type="Ljung")

par(mfrow=c(1,2))
Acf(r, main = "ACF of Log Return")
pacf(r, main = "PACF of Log Return")
eacf(r)
```

```
# Fitting ARIMA Model
fit1 = auto.arima(r)
fit1        #2,0,3
coeftest(fit1)
autoplot(fit1$residuals)       #check for heteroscedacity if yes then garch
Acf(fit1$residuals)
Box.test(fit1$residuals, lag=10, type = "Ljung-Box")

fit0 = Arima(r, order=c(1, 0, 2))  #better than auto
fit0
coeftest(fit0)
autoplot(fit0$residuals)
Acf(fit0$residuals)
Box.test(fit0$residuals, lag=10, type = "Ljung-Box")

fit2 = Arima(r, order=c(0, 0, 2), include.drift = T)
fit2
coeftest(fit2)
autoplot(fit2$residuals)
Acf(fit2$residuals)
Box.test(fit2$residuals, lag=10, type = "Ljung-Box")

#forecast
autoplot(forecast(fit0), h = 20)

#backtest for fit 1 and 0
backtest(fit1, r, h=1, orig=.8*length(r))
backtest(fit0, r, orig=.8*length(r), h=1)
backtest(fit2, r, orig=.8*length(r), h=1)

#test for arch effects
autoplot(fit0$residuals, main="Residuals of Arima(1, 0, 2) model")
lag.plot(fit0$residuals, main="Lag Plot of Arima(1, 0, 2) residuals", pch=19, cex=.1)

autoplot(fit0$residuals^2, main="Residuals Squared of Arima(1, 0, 2)")
Acf(fit0$residuals^2,lag.max = 100)
autoplot(abs(fit0$residuals), main="Abs(residuals) in Arima(1, 0, 2)")

#do a t test for change in mean
t.test(r)

# Change in overall mean is clearer, but peaks and spikes in variance are not
autoplot(abs(fit0$residuals))
acf(abs(fit0$residuals))
```

```
#garch
fit2=garchFit(~arma(0,2)+garch(1,1),data=r,trace=F)
summary(fit2)
fit2

fit3=garchFit(~arma(1,2)+garch(1,1),data=r,trace=F)
fit3
summary(fit3)

gRes2 = ts(residuals(fit3, standardize=T))
autoplot(gRes2)
jarque.bera.test(gRes2)

plot(residuals(fit3)^2, type="l", ylim=c(0, .01))
lines(fit3@h.t, col="red")

plot(residuals(fit3), type="l", ylab="ARMA Residuals")
lines(1.96 * sqrt(fit3@h.t), col="red", lw=2)
lines(-1.96 * sqrt(fit3@h.t), col="red", lw=2)

plot(ts(fit3@h.t))

f = predict(fit3, n.ahead=20, plot = T)
plot(f$standardDeviation, type = "l")
plot(f$meanForecast, type="l")
points(f$meanForecast, col="blue", pch=16)

Acf(gRes2)          # Higher order autocorrelation, but fairly minor
Acf(gRes2^2)

qqnorm(gRes2)
qqline(gRes2, col="red")

#ugarch
library(rugarch)
res = r
gFit3.spec = ugarchspec(variance.model=list(garchOrder=c(1,1)),
                        mean.model=list(armaOrder=c(1, 2)))

gFit3 = ugarchfit(spec=gFit3.spec, data=res)
gFit3
coeftest(gFit3)

gRes3 = residuals(gFit3, standardize=T)
autoplot(gRes3)
jarque.bera.test(gRes3)
```

```
#skewness(gRes2)  #information criteria down = good
#kurtosis(gRes2)
gFit3.rolltest = ugarchroll(gFit3.spec, data = res, n.start = 200, refit.window = "moving", refit.every
= 5)
gFit3.rolltest
report(gFit3.rolltest, type = "fpm")

