Modeling Financial Interval Time Series

Submitted by: Sweta Rai

Roll Number: C91/STS/181004

Year: 2018-2020

Date: 31-05-2020

Carried out at: Department of Statistics, University Of Calcutta

Under the Supervision Of: Prof. A. Chattopadhyay,

Prof. A. Sengupta, Prof. R. Bhattacharya

Contents

INTRODUCTION:

In financial economics, a large number of models are developed based on the daily closing price. When using only the daily closing price to model the time series, we may discard the valuable intraday information, such as the High, Low, Open price and Volume of shares for the day.

In this study, we propose an interval time series model, including the daily maximum and minimum prices and then apply the proposed model to forecast the interval of price. Here we can treat the maximum and the minimum prices as an interval valued observations.

A method to approach the interval time series is considering the maximum and the minimum process as vector of dimension 2×1. This leads to the vector autoregressive (VAR) model .The VAR model is one of the most successful, flexible, and easy to use models for the analysis of the multivariate time series. The VAR model has proven to be especially useful for describing the dynamic behavior of the economic and financial time series and for forecasting.

Volatility in capital market means the gap between increase or decrease of a stock price which is highly fickle and there would be a moment where the volatility will go up and down. High volatility means that the stock price increases and decreases significantly within a second. The volatility in capital market is significantly affects the return of an investment.

To account the volatility in the time series, we use **Principal component analysis**. PCA which is a method for dimensionality reduction of the data, is used in different fields such as statistical variables analysis, data compression, and visualization of high dimensional data. PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. Therefore, to account the volatility we fit a GARCH model on the **first principal component series (PC1)**.

The purpose of this project is to model the interval time series by fitting a VAR model and account the volatility by fitting GARCH model over the PC1 series and to forecast the interval(by interval we mean the Low and the High price) and the volatility of the shares.

OBJECTIVE:

- To study the dynamic relationships between the variables
- To analyse the interval valued observations and model it
- To estimate and forecast the volatility (if exist)
- To forecast the interval valued observations

DATA:

In this study, I have used intraday daily data of the Tata Consultancy Services Ltd (TCS) and from Jan-2015 to Dec-2019.

TATA Consultancy Services Limited (TCS) is the biggest Indian multinational information technology (IT) service and consulting company.

We have collected historical data of TCS NSE for the above years.

DATA SOURCE:

Link: https://in.finance.yahoo.com/quote/TCS.NS

>head(data)

Date	Open	High	Low	Close	Adj Close	Volume
1/1/2015	1283.5	1283.5	1270.5	1272.78	1092.68	366830
1/2/2015	1275.5	1295.47	1275.3	1289.72	1107.224	925740
1/5/2015	1290.5	1299.95	1262.32	1270.12	1090.397	1754242
1/6/2015	1264.55	1264.55	1220	1223.3	1050.202	2423784
1/7/2015	1235	1239.57	1203.72	1208.85	1037.797	2636332

MULTIVARIATE TIME SERIES ANALYSIS:

Multivariate time series analysis considers simultaneously multiple time series and deals specifically with dependent data. Let $\boldsymbol{z_t} = (z_{1t}, z_{2t}, \dots, z_{kt})'$ be a k-dimensional time series observed at equally spaced time points.

We decompose a multivariate time series as

$\boldsymbol{z_t} = \boldsymbol{\mu_t} + \boldsymbol{a_t}$, where

$\boldsymbol{\mu_t}$ follows some of the multivariate model and $\boldsymbol{a_t}$ is the innovation.

Here is the **time series plot** for the High and Low prices of intraday data from TCS (NSE). recorded from 1st Jan,2015 to 31st Dec,2019.

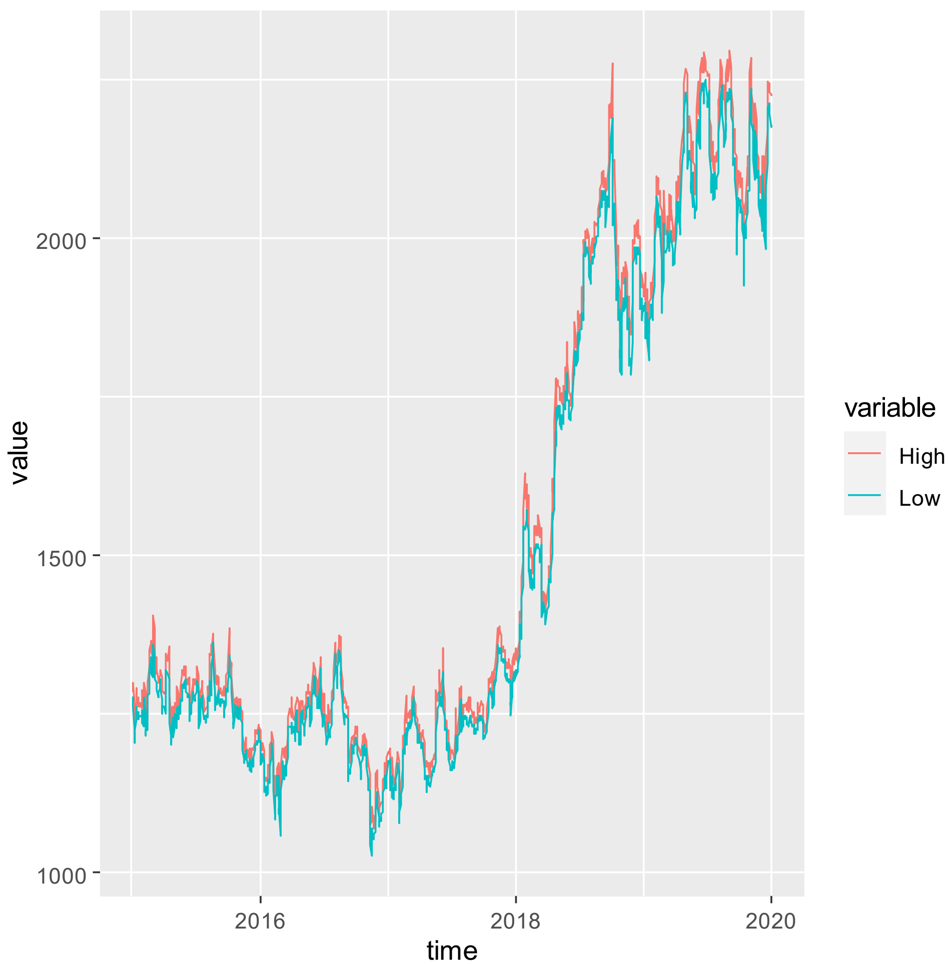


Figure: Time plot of the intraday High and Low stock market values of TCS from Jan, 2015 to Dec, 2019.

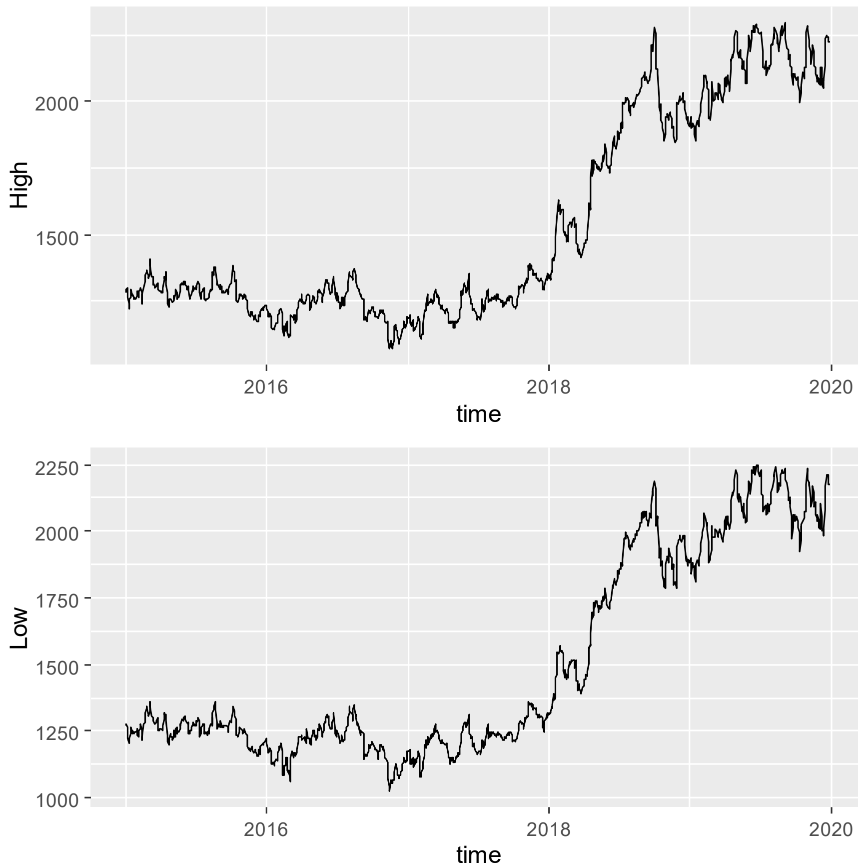


Figure: Time plots of High and Low prices of intraday data of TCS (NSE) from 2015 to 2019.

Here z_{1t}: time series corresponding to High price and z_{2t}: time series corresponding to Low price.

Summary for two series:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
High:	1070	1245	1312	1534	1946	2296
Low:	1026	1224	1285	1502	1901	2251

- **Linear trend** pattern visible from the above graphical representation.
- We **detrended** the series by taking first order differencing.
- We are working with logarithm of the **growth rates** or **returns** (without multiplier 100) of the two processes.

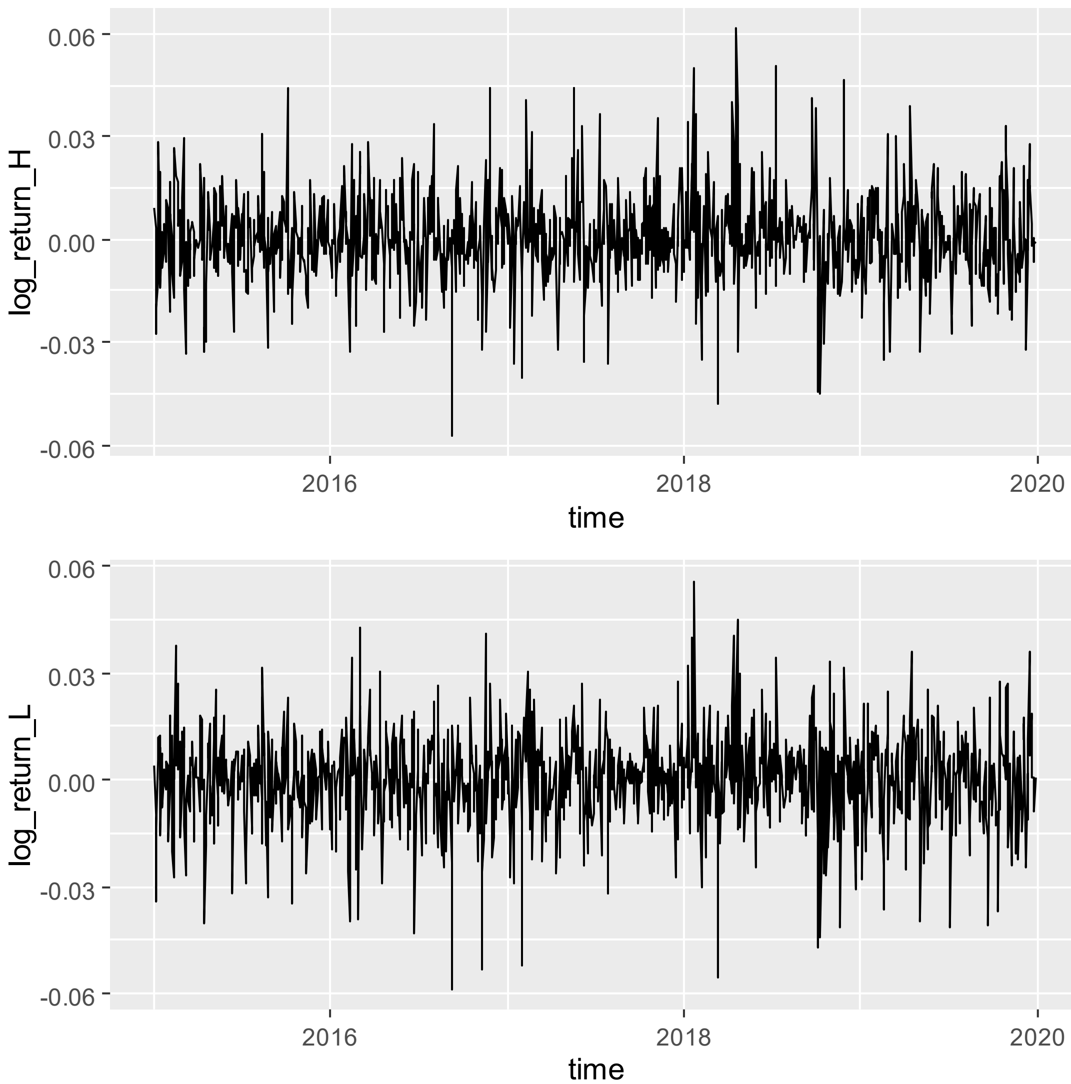


Figure: Time plots of the growth rate/returns of High and Low process (in logarithm) from the year2015 to 2019.

Now, **linear trend** is removed. Also from the above time plots we can say there is no evolutionary component in the series.

S TATIONARITY CHECK:

$$VAR(p)\ model: \qquad \boldsymbol{z_t} = \boldsymbol{\phi_0} + \sum_{i=1}^{p} \boldsymbol{\phi_i}\, \boldsymbol{z_{t-i}} + \boldsymbol{a_t},$$

where $\boldsymbol{a_t}$ is sequence of **iid** random vectors with mean **zero** and covariance matrix Σ_a, positive definite. Above model can be also written as

$$\phi(B) z_t = \boldsymbol{\phi_0} + \boldsymbol{a_t},$$

$$where\ \phi(B) = \boldsymbol{I_K} - \sum_{i=1}^{p} \boldsymbol{\phi_i} B^i \text{ is matrix polynomial of degree } p.$$

Necessary and sufficient condition for the **stationarity** of $VAR(p)$ series $\boldsymbol{z_t}$ is that all solutions of the determinant equation $|\phi(B)| = 0$ must be greater than 1 in modulus.

MULTIVARIATE UNIT-ROOT PROCESSES

Let $\Delta \boldsymbol{z_t} = \boldsymbol{z_t} - \boldsymbol{z_{t-1}} = (1-B)\boldsymbol{z_t}$ be the first differenced series of $\boldsymbol{z_t}$.The AR matrix polynomial can be rewritten as

$$\begin{aligned}\phi(B) &= I_k - \phi_1 B - \phi_{\boldsymbol{2}} B^2 - \cdots - \phi_p B^p \\ &= I_k - (\phi_1 + \phi_2 + \cdots + \phi_p)B + (\phi 2 + \cdots + \phi_p)B(1-B) \\ &\quad + (\phi_3 + \cdots + \phi_p)B^2(1-B) + \cdots + \phi_p B^{p-1}(1-B)\end{aligned}$$

Define $\Pi^* = \phi_1 + \phi_2 + \cdots + \phi_p, \Pi = \Pi^* - I_p$ and $\phi_j^* = -(\phi_{j+1} + \cdots + \phi_p),\ j = 1,2,\ldots,(p-1)$.

Therefore, the **AR matrix polynomial** can be written as

$$\phi(B) = I_k - \Pi^* B - (\phi_1^* B + \cdots + \phi_{p-1}^* (B)^{p-1})(1-B)$$

and

$$VAR(p) \text{ can be written as } \boldsymbol{z_t} = \Pi^* \boldsymbol{z_{t-1}} \Pi + \sum_{j=1}^{p-1} \phi_j^* \Delta \boldsymbol{z_{t-j}} + \boldsymbol{a_t} \,,$$

Or equivalently, we have

$$\Delta \boldsymbol{z_t} = \Pi \boldsymbol{z_{t-1}} + \sum_{j=1}^{p-1} \phi_j^* \Delta \boldsymbol{z_{t-j}} + \boldsymbol{a_t}$$

The null hypothesis of interest is

$$H_0 : \Pi = 0 \; or \; equivalently \; \Pi^* = \phi_1 + \cdots + \phi_p = I_k$$

$i.e.$ testing the hypothesis that each component of z_{it} of z_t has a unit root.

AUGUMENTED DICKEY-FULLER (ADF) TEST

For AR(p): $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \varepsilon_t, \varepsilon_t \; follows \; WN(0, \sigma^2)$

Characteristic polynomial: $1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p$, where B is the backward operator.

To test $H_0 \; vs \; H_1$, where

H_0: process is Non Stationarity

H_1: Not H_0 (Stationarity)

FOR HIGH PROCESS

Dickey-Fuller = -10.53, Lag order = 10, p-value = 0.01

FOR LOW PROCESS

Dickey-Fuller = -10.554, Lag order = 10, p-value = 0.01

As p-value < 0.05, we reject the null hypothesis.

i.e., both the series are **Stationary.**

LINEAR DYNAMIC DEPENDENCE:

To study the linear dynamic dependence of a stationary time series $\boldsymbol{z_t}$, we define lag l cross-covariance matrix

$$\Gamma_l = Cov(\boldsymbol{z_t}, \boldsymbol{z_{t-l}}) = E(\boldsymbol{z_t} - \boldsymbol{\mu})(\boldsymbol{z_{t-l}} - \boldsymbol{\mu})', \text{ where}$$

$\boldsymbol{\mu} = E(\boldsymbol{z_t})$ is mean vector of $\boldsymbol{z_t}$.

For $l = 0$, we have the covariance matrix Γ_0 of $\boldsymbol{z_t}$. Denote $(i,j)^{th}$ element of Γ_l as $\gamma_{l,ij}$,

$$\gamma_{l,ij} = Cov(\boldsymbol{z_{i,t}}, \boldsymbol{z_{j,t-l}}).$$

Therefore, for a positive lag l, $\gamma_{l,ij}$ can be regarded as a **measure of linear dependence** of the i^{th} component z_{it} on the l lagged value of the j^{th} component of z_{jt} .

For a stationary multivariate series $\boldsymbol{z_t}$, the lag l Cross-correlation matrix (CCM) ρ_l is defiend as

$$\rho_l = D^{-1}\Gamma_l D^{-1}, \text{ where}$$

$D = diag\{\sigma_1, \sigma_2, \dots, \sigma_k\}$ is the diagonal matrix of the standard deviations of the component of $\boldsymbol{z_t}$.

To summarize the information we consider the $\boldsymbol{k^2}$**plot** of the element of ρ_l for $l = 1,2,\dots,m$, where m is a pre-specified positive integer:

Specifically, for each $(i,j)^{th}$ position, we plot $\rho_{l,ij}$ vs l . This plot shows **the linear dynamic dependence of z_{it} on $z_{j,t-l}$**, $l = 0,1,\dots,m$ (generalization of the **autocorrelation function).**

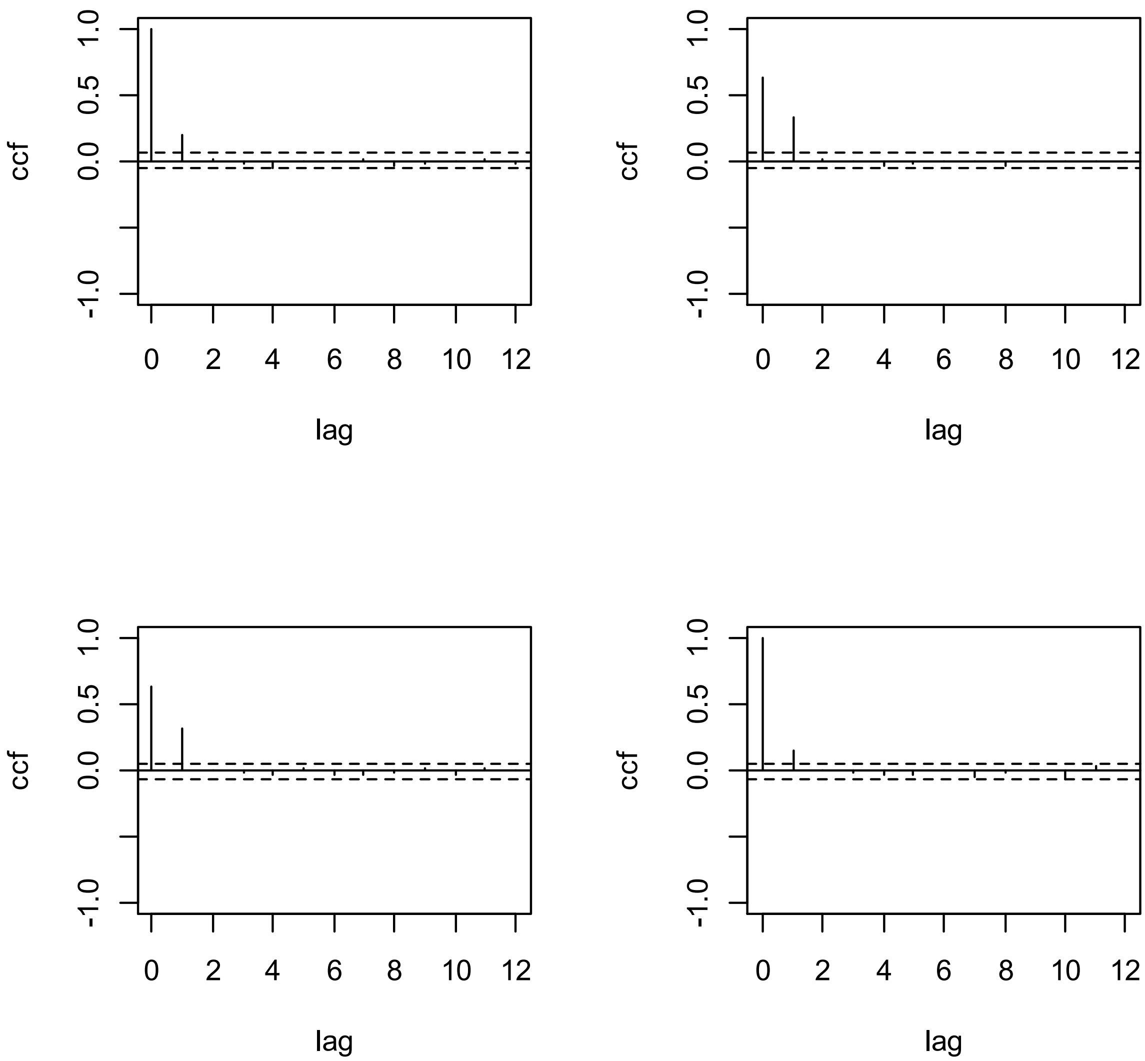


Figure: Sample cross-correlation plots for log of returns for the maximum and minimum process. The dashed lines indicate pointwise 95% confidence intervals.