#backtest GARCH

length(r)
testLen = floor(length(r) * .90)   # Train on 98% of thes
testLen

backtestGarch(fit3, r, testLen, 1)
```

Shadhana Palaniswami

```
# ---------------------------------- Visualizing time series ---------------------------------
class(HDFC$Date)
HDFC$Date = as.Date(HDFC$Date, format="%d-%m-%y")
class(HDFC$Date)
head(HDFC)

TS_HDFC = ts((HDFC$VWAP), start=c(2009, 9),end =c(2020,8),frequency = 21)
autoplot(TS_HDFC,xlab = "date")
HDFClg=log(TS_HDFC)
lgret_HDFC=diff(HDFClg)
autoplot(lgret_HDFC,xlab="Date", ylab="")
qqnorm(lgret_HDFC)
#---------------------ACF PLOT-------------
Acf(lgret_HDFC,lag.max = 50)
pacf(lgret_HDFC,lag.max = 50)
eacf(lgret_HDFC)
adfTest(lgret_HDFC)
kpss.test(lgret_HDFC)
#----------------------building ARIMA model----------
fit1 = Arima(lgret_HDFC, order = c(0,0,1), seasonal = list(order = c(1,0,0), period=11))
summary(fit1)
coeftest(fit1)
fit2 = Arima(lgret_HDFC, order = c(0,0,1))
fit2
coeftest(fit2)

plot(fit1$residuals)
Acf(fit1$residuals)

adfTest(fit1$residuals)
kpss.test(fit1$residuals)
kpss.test(lgret_HDFC)

Box.test(fit1$residuals, lag=10, type="Ljung")

fit5 = auto.arima(lgret_HDFC)
fit5
coeftest(fit5)

plot(fit5$residuals)
Acf(fit5$residuals)

adf.test(fit5$residuals)
kpss.test(fit5$residuals)
library(fGarch)
```

```
summary(gFit <- garchFit(~ arma(0,2) + garch(1,1), data = lgret_HDFC, trace = F))
gFit
gRes2 = ts(residuals(gFit, standardize=T)) # Standardize to get garch residuals
autoplot(gRes2)

plot(residuals(gFit), type="l", ylim=c(-.5, .5), ylab="ARMA Residuals")
lines( sqrt(gFit@h.t), col="red", lw=3)
lines(sqrt(gFit@h.t), col="red", lw=3)

plot(ts(gFit@h.t))
Box.test(gRes2, lag=20, type="Ljung")
qqnorm(gRes2)
qqline(gRes2)
jarque.bera.test(gRes2)
adfTest(gRes2, lags=20, type='ct')
kpss.test(gRes2)