- Clearly from the above CCM plot, ρ_l significant for $l = 1$.
- Indication of dynamic dependence in High and Low process.

Cross –covariance and correlation matrix:

```
Covariance matrix:
          High       Low
High   0.000155   0.000101
Low    0.000101   0.000164

CCM at lag:  0
       [,1]    [,2]
[1,]  1.000   0.635
[2,]  0.635   1.000

CCM at lag:  1
       [,1]    [,2]
[1,]  0.191   0.332
[2,]  0.325   0.150
```

TEST FOR ZERO-CROSS CORRELATION:

MULTIVARIATE PORTMANTEAU TEST

To detect the existence of linear dynamic dependence in the data, i.e. to test

$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0$ vs alternative $H_1: \rho_i \neq 0$, for some $i, 1 \leq i \leq m$

Multivariate Ljung–Box test statistic is

$$Q_k(m) = \frac{T^2 \sum tr\left(\widehat{\Gamma}_l \, \widehat{\Gamma_0}^{-1} \widehat{\Gamma}_l \widehat{\Gamma_0}^{-1}\right)}{T-l} \backsim^{asy} \chi^2_{mk^2},$$

where T is the sample size and m is a positive integer.

Multivariate Ljung-Box Statistics/Portmanteau test:

m	Q(m)	df	p-value
1	349	4	0
2	349	8	0
3	351	12	0
4	354	16	0
5	359	20	0

The statistic rejects the null hypothesis of zero cross-correlations. All p-values < 0.05, confirming that **the mutivariate series has some non zero CCM's**.

MODEL FITTING:

The multivariate time series $\boldsymbol{z_t}$ follows a VAR model of order p, VAR(p), if

$$\boldsymbol{z_t} = \boldsymbol{\phi_0} + \sum_{i=1}^{p} \phi_i \, \boldsymbol{z_{t-i}} + \boldsymbol{a_t} \;,$$

where $\boldsymbol{\phi_0}$ is a k dimensional constant vector and ϕ_i are $k * k$ matrix for $i > 0, \phi_p \neq 0$ and $\boldsymbol{a_t}$ is sequence of **iid** random vectors with mean **zero** and covariance matrix Σ_a, positive definite.

With the B back-shift operator, the model becomes

$$\phi(B)\boldsymbol{z_t} = \boldsymbol{\phi_0} + \boldsymbol{a_t}, \; where \; \phi(B) = I_K - \sum_{i=1}^{p} \phi_i B^i$$ matrix polynomial of degree p, with ϕ_l as lag l AR coefficient matrix.

Three principle followed in fitting a model:

i. Identification
ii. Estimation
iii. Diagnostic Checking

➢ IDENTIFICATION

For VAR models, model specification is to select the order p.

1) SEQUENTIAL LIKELIHOOD RATIO TEST

The idea of the approach is to compare a VAR(l) model with VAR($l-1$) model.
To test $H_0 : \phi_l = 0 \ \ ag \ \ H_1 : \phi_l \neq 0$ this is a problem of nested hypothesis and a natural test statistics to use is the likelihood ratio test.

Considering multivariate linear regression framework for a VAR(l) model. Let $\beta_l{}' = (\phi_0, \phi_1, \dots, \phi_l)$ matrix of coefficient parameters of a VAR(l) model and $\Sigma_{a,l}$ corresponding innovation covariance matrix.

Under the normality assumption, the likelihood ratio for the testing problem is

$$\Lambda = \frac{\max \ \ (\beta_{l-1}, \Sigma_{l-1})}{\max L(\beta_l, \Sigma_l)} = \left(\frac{|\widehat{\Sigma_{a,l}}|}{|\widehat{\Sigma_{a,l-1}}|}\right)^{\frac{T-l}{2}}$$

The likelihood ratio test is equivalent to rejecting H_0 for large values of $-2\ln(\Lambda)$.

A commonly used test statistics is $\mathrm{M}(l) = (T - l - 1.5 - kl) \ln\left(\frac{|\widehat{\Sigma_{a,l}}|}{|\widehat{\Sigma_{a,l-1}}|}\right) \sim^{asy} \chi^2_{k^2}$

2) INFORMATION CRITERIA

All criteria are likelihood based and consist of the two components, i.e. goodness of fit and parsimony. Three criteria functions we are using here to determine VAR order:

AIC, BIC and Hanan and Quinn (HQ) criteria.

Table for Order selection of VAR model for the given time series:

P	AIC	BIC	HQ	M(p)	p-value
0	-18.0104	-18.0104	-18.0104	0 0.0000	0.0000
1	-18.3151	-18.2985	-18.3089	377.4346	0.0000
2	-18.3795	-18.3462	-18.3669	85.7868	0.0000
3	-18.4074	-18.3574	-18.3886	41.6108	0.0000
4	-18.4309	-18.3643	-18.4058	36.2190	0.0000
5	-18.4681	-18.3848	-18.4368	52.6691	0.0000
6	-18.4748	-18.3749	-18.4372	15.8647	0.0032
7	-18.4801	-18.3635	-18.4362	14.2035	0.0067
8	-18.4779	-18.3446	-18.4277	5.1315	0.2741
9	-18.4790	-18.3291	-18.4226	9.1760	0.0568
10	-18.4946	-18.3281	-18.4319	26.3931	0.0000
11	-18.4914	-18.3082	-18.4225	3.9743	0.4095
12	-18.4916	-18.2917	-18.4164	7.9199	0.0946

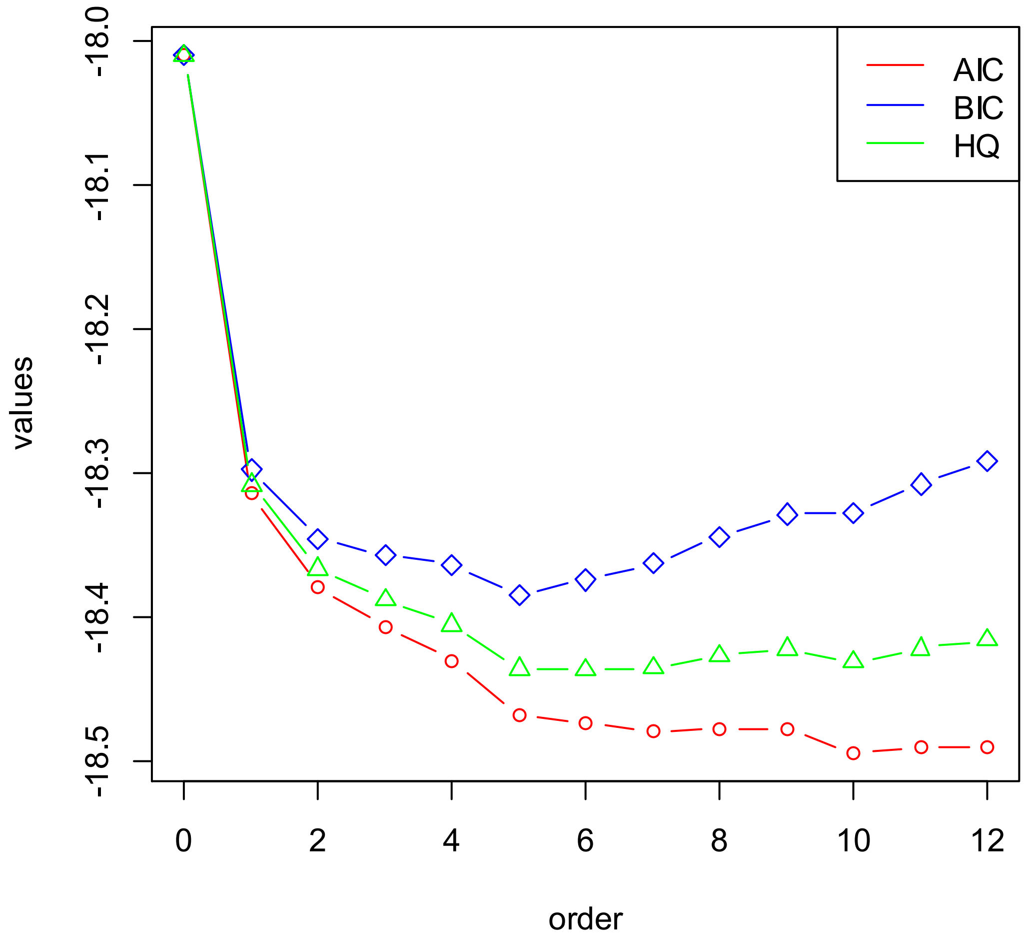


Figure: Plot showing values of the three information criteria for different order choices p.

- Sequential M-statistic selects the order p=5.
- AIC shows a relatively close values for p = {4, 5, 6}.
- BIC shows a clear minimum at p=5 with p= {3, 4, 6} as close, whereas HQ shows a minimum at p=6.
- In summary, a VAR(5) model may serve suitable for the given data.

➢ ESTIMATION

The model fitted is $VAR(5)$ and model is estimated by the Least Square Method. Parameters of interest are $\{\boldsymbol{\phi_0}, \phi_1, \dots, \phi_5\}$ and Σ_a .

Constant term:

Estimates: 0.0003679246 0.000373295
Std.Error: 0.0003322728 0.0003395285

AR coefficient matrix:

AR(1)-matrix

	[,1]	[,2]
[1,]	-0.126	0.458
[2,]	0.529	-0.192

standard error

	[,1]	[,2]
[1,]	0.0400	0.0389
[2,]	0.0409	0.0398

AR(2)-matrix

	[,1]	[,2]
[1,]	-0.278	0.186
[2,]	0.172	-0.276

standard error

	[,1]	[,2]
[1,]	0.0458	0.0450
[2,]	0.0468	0.0459

AR(3)-matrix

	[,1]	[,2]
[1,]	-0.164	0.193
[2,]	0.189	-0.167

standard error

	[,1]	[,2]
[1,]	0.0469	0.0462
[2,]	0.0479	0.0472

```
 AR( 4 )-matrix
        [,1]     [,2]
[1,] -0.154   0.0962
[2,]  0.139  -0.1776

 standard error
        [,1]      [,2]
[1,]  0.0455  0.045
[2,]  0.0465  0.046

 AR( 5 )-matrix
        [,1]       [,2]
[1,] -0.0333   0.0526
[2,]  0.1648  -0.1466
standard error
        [,1]       [,2]
[1,]  0.0397   0.0390
[2,]  0.0406   0.0399

Residuals covariance matrix:
          [,1]                 [,2]
[1,] 1.330466e-04   9.615421e-05
[2,] 9.615421e-05   1.389205e-04


 det(SSE) =  9.237266e-09
 AIC =  -18.46745
 BIC =  -18.38417
 HQ  =  -18.43611
```

➢ DIAGNOSTIC CHECKING

A fitted model is said to be adequate if

- All the fitted parameters are statistically significant at 5% level of significance.
- The residuals have no significant serial or cross-sectional correlations.
- There exist no structural changes or outlying observations.

TESTING FOR ZERO PARAMETERS

An approach to simplify a VAR(p) model is to remove insignificant parameters. Given a specified significant level, we can identify the target parameters for removal (By target parameters we mean those parameters whose individual t-ratio is less than the critical value of the normal distribution with type I error α).

To check the above we perform χ^2 test.

```
Number of targeted parameters:  4
Chi-square test and p-value:  3.313466    0.5068016
```

As p-value>0.05, there is no insignificant parameters in the fitted VAR model.

RESIDUAL CROSS-CORRELATIONS

The residuals of an adequate model should behave like a white noise series. Checking the serial and cross-correlations of the residuals thus becomes an integral part of the model checking.

Let $\hat{A} = Z - X\hat{\beta}$ – residual matrix of a fitted $VAR(p)$ model, i^{th} row of $\hat{A}$ contain

$$\widehat{a}_{p+i} = Z_{p+i} - \widehat{\phi}_0 - \Sigma_{i=1}^{p} \widehat{\phi}_i z_{t-i}.$$

The lag l cross-covariance matrix of the residual series is defined as

$$\hat{C}_l = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{a}_t \, \hat{a}_{t-1}{}' \quad \text{and}$$

$$\hat{R}_l = \hat{D}_l^{-1}\hat{C}_l\hat{D}_l^{-1}, where\ \hat{D} = \sqrt{diag}\left(\hat{C}_0\right)$$

<u>Cross-covariance and correlation matrix for the residual series</u>:

```
 Covariance matrix:
            High       Low
High  1.33e-04  9.62e-05
Low   9.62e-05  1.39e-04

 CCM at lag:  0
        [,1]      [,2]
[1,]  1.000    0.707
[2,]  0.707     1.000

 CCM at lag:  1
        [,1]        [,2]
[1,] -0.00080   0.00631
[2,]  0.00187  -0.00331
```

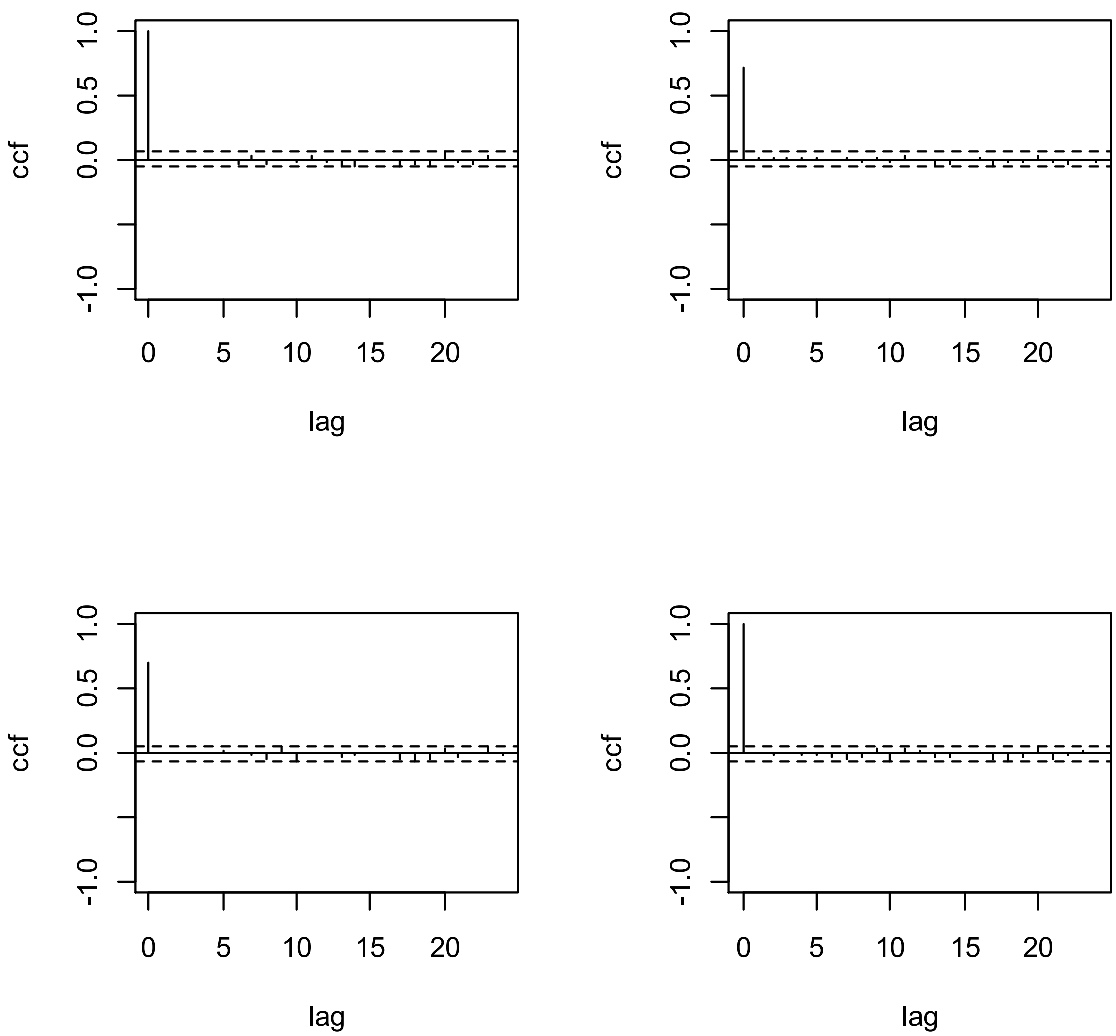


Figure: Residual cross-correlation matrices of the $VAR(5)$model for the logarithm of growth rates of High and Low stock market values of TCS intraday data from the Jan, 2015 to Dec, 2019.

From the above CCM plot of the Residual series, we see no significant dynamic dependence of the residual series.

MULTIVARIATE –PORTMANTEAU TEST

Let R_l be the theoretical lag l cross- correlation matrix of innovation $\boldsymbol{a_t}$. The hypothesis of interest in model checking is

$$H_0: R_1 = R_2 = \cdots = R_m = 0 \;\; vs \quad H_1: R_j \neq 0, \text{ for some}$$

$1 \leq j \leq m, m$ is a prespecified positive integer.

Portmanteau Test Statistics:

$$Q_k(m) = T^2 \sum_{l=1}^{m} \frac{1}{T-l} tr(\hat{C}_l{}'\hat{C}_0^{-1}\hat{C}_l\hat{C}_0^{-1}) \sim^{asym} \chi^2_{(m-p)k^2},$$

where the adjustment pk^2 in the degrees of freedom of the chi-square distribution is set to the number of estimated AR parameters.

Ljung-Box Statistics:

m	Q(m)	p-value
1.000	0.574	1
2.000	1.926	1
3.000	4.031	1
4.000	8.519	1
5.000	18.803	1
6.000	36.517	0
7.000	43.748	0
8.000	46.570	0

As p-value <0.05 , for m>5. There is cross-correlation in the residual series. The residuals are not random, i.e. does not behaves like white noise.

RESIDUAL PLOTS:

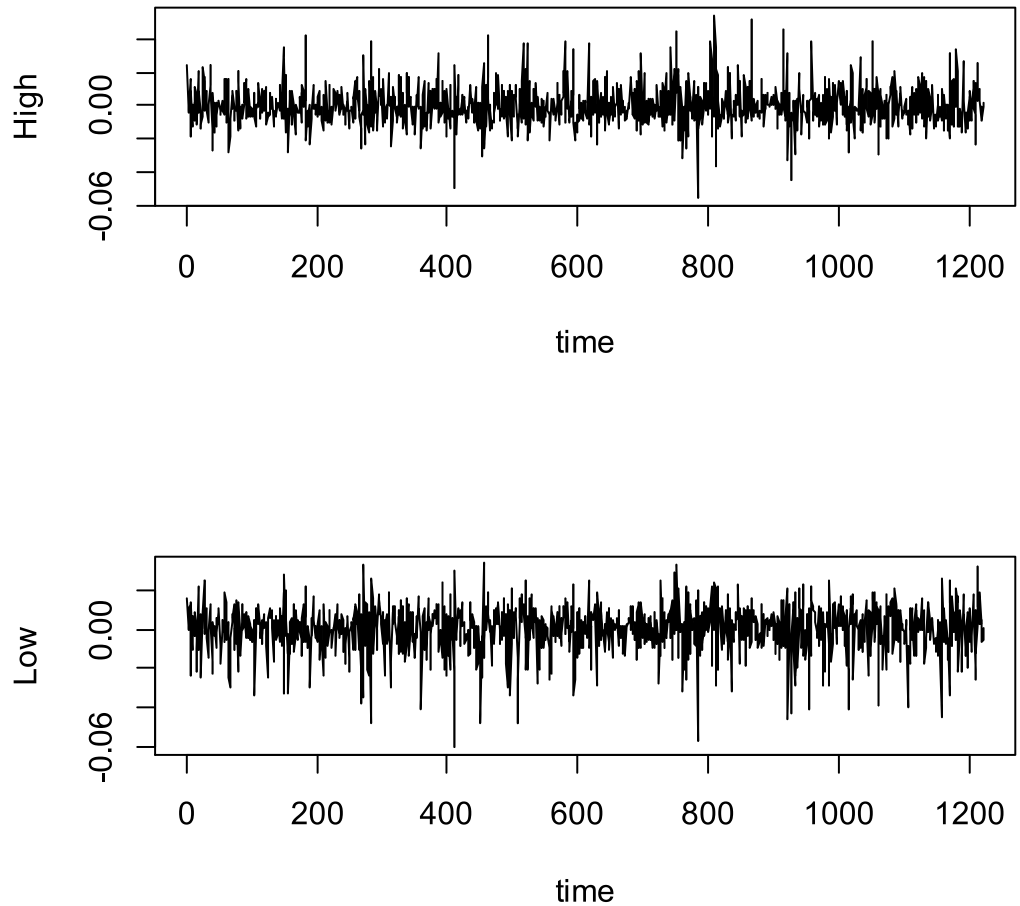


Figure: Residual plots of the fitted VAR(5) model for the intraday growth rates of High and Low stock market prices of TCS from the 1^{st} Jan of 2015 to the 31^{st} of 2011.