f = predict(gFit, n.ahead=36)
plot(f$meanForecast, type="l")
points(f$meanForecast, col="blue", pch=16)
f = predict(gFit, n.ahead=20, plot=T)
n = length(lgret_HDFC)
b1 = backtest(fit1, lgret_HDFC, h=1, orig=.9*n)
b2 = backtest(fit5, lgret_HDFC, h=1, orig=.9*n)
b3 = backtest(gFit, log_returns, h=1, orig = .9*n)
plot(residuals(gFit)^2, type="l")
lines(gFit@h.t, col="red")
plot(residuals(gFit), type="l", ylim=c(-.1, .1))
lines(sqrt(gFit@h.t), col="red", lw=2)
lines(-sqrt(gFit@h.t), col="red", lw=2)
f = forecast(fit1)
plot(f)
```

Individual report: Vallabha Datta Penmetcha

I worked on TCS stock market. After doing some background research about the variables I decided to analyze the daily percentage deliverables of TCS stock market. It indicates the % of shares that are marked for delivery. If % deliverable is high it indicates that the shares are interested by the investors to buy and hold it so, it is one of the important factors that is used to analyse the stocks. Firstly, I performed data preprocessing to check the correctness of the columns and analyze the distribution of percentage deliverable using histogram. During the data cleaning stage, I excluded the trader's column from the dataset because it has about 50% of null values. After data cleaning I created a time series for % deliverables and analyzed it's distribution using time series and decomposition. The series looks like trend stationary so, performed unit root test, autocorrelation test and normality test for the series to confirm the trend and autocorrelation. Since the series is trend stationary I did not perform first difference and log returns because it leads to overdifferencing.

In the modeling stage firstly, I fit a regression model using time as the independent variable. After fitting the regression model, the model is evaluated by goodness of fit and residual analysis. The residual analysis of the regression model shows that the residuals do not exhibit white noise behaviour. The series has autocorrelation and seems to have ARIMA behaviour, so I fit a model2 with ARIMA(1,0,2). The order of the model is chosen by analyzing acf, pacf and eacf plots. The residual analysis of model2 indicates that the residuals show white noise behaviour and it was a good fit. The residuals of the fitted model show ARCH effect hence, I fit model3 using GARCH(1,1) with ARIMA(1,1). The order of the GARCH model is selected by analyzing acf, pacf and eacf plots.

After fitting all the models, the model performance of all the models are evaluated by performing residual analysis and goodness of fit. Of all the three models, GARCH has less AIC, BIC values and high likelihood ratio, so I selected GARCH as the final model and performed forecasting for the next 20 months. Finally, analyzed the forecast and concluded that the volatility of the percentage deliverables will persist for a short duration after reaching its peak. Since the volatility persists for a short time it is better to buy the stocks after when it returns to normal.

I was always interested in analyzing the data of the financial sector and this interest made me take a time series analysis course. Time series forecasting plays a major part in most of the domains. I feel that the course was well organized. Apart from the lectures, the example files helped a lot to understand the subject. The assignments were not only useful to exercise the classroom concepts but also to do the real world project. All the topics were new and refreshing for me. Overall, the class was very informative and had a lot of fun. It was a good learning experience.

Individual Report: Milin Desai

TCS (Tata Consultancy Services) is the one of the giant tech companies based in India which is founded by leading business tycoon Ratan Tata. Also, it is one of the leading IT solution provider companies in the world. Here, we are analyzing the stock data and how it evolved over the years. And a journey from zero to hero of the company through pattern in stock price. Mainly we are focusing on the closing price.

First step I started with exploring the data knowing what is in there and how we can leverage it.

I addressed the basic problem like is there any missing value in the data, or how many NA available are there. How we can avoid in the way that it won't affect the quality.

Then I analyzed the variable and moved forward working with the closing price of the TCS stocks, I've plot histograms and tried so it was bi-,model so I used the first difference. In addition, I got to know that in the data there are big jumps because of the splitting of the stocks so many people can buy the stocks. We didn't have an idea about this but meeting with Dr. McDonald is enlightened about this. I love this thing about him. You will always learn something new and something valuable when you speak with him. that is one of the reasons for taking another class with him.

Then I tried to fit the model, checked skewness, kurtosis and jarque.bera test also before that checked a box test. Try decomposing graphs so we can know if there are any hidden patterns in the data. Randomness, seasonal component likewise.

I attempted fitting models using acf, eacf, pacf found the best order for the model and found the best model using ARIMA and also cross check with the AUTO ARIMA. P values were not significant so reject the null hypothesis. Then run a back test, back testing and then the forecasting of the variable.

Stocks data always fascinated me and when I took this subject that was the first thing I wanted to apply time series concepts to the stocks data and wanted to see how I can curtail it and build a model with this. So I started my quest for finding the best dataset for time series.
I love Dr. McDonald's teaching style and also the way he gets creative so that more students can interact with them. Course structure is really organized and easy to access and you will receive future resources for outside class. I've learned a lot of new things . Overall all it was challenging but total brain

Individual Report: Syed Qavi

My group worked on two different datasets and I was assigned to the HDFC stock dataset. I specifically looked at three variables and decided to focus on the Close price of stock. First, I analyzed this variable and fitted the variable vs time, transformed it using Box Cox transformation and got the best ARIMA model by analyzing the ACF, PACF, eacf and the residuals for white noise then backtesting the different models to get the best fit model. I did a forecast of the ARIMA model to see the forecast behavior. Then I transformed the original series into a log return so I can model the volatility of the returns. I fitted the a GARCH(1,1) model with the ARIMA model that I had previously calculated, then I analyzed the alpha and beta parameters or the lagged squared residuals and the lagged variance. I plotted different residual plots of the GARCH model and forecast along with the backtest to see if the GARCH model is a good fit. Then I fitted a ugarch model with the specification to see if I can improve the model in terms of goodness of fit. Also ran the backtest version of the ugarch to see if it perfomed better than the standard garch fit model. I want to try other GARCH model as well. So I performed an iGARCH model since the alpha and beta were close to so I want to see if I can improve the GARCH model by forcing the alpha and beta to equal 1 but it didn't improve the model so it was not added to the report; it was similar to the ugarch model. Also, I want to try fitting the model with an eGARCH model and base the calculation on the prior standardized shocks to see if I can improve the model, which also didn't improve the model too much to be included in the report.

This was the first time I did a deep analysis of a stock dataset and I learned a lot from the time series datasets. First, I noticed that the time plot was always non stationary and we had to convert to log return to understand what was going on. I also learned that the log returns and the residuals always have volatility and GARCH would be helpful to analyze it. I knew that it was hard to predict stocks and if some came up with a way to predict them accurately then they would be rich but I hadn't realized how difficult it would be until I started doing the forecasts and noticed that the forecasts don't really tell you a whole lot. But they do give some idea for only a short period of time and after that you can see how unpredictable the stock dataset can be. I also noticed that the forecast cannot be trusted and even if you do get a forecast, the market can suddenly crash, or some unforeseeable event can cause the market to act in a volatile way. This is best seen in the time plot of the stock price. For example, the HDFC stock price was following a trend of some sort which looked like a random walk with drift then suddenly, the market crashes, causing the prices to plummet. After the crash, the prices rose steadily from the spot where the crash ended. I guess there is a ceiling to how high the stock price can go.

I also learned a lot about data analysis and how it is a process that requires you to be thorough in analyzing the different tests. Sometimes one test can tell you this is a good model but another test can tell you similar, so you have to be careful to pick the best model. ARIMA models can sometimes be tricky since you can have two models who are different but when checking the goodness of fit can be similar but you can't tell which one is better until you do some additional tests such as back test and forecast. I also noticed that GARCH models can be very complex and hard to understand sometimes. The primary purpose of the GARCH model is just to model the volatility

Individual Report :Shadhana Palaniswami

I worked on the HDFC stock data .My variable of interest was VWAP,since it is used a tool to confirm the trend of the stock .Initial exploratory analysis showed that there are no missing values or values with values that had 0 as the value. But we noticed that the time series formed without transformation was stationary with no trend. I applied log returns transformation to obtain a uniformly distributed time series. The time series has monthly frequency.The following is the plot received when I tried to plot the time series of the HDFC weighted Average Variable.We can see that the time series is stationary and is normally distributed after applying log returns to the data set.I first plotted the ACF,PACF and the EACF plots to determine the order of the ARIMA model I was going to build.Looking at the plots above we can definitely say there is seasonality in the series ,after lag 10 .It looks like the seasonality appears at the lag 11.The seasonality appears to be weak here .This is not going to affect the trading on a larger scale and we cannot count on this for a opportunity for trading here. The best ARIMA model that I was able to built ,which also could capture the seasonality had the order(1,0,0) with just one MA component and seasonality order of (0,0,1).The auto-ARIMA gave me a model of order(0,0,1)and seasonality order(2,0,0),but it did not pass the coefficient test and also filed to capture the seasonality .Here we can see that the final ARIMA model passes the coefficient test also.The ACF plot and the autoplot suggests that there is no autocorrelation the Box-Ljung's test also agrees to it.We can see that the p value is greater