There are uneven fluctuations in the residual series which implies **presence of <u>conditional heteroscedasticity</u>** in the vector time series.

TESTING CONDITIONAL HETEROSCEDASTICITY:

Usually we assume the innovations $\boldsymbol{a_t}$ of a multivariate time series $\boldsymbol{z_t}$ are serially uncorrelated and have zero mean and positive definite covariance matrix. Let F_{t-1} denote the $\sigma - field$ generated by the past data $\{\boldsymbol{z_{t-i}}, i = 1,2, \ldots\}$.

We have $E(\boldsymbol{a_t}|F_{t-1}) = 0\ and\ E(\boldsymbol{a_t}\boldsymbol{a_t}'\,|\,F_{t-1}) = \Sigma_a$ positive definite, which is a constant matrix.

Most financial time series have **conditional heteroscedasticity**. Let $\Sigma_t = Cov(\boldsymbol{a_t}|F_{t-1})$ – conditional covariance matrix of $\boldsymbol{z_t}$ given F_{t-1} .

PORTMANTEAU TEST

If $\boldsymbol{a_t}$ has no conditional heteroscedasticity, then its conditional covariance matrix Σ_t

is time-invariant. This implies Σ_t, hence $\boldsymbol{a_t^2}$ does not depends on $\boldsymbol{a_{t-i}^2}, for\ i > 0$.

To test $H_0\colon \rho_1 = \rho_2 = \cdots = \rho_m = 0 \quad vs\ alternative\ H_1\colon \rho_i \neq 0$, for some $i \leq m$, where ρ_i is lag i cross-correlation matrix of $\boldsymbol{a_t^2}$.

Test Statistics

$$Q_k^*(m) = T^2 \sum_{i=1}^{m} \frac{1}{T-i} b_i{}'(\hat{\rho}_0^{-1} \otimes \hat{\rho}_0^{-1})\, b_i\ ,$$

where T –denotes sample size and $b_i = vec(\hat{\rho}_i{}')$ with $\hat{\rho}_i$ lag i sample cross-correlation matrix of $\boldsymbol{a_t^2}$.

Here $Q_k^*(m)$ is **asymptotically** equivalent to the **multivariate generalization of the lagrange multiplier (LM) test** of Engle for conditional heteroscedasticity.

RANK-BASED TEST

Let R_t be the rank of $e_t, where\ e_t = a_t{}' \Sigma\, a_t,\ \Sigma -$denotes the unconditional covariance matrix of $\boldsymbol{a_t}$.

The lag l rank autocorrelation of e_t is

$$\tilde{\rho}_l = \frac{\sum_{t=l+1}^{T}(R_t - \bar{R})\,(R_{t-1} - \bar{R})'}{\sum_{t=l+1}^{T}(R_t - \bar{R})^2}$$

$$, l = 1,2, \ldots \quad where\ \bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t = \frac{T+1}{2}\ and\ \sum (R_t - \bar{R})^2 = \frac{T(T^2-1)}{12}$$

Test statistics

$$Q_R(m) = \sum_{i=1}^{m} \frac{\left(\tilde{\rho}_i - E(\tilde{\rho}_i)\right)^2}{Var(\tilde{\rho}_i)} \sim^{asym} \chi^2_{m^2},$$
$$provided\ e_t has\ no\ serial\ dependence.$$

Q(m) of squared series(LM test):
Test statistic: 47.95817 p-value: 6.316805e-07

Rank-based Test:
Test statistic: 125.7197 p-value: 0

As p-value < 0.05, there is **heteroscedasticity** in the residual series.

VOLATILITY:

Volatility is a statistical measure of the dispersion of returns for a given security or market index. Conditional heteroscedasticity exists in finance because asset returns are volatile. There are many different ways to measure volatility, indicating beta coefficients, option pricing, and standard deviations. More volatile assets are considered riskier than less volatile assets because the price is expected to be less predictable.

For the TCS(NSE) stock values from the year 2015 to 2019, to account the dynamic volatility of the corresponding stock price: we first use Principal component analysis (PCA) over the variables: Open, High, Low and Close price to reduce the multivariate setup to univariate.

PCA identifies a small number of principal components that explain most of the variation in a data set. This method is often used for dimensionality reduction and analysis of the data.

Here taking the first Principal Component (PC1) which explains data variability the most, we fit a suitable GARCH model on the PC1 series to account the volatility.

Here the PC1 series (standardized) serves as some sort of weighted average of the Open, High, Low and Close price of the stock values.

Summary of the PCA (on the standardized data):

	PC1	PC2	PC3	PC4
Standard deviation	1.9994	0.03983	0.02692	0.01233
Proportion of Variance	0.9994	0.00040	0.00018	0.00004
Cumulative Proportion	0.9994	0.99978	0.99996	1.00000

Proportion of variability explained by the first principal component PC1 is 0.9994, that explains taking PC1 series for model fitting is reasonable.

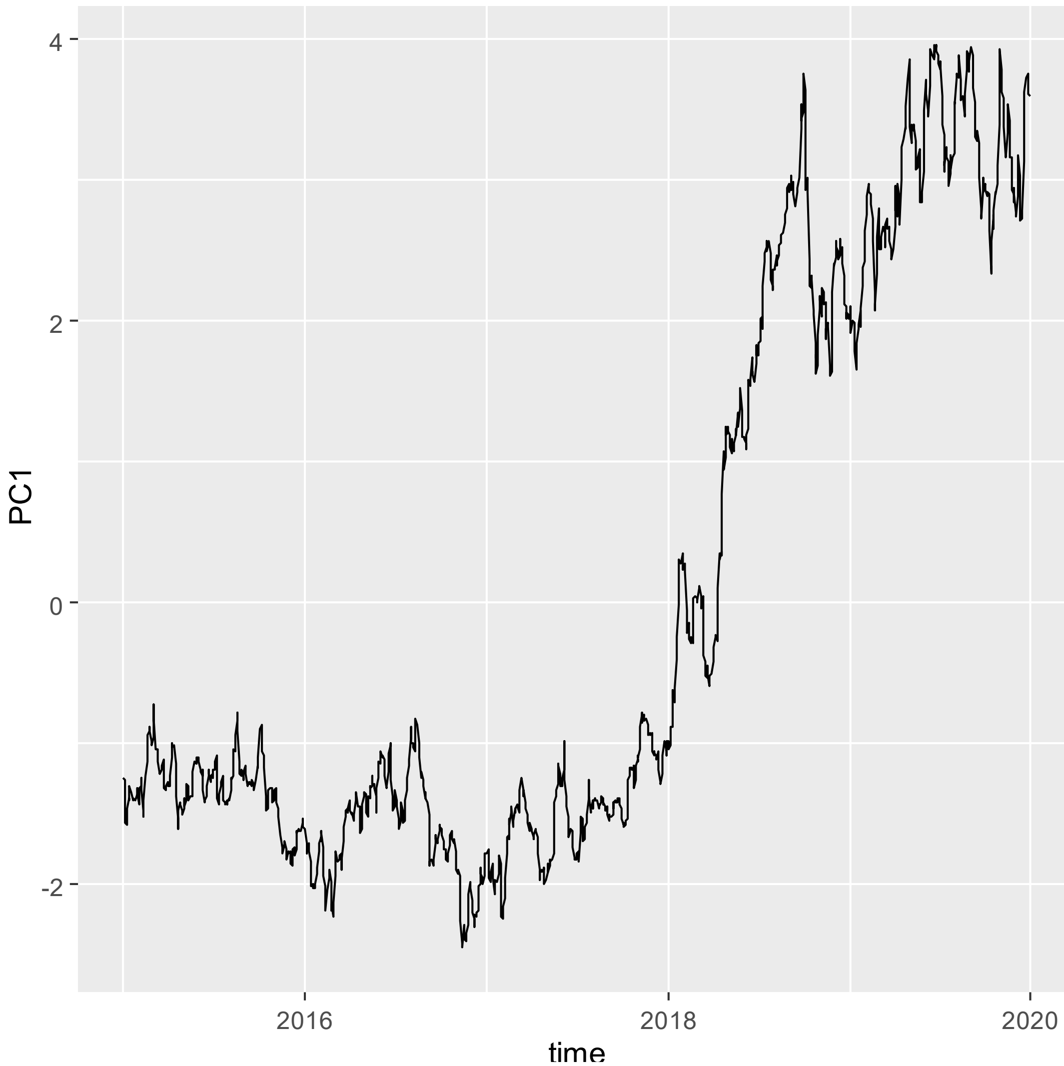


Figure: Time plot for the 1st PC series for TCS stock market values from Jan,2015-Dec,2019.

From the above plot, we can clearly see presence of trend. Therefore, we have detrend the PC1 series before proceeding further.

Plot for the Detrended Series:

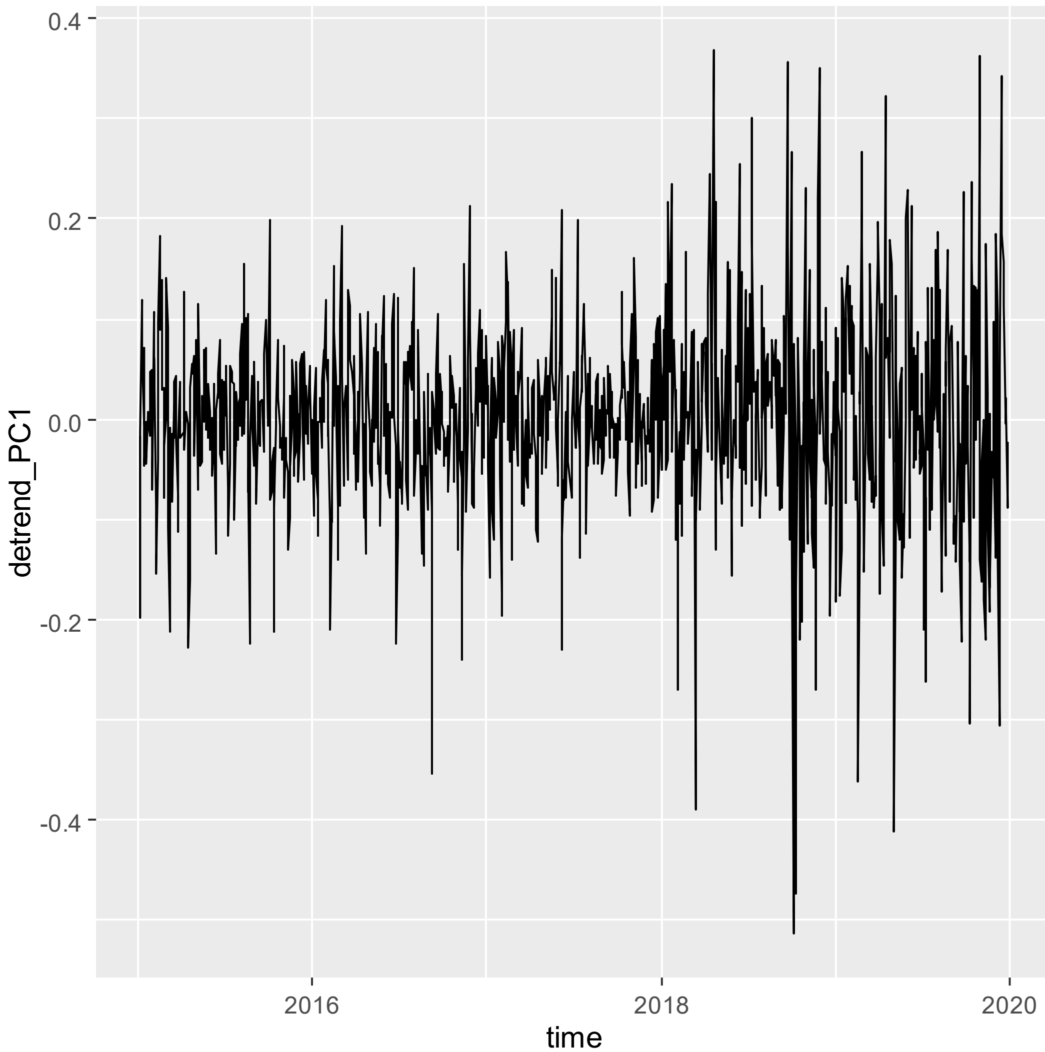


Figure: Time plot for the detrended PC1 series for the given data set.

There is **uneven fluctuations** in the series which imply presence of **conditional heteroscedasticity**.

BUILDING AN GARCH(p, q) MODEL

We fit the model

$$\mathrm{X_t} = \sigma_t \mathrm{u_t},\ \mathrm{u_t}\ \text{follows}\ \mathrm{WN}(0, \sigma^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_t^2 + \cdots + \beta_q \sigma_{t-q}^2.$$

ACF AND PACF PLOT:

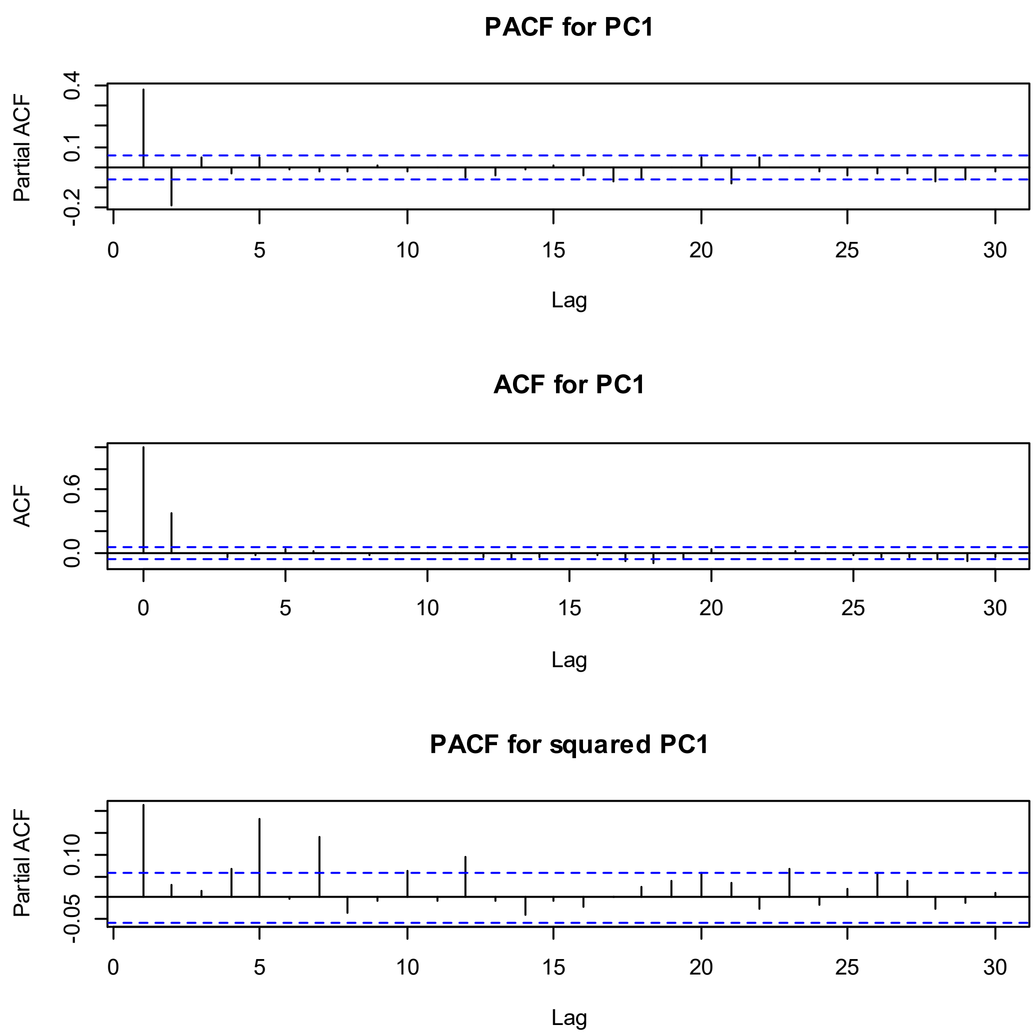


Figure: PACF and ACF plot for the detrended PC1 series and PACF for the squared series for the given data set from Jan,2015 to Dec,2019.

STATIONARITY CHECK

Augmented Dickey-Fuller Test

lag	ADF	p-value
0	-23.4	0.01
1	-23.5	0.01
2	-18.5	0.01
3	-16.8	0.01
4	-14.4	0.01
5	-13.5	0.01
6	-12.8	0.01

As p-values are < 0.05, i.e. the PC1 process is **Stationary**.

GENERALISED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY (GARCH(p,q)) MODEL:

GARCH model (Generalized Autoregressive Conditional Heteroscedastic) model is a generalized form of ARCH. GARCH model allows the conditional variance based on the conditional variance of the previous lag. The equation of conditional variance becomes as presented

$$\mathrm{X_t = \mu_t + \sigma_t u_t,\ u_t\ follows\ WN(0, \sigma^2)}$$

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2,$$

$$\alpha_j \geq 0, j = 1(1)p, \beta_k \geq 0, k = 1(1)q,\ \sum_{i=1}^{\max(p,q)} (\alpha_j + \beta_j) < 1\ ,$$

where X_t stationary process and $\sigma_t^2 -$ conditional variance. Here mean $\boldsymbol{\mu_t}$ is model by fitting a suitable **ARMA(p, q) model** with order specified from the ACF and PACF plot of the PC1 series, we see a significant spike at h=2 and h=1 in the PACF and ACF plot respectively, i.e. we get p=2 and q=1, fit ARMA(2,1).

Three principles for model fitting are

- Identification
- Estimation
- Diagnostic Checking

For **identification** of p, q in GARCH model, checking the PACF and ACF plot of the squared PC1, we see that there is a significant spike at h=1, get p=1 and q=1, i.e. fit GARCH(1, 1).

ESTIMATION:

GARCH Model Fit

Conditional Variance Dynamics

GARCH Model : sGARCH(1,1)
Mean Model : ARMA(2,1)
Distribution : std

Parameters	Estimate	Std.Error	t-value	p-value
mu	0.003505	0.002727	1.28544	0.198640
ar1	-0.116374	0.139003	-0.83721	0.402476
ar2	0.049305	0.068881	0.71579	0.474119
ma1	0.562772	0.134314	4.18998	0.000028
Omega	0.000066	0.000048	1.37161	0.170184
alpha1	0.046245	0.019107	2.42035	0.015506
beta1	0.945754	0.023779	39.77185	0.000000
Shape	5.348288	0.785408	6.80956	0.000000