than 0.05 which means it fails to reject the null hypothesis that there is white noise .Hence we can say that the residuals form a white noise .The Dickey Fuller test reject the null hypothesis that the residuals are non-stationary. Similarly the KPSS test fails to reject the null hypothesis that residuals are stationary .So we can conclude that the residuals form a white noise .Next . I then plotted the volatility and change of mean for the best ARIMA model I built .There was change in volatility ,which could be modeled,so I went ahead and performed GARCH modeling .This resulted in forming a GARCH model of ARMA order (0,1).I then performed Back testing for both the models. The MAPE value for the ARIMA model was 1.3 and that of GARCH was 1.09 .So it is evident that the GARCH model performs better and hence I plotted the prediction for the GARCH model.The forecast of the ARIMA model ,had the forecast fall within the mean range of the original series and the same for the forecast of the GARCH model .The beta value of the model was nearly 1 which indicates that we have long persistence.Also since the GARCH forecast indicated that the volatility is going to remain low,it means it is a good time to buy stocks .

.I learned that for financial data it is better to convert them to their log returns .I also , learned that sometimes the manually built ARIMA model performs better than the auto ARIMA ,hence it is always safe to perform manual building of the ARIMA models even if we have the option for auto arima .I also learnt that sometimes the seasonality can be too weak to capture .And seasonality is not very much in financial data.It is either null or very weak .I also learned that GARCH is the method for modeling volatility .I also learnt when we can use GARCH modeling after we have built the ARIMA model .

GARCH modelling was a bit challenging ,but with the help from professors lectures and also help dur9ng the meeting in his office hours helped me to understand the concept better .I have always been curious to learn why time series data are handled differently and how it should be handled .This project gave me a chance to the full understanding of the process .Since I was able to perform a variety of models on it and analyze it .I also learnt the significances of the omega,alpha and beta in the modelling of GARCH modelling and I think they give us a more deep insight of the time series data .I also learnt that forecasting the stock data cannot really tell a lot ,since some of the unforeseeable events like recession ,market crash are not taken into account we cannot fully depend on this analysis for making a confirmed decision .

Individual Report - Dharun Selvan

The variable allotted to me is variable ‘high’ on HDFC data set. In exploratory analysis I found that the data was not distributed properly so I went on and did log transformation and I found quite a lot of improvement on the histogram.The transformed data looks stationary in qq plot and that was confirmed using box test. When I did the decomposition plot, I could see a good upward trend.I was able to see some seasonality all over the plot and the random row show some very good randomness in the plot.

Then I created the time series for the variable. The auto plot did look good, so I took the difference of that i.e. log returns. The autoplot of the log returns is good and from there I started my analysis. The Acf plot and pacf plot of the log returns suggest to me the model would be AR so I went on and took eacf of the log returns. Where I could see some seasonality but I started with arima model my first model is arima (1,0,0) Based on my eacf plot .That the P value of ADF test and P value of kpss test of the residuals suggest me that the model stationary but acf plot of residual as well as the acf plot of residual^2 was correlated. Similar results for arima(1,0,1).

Since it showed seasonality I tried Sarima models. SARIMA (0,0,1)(2,0,0) [21] and SARIMA (1,0,1)(2,0,0) [21]. On residual analysis both variable showed nice acf plot on residual and acf plot on residual^2 without any correlation The ADF test have P value less than 0.05 and the kpss test have p value greater than 0.05 which means the models are stationary on comparing both models the first model have better results in back testing then the second model so I chose the first model for my forecasting.

After module 7 and 8 I thought of trying garch model as well since Syed already did the Garch model for variable ‘close’ I tried for variable high but the garch model didn't fit properly for my variable. The results were not good. Although the P value of ADF test and kpss test show the model is stationary where ADF plot of residual and residual^2 was not good. Thus I conclude that SARIMA (0,0,1)(2,0,0) [21]is the best model for the variable high.

I did a forecast for this model. I found out in late 2020 values are going little lower but after that it has some nice peaks later especially in 2021 March the values are quite high and it is safer to sell the stocks on that time soon after that the value moves towards the mean and it becomes a flat line that shows the output is as expected.

In the whole project I learned how to work with seasonal data, the HDFC data set is one challenging data that required different models for different variables. In my variable especially the benchmarks were very close to compare and to select the best model. The feedback on my reflection or module 7 and 8 by the professor where he mentioned using the domain knowledge to find the best model actually helped me in this project. I learned that even a small curve after the log transformation in the auto plot will give a huge to differences in the analysis