LogLikelihood : 1442.143

Information Criteria

Akaike -2.3376
Bayes -2.3043
Hannan-Quinn -2.3251

Hence the fitted model is

$$\mathrm{X_t = \mu_t + \sigma_t u_t,\ u_t\ follows\ WN(0, \sigma^2)}$$

$$\mu_t = 0.003505 - 0.116374\ X_{t-1} + 0.049305\ X_{t-2}\ + 0.562772\ u_t$$

$$\sigma_t^2 = \ 0.000066\ + 0.046245 X_{t-1}^2 + +0.945754\ \sigma_{t-1}^2\ .$$

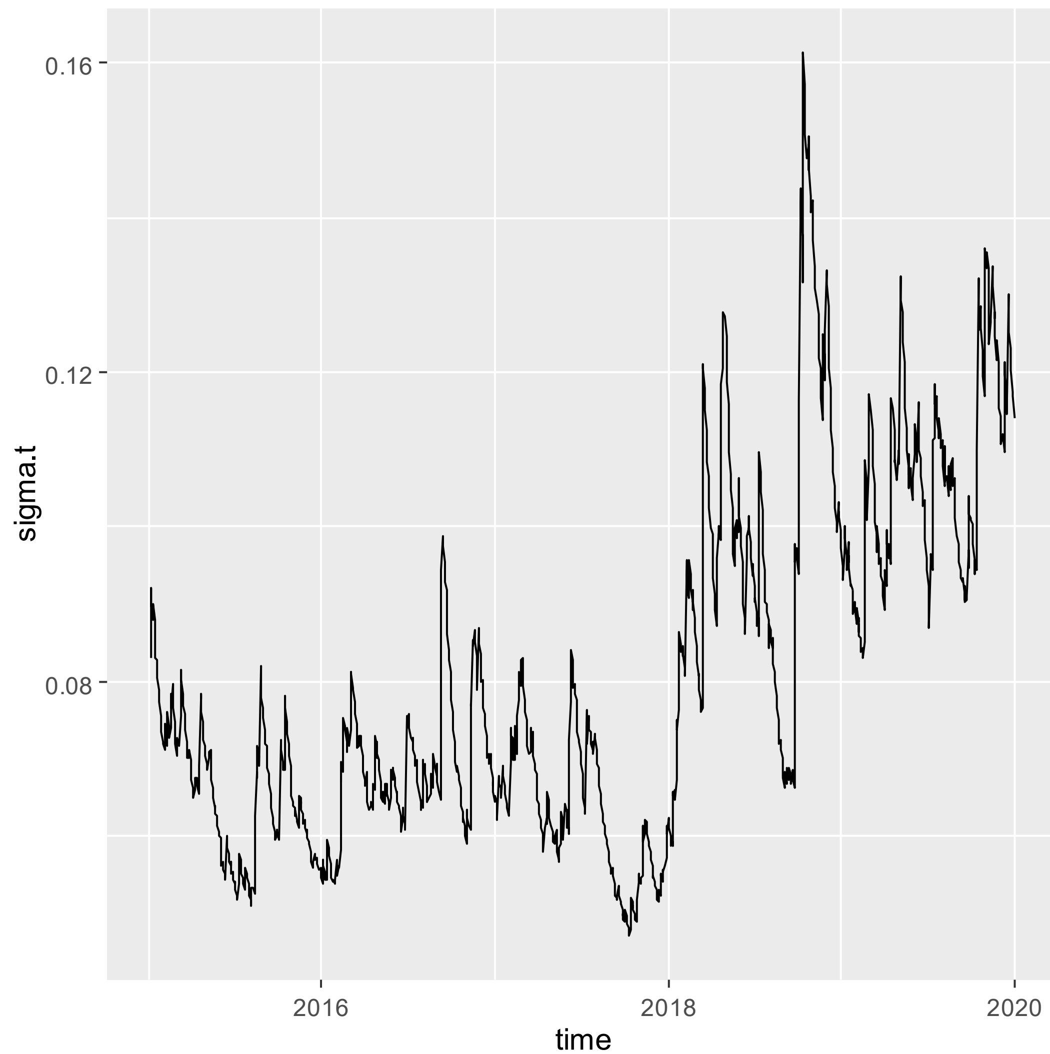


Figure: Plot for the estimated values of Standard deviation obtained by fitting GARCH(1,1) for the PC1 series of the TCS.NSE stock values, from Jan,2015 to Dec,2019.

DIAGNOSTIC CHECKING:

To check the efficacy of the model we test ARCH Effect, the serial correlation, the mean zero of the residuals. The normality of the residuals using histogram, QQ plot and the Ljung –Box test have been used.

TEST FOR ARCH/GARCH BEHAVIOR IN STANDARDIZED RESIDUALS

Weighted Ljung-Box Test on Standardized Residuals

Lag	Statistic	p-value
Lag[1]	0.7648	0.3818
Lag[2*(p+ q)+(p+ q)-1][8]	2.2999	1.0000
Lag[4*(p+ q)+(p+ q)-1][19]	4.4437	0.9488

d.o.f=3

To test H_0 vs H_1 , where

H_0: No serial correlation

H_1: Serial correlation

As p-value>0.05, there is no evidence of serial correlation in the residual series.

Weighted Ljung-Box Test on Standardized Squared Residuals

Lag	Statistic	p-value
Lag[1]	2.478	0.1154
Lag[2*(p+ q)+(p+ q)-1][5]	3.759	0.2857
Lag[4*(p+ q)+(p+ q)-1][9]	7.711	0.1466

d.o.f=2

As p-value>0.05, there is no evidence of serial correlation in the squared residual series.

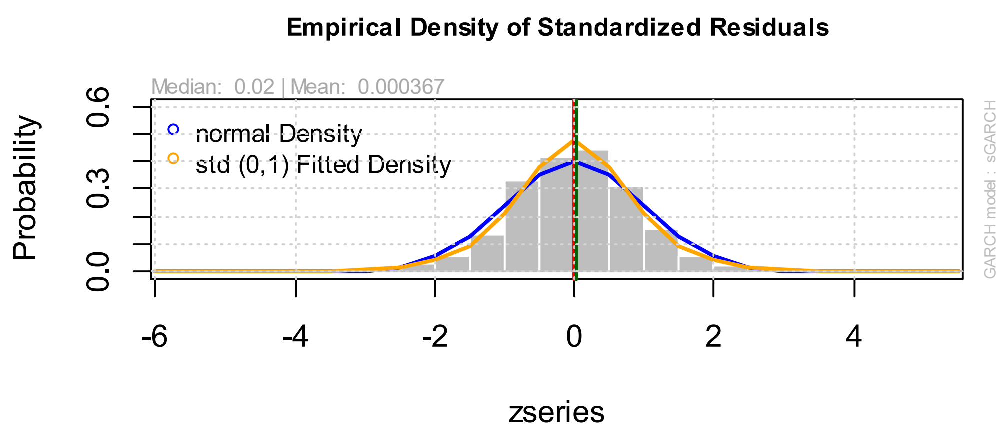
Weighted ARCH LM Tests

Lag	Statistic	Shape	Scale	p-value
ARCH Lag[3]	0.7583	0.500	2.000	0.3839
ARCH Lag[5]	2.1069	1.440	1.667	0.4481
ARCH Lag[7]	5.0373	2.315	1.543	0.2208

To test H_0 *vs* H_1 , where H_0: Residuals are homoscedastic. No ARCH Effect

H_1: Residuals are heteroscedastic

As p-value> 0.05, we don't reject H_0. Hence, there is **no ARCH Effect** in the residual series.



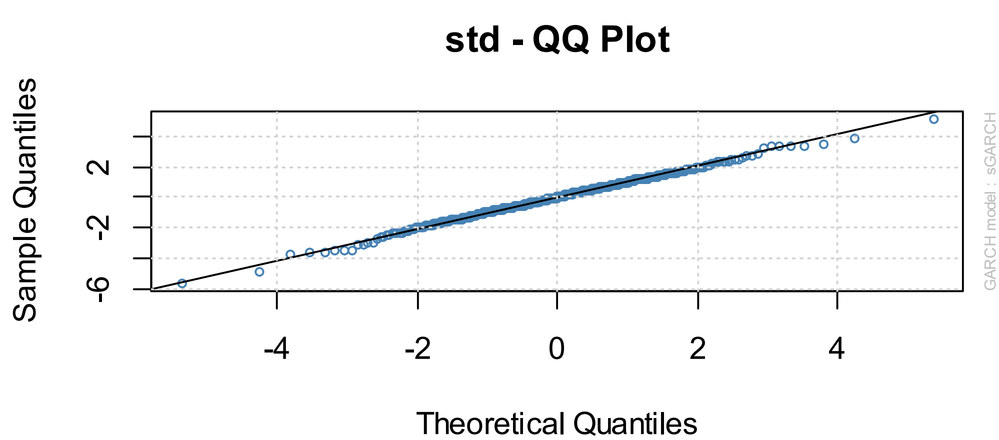


Figure: Histogram and QQ-plot of the residuals obtained after fitting the GARCH(1, 1) model.

From the above histogram and QQ- plot of the residuals, we observe that the residuals are normally distributed with mean zero, i.e. the normality assumption is good. From the Ljung-Box test and the LM test of the residuals series, we get evidence to support the fitted model, i.e. overall the above fit is adequately good. Hence, we get reason to support forecast from the fitted GARCH(1,1) model.

FORECAST:

Forecasting is a method to predict the future by evaluating the information and data in the prior period. We predict the **Interval time series** using VAR(5) model and the **Volatility** of the stock values of TCS using GARCH(1,1).

The forecasted values of the High and Low price and the predicted conditional SD for the next 15 days are given below:

Time point	Low	High
1	2180.677	2222.973
2	2184.360	2220.948
3	2186.521	2221.672
4	2188.526	2223.267
5	2190.331	2225.465
6	2190.113	2227.869
7	2190.904	2228.513
8	2192.143	2228.948
9	2193.268	2229.717
10	2194.449	2230.760
11	2195.647	2232.058
12	2196.545	2233.378
13	2197.559	2234.423
14	2198.637	2235.402
15	2199.713	2236.411

Time point	Sigma
1	0.1113391
2	0.1111893
3	0.1110406
4	0.1108928
5	0.1107460
6	0.1106002
7	0.1104554
8	0.1103115
9	0.1101687
10	0.1100267
11	0.1098858
12	0.1097457
13	0.1096066
14	0.1094685
15	0.1093313

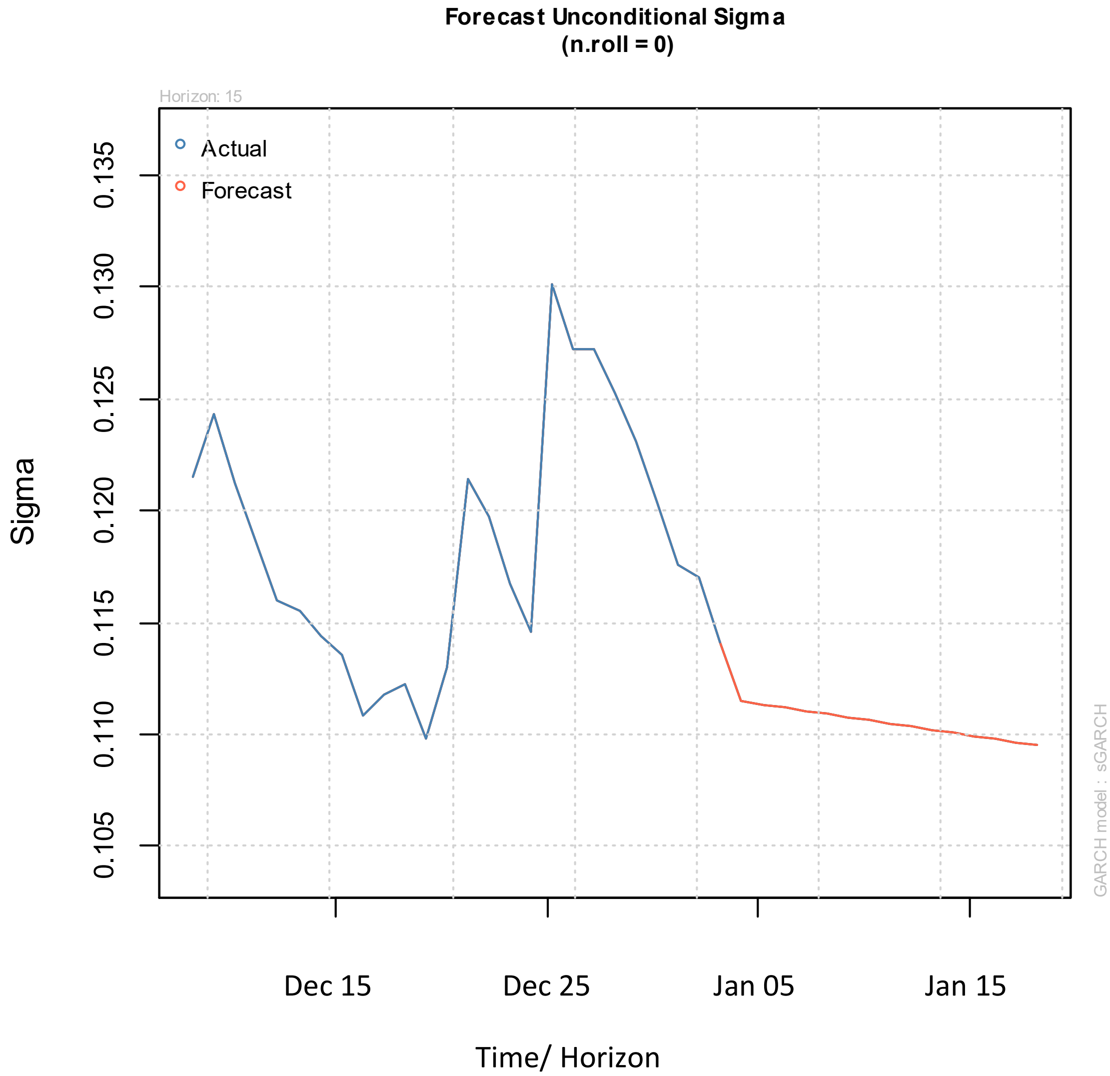


Figure: Graphical representation for the forecasted volatility.

CONCLUSION:

In this project the aim was to model the TCS(NSE) stock value interval time series using the VAR model. The result of the project shows that the fitted VAR(5) model is adequately good for prediction of the interval i.e. predict the High and Low value simultaneously. Since predicting the High and Low prices of the stock value, thus we are providing the interval of price movement for the day.

But there is evidence of presence of heteroscedasticity in the residual time series of VAR model. To account the heteroscedasticity in TCS stock values we perform Principal component analysis over the Open, High, Low, Close price and take the first principal component series(which explains more than 99% of the total variation of the data) to fit a univariate GARCH(1,1) model. The fluctuation of price of the stock values is well accounted by the GARCH(1,1) model, which has the ability to forecast volatility with certain probability.

R-CODE:

```
intsall.packages("MTS")

install.packages("forecast"))

install.packages("vars")

install.packages("tseries")

install.packages("rugarch")

install.packages("gridExtra")

library(ggplot2)

library(tseries)

library(vars)

library(MTS)

library(forecast)

library(gridExtra)

####Calling the data file
```

```
data=read.csv(file.choose())
data=data[order(as.Date(data$Date,format="%m/%d/%Y")),]
head(data)

Date=as.Date(data$Date,format="%m/%d/%Y")
df=data.frame(Date,data$High,data$Low,data$Close,data$Open)
names(df)=c("Date","High","Low","Close","Open")
head(df)

#data summary
summary(df$High)
summary(df$Low)

####Plot of the original time series
plot1=ggplot(df, aes(x = Date,y = High))+
  geom_line()+xlab("time")

plot2=ggplot(df, aes(x = Date,y = Low))+
  geom_line()+xlab("time")
grid.arrange(plot1, plot2, nrow=2)

library(tidyr)
library(dplyr)
library(reshape2)
meltdf=melt(df,id=c("Date","Open","Close"))
```

```
ggplot(meltdf,aes(x=Date,y=value,colour=variable,group=variable)) +
geom_line()+xlab("time")


#From the above plot it can be clearly seen that there is presence of trend pattern

#work with the log of the return values or growth rate


df$Max=log(df$High)

df$Min=log(df$Low)

head(df)


####converting the data in matrix form


mat=cbind(df$Max,df$Min)

colnames(mat)=c("High","Low")


mat=diffM(mat) ##to detrend

T=length(df$Date)


df$log_return_H[2:T]=mat[,1]

df$log_return_L[2:T]=mat[,2]

df=df[-1,]

head(df)


####plot for the log of returns or log of growth rate without the multiplier 100
```

```
plt1=ggplot(df, aes(x = Date,y=log_return_H))+
  geom_line()+xlab("time")


plt2=ggplot(df, aes(x = Date,y = log_return_L))+
  geom_line()+xlab("time")
grid.arrange(plt1, plt2, nrow=2)


####To study the linear dynamic dependence in data
ccm(mat,level=T)
str(ccm(mat))
#Sample cross-correlation plots for log of returns from TCS data.
#Detect the existence of the linear dynamic dependence in the data to test hypothesis
H0:row1=row2=0 vs alternative


####Check for stationarity
library(aTSA)
apply(mat,2,adf.test)


#Both series are stationary and hence there is no requirement for Cointegration test


####Testing zero cross correlation
mq(mat,lag=6)


#VAR(p) Model Order Selection Criterion
#Computes information criteria and sequential Chi-square statistics for VAR process
```

```
fitt=VARorder(mat,maxp=12,output=T)
```

####Figure shows the time plots of the three information criteria

```
plot(0:12,fitt$aic,ylab="values",xlab="order",type="b",col="red")
lines(0:12,fitt$bic,type="b",pch=5,col="blue")
lines(0:12,fitt$hq,type="b",pch=2,col="green")
legend("topright",c("AIC","BIC","HQ"),col=c("red","blue","green"),lty=1)
```

#Fit the VAR model corresponding to the BIC and M statistics

#In summary,a VAR(4) or VAR(5) model may serve as a starting model for the given data.

#From the M statistics we take order of VAR as p=5.

####Fitting VAR(5) model

#Perform least squares estimation of a VAR model

```
fit=VAR(mat,p=5)
#str(fit)
```

####Testing Zero Parameters

#To perform the chi-square test for zero parameters, we use the command VARchi in the MTS package

```
VARchi(mat,p=5,thres=1.96)
```

```
####This stage of the analysis deals with the diagnostic checking process.
resi=residuals(fit)   # Obtain the residuals of VAR(5)fit.

MTSdiag(fit,gof=8,level=T,adj=20)

#adjustment in df of chi-square distr is set to no of estimated AR parameters.
#Multivariate Portmanteau Statistics
#Residual Cross-Correlations
#Residuals of an adequate model should behave like a white noise series.
#Checking the serial and cross-correlations of residuals thus becomes an integral part of model checking
#If residuals has no condi heteroscedasticity then its condi-cov matrix is time-invariant,

ccm(resi^2)
MarchTest(resi)

####prediction for the next 15 days
prediction=VARpred(fit,h=15)
#str(prediction)

pred=as.matrix(prediction$pred)

High_L=data$High[length(data$High)]
Low_L=data$Low[length(data$High)]
```

```
pred_High=0
pred_Low=0

pred_High[1]=exp(pred[1,1])*High_L
pred_Low[1]=exp(pred[1,2])*Low_L

for(i in 2:15)
{
 pred_High[i]=pred_High[i-1]*exp(pred[i-1,1])
 pred_Low[i]=pred_Low[i-1]*exp(pred[i-1,2])
}
pred_High
pred_Low
####To account volatility
####Univariate reduction of multivariate time series
####PCA for the TCS Data for variables Open, High, Low, Close

head(df)
dat_PCA=cbind(df$Open,df$High,df$Low,df$Close)
colnames(dat_PCA)=c("Open","High","Low","Close")

####Principal Component Analysis on dat_PCA
pca=prcomp(dat_PCA,center = TRUE,scale. = TRUE)
summary(pca)
```

```
#screeplot(pca)

#str(pca)

matrix_PC=as.matrix(pca$x)


####Extracting the 1st Principal Component

PC1=matrix_PC[,1]  #1st PC


df$PC1=PC1

head(df)


ggplot(df, aes(x = Date,y = PC1))+
  geom_line()+xlab("time")


#Presence of trend

#taking forward difference to detrend PC1 series


detrend_PC1=diff(df$PC1)

df_detrend=data.frame(df$Date[-1],detrend_PC1)

names(df_detrend)=c("Date","detrend_PC1")

head(df_detrend)


ggplot(df_detrend, aes(x =Date,y=detrend_PC1))+
  geom_line()+xlab("time")


#uneven fluctuation
```

```
#indication to volatility


####Stationarity Check


library(aTSA)

adf.test(detrend_PC1)

#Stationary


#For determining the order of the mean model and the GARCH model corresponding to PC1 series

par(mfrow=c(3,1))

pacf(detrend_PC1,main="PACF for PC1")

acf(detrend_PC1,main="ACF for PC1")

pacf(detrend_PC1^2,main="PACF for squared PC1")


#GARCH(p,q)

#p=1,q=1


library(rugarch)

#specify GARCH(1,1) with ARMA(2,1) mean equation


garchSpec= ugarchspec(variance.model=list(model="sGARCH",garchOrder=c(1,1)),

        mean.model=list(armaOrder=c(2,1)),distribution.model="std")


####estimate GARCH(1,1)
```

```
garchFit= ugarchfit(spec=garchSpec,data=detrend_PC1)


#QQ-plot and  histogram corresponding to the residuals series
par(mfrow=c(2,1))
plot(garchFit,which=8)
plot(garchFit,which=9)


#str(garchFit)
residuals=garchFit@ fit$residuals
df_detrend$sigma.t=sigma(garchFit)
head(df_detrend)
length(df_detrend$Date)


#Plotting the estimated sd
ggplot(df_detrend, aes(x=Date,y=sigma.t))+
  geom_line()+xlab("time")


#Volatility forecast
#Compute h-step ahead forecasts for h=1,2,...,15


Forecast=ugarchforecast(garchFit,n.ahead=15)
str(Forecast)
forecast_sigma=Forecast@ forecast$ sigmaFor
plot(Forecast,which=3,main="Forecast",lwd=3) #Plot of the forecasted volatility
```

ACKNOWLEDGEMENT:

I would like to express my special thanks of gratitude to the professors of my group Prof. A. Chattaopadhyay, Prof. A. Sengupta and Prof. R Bhattacharya as well as all my other professors- Professor S. Sen Roy, Professor G. Chattopadhyay, Professor M. Pal and Professor B. Ganguly who have guided me during the course of this project on topic "Modeling Financial Interval Time Series", over the Stock market values of TCS(NSE), from Jan,2015- Dec,2019", which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them.

REFERENCE:

The following books and websites have been consulted during the course of this project:

- Applied Multivariate Statistical Analysis by Johnson and Wichern
- https://faculty.washington.edu/ezivot/econ589/univariateGarch2012powerpoint.pdf
- C. Chatfield: The Analysis of Time Series – An Introduction
- https://www.sciencedirect.com/science/article/pii/S0275531919306348
- P.J.Brockwell & R.A. Davis: Introduction to Time Series Analysis & Forecasting
- A. Pankratz: Forecasting with Unique Box-Jenkins Model
- https://www.researchgate.net/publication/5113512_Modeling_and_Forecasting_Volatility_in_Indian_Capital_Markets
- https://www.researchgate.net/publication/331104930_Modeling_financial_interval_time_series/fulltext
- Ruey S. Tsay - Analysis of Financial Time Series, Third Edition (Wiley Series in Probability and Statistics)-John Wiley & Sons (2010)
- Gareth James, Daniela Witten, Trevor Hastie &Robert Tibshirani - An Introduction to Statistical Learning with Applications in R
- Ruey S. Tsay: Multivariate Time Series Analysis With R and Financial Applications.
- Bollerslev, T.(1986) ' Generalised autoregressive conditional heteroscedasticity ´ Journal of Econometrics, 31(3), 307-327.