How COVID 19 pandemic impact on stock prices and find out correlation between COVID 19 and stock price and develop machine learning or statistical model for predict the same.

Covid19 data collected from https://covid19.who.int/ provided by WHO. The number of confirmed cases over time is chosen for finding correlation. The proxy for the Indian stock market is the S&P CNX Nifty 50 index (N50) obtained from National Stock Exchange of India. Used the daily data for both variables from 31-01-2006 to 22-05-2020. Computed the continuously compounded daily returns by calclulating the difference in the logarithmic values of two consecutive prices: $r_{i,t} = \ln(P_{i,t} / P_{i,t-1}) \times 100$, where $r_{i,t}$ indicates the continuously compounded percentage daily returns for index i at time t , while $P_{i,t}$ indicates the price level of index i at time t .

Language used: R

Libraries :rmgarch,xts,psych,matrixTests,tseries,fDMA,urca,riskR,readxl,MTS

In just weeks, the Corona virus pandemic has shaved off nearly a third of the global market cap. The spread of the virus has triggered panic across the world and shaken the confidence of investors. On 28th February the Indian indices registered a 3.5% fall which was the second-biggest fall in the history of the Sensex. The stock market has historically been prone to fear psychoses, and this is one such instance. The automobile and healthcare industry are significant stakeholders in the Indian stock market. If their operations and production get affected due to the Coronavirus outbreak and China’s lockdown, it could lead to reduced investor faith in the market.

Methodology

Heteroskedasticity describes the irregular pattern of variation of an error term in a statistical model. GARCH processes, being autoregressive, depend on past squared observations and past variances to model for current variance. Here we go for a Multivariate approach MGARCH to estimate the time varying variance. AR MGARCH and VARMA MGARCH are employed.

$$r_{it} = m_{i0} + \sum_{j=1}^{3} m_{ij} r_{jt-1} + \varepsilon_{it},\ \varepsilon_{it} | \Omega_{it-1} \sim N(0, h_{it}), \quad i = 1,2,3, \tag{1}$$

$$\varepsilon_{it} = v_{it} h_{it}^{1/2},\ \ v_{it} \sim N(0,1) \tag{2}$$

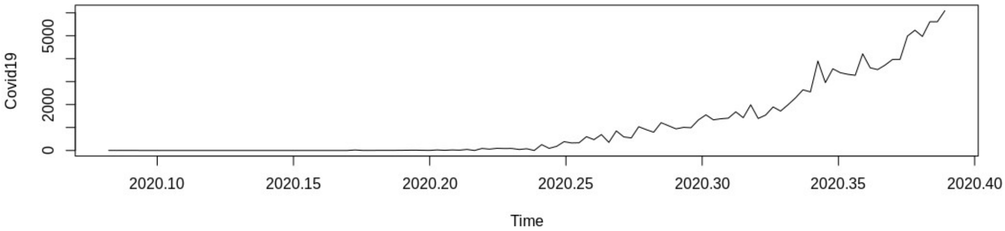
$$h_{it} = c_{it} + \sum_{j=1}^{3} \alpha_{ij} \varepsilon_{jt-1}^2 + \sum_{j=1}^{3} \beta_{ij} h_{jt-1} + d_i \varepsilon_{it-1}^2 I(\varepsilon_{it-1}), \tag{3}$$

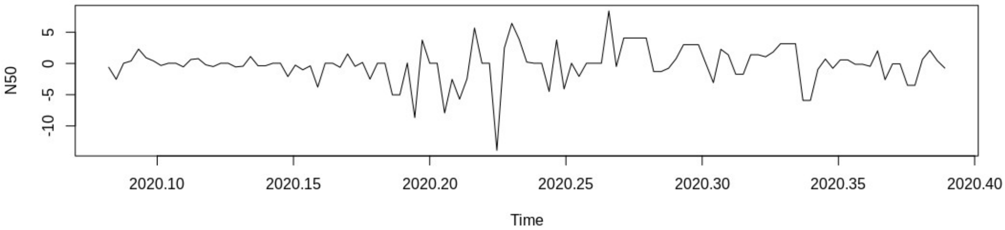
Ri,t : return of the series

Eq. (2) represents the relation between the error term and conditional variance. In Eq. (3), the dummy variable I is equal to one if $\varepsilon_{it-1} < 0$ and 0, otherwise. For this specification, a positive value of d implies that negative shocks tend to increase the variance more than the positive ones, referred to as the leverage effect(Asymmetric effect).

Both DCC and CCC are employed with each Asymmetric and Symmetric VARMA/AR MGARCH.

Daily evolution of Covid and N50 during the sample period taken into consideration is plotted below.





Implementation and Results

Descriptive statistics

```
[1] "Covid19"
   vars   n    mean     sd median trimmed    mad min  max range skew kurtosis     se
X1    1 113 1048.19 1566.3     88   727.1 130.47   0 6088  6088 1.55     1.36 147.34
[1] "N50"
   vars   n  mean   sd median trimmed mad   min max range skew kurtosis   se
X1    1 113 -0.17 2.95   0.03       0 1.2 -13.9 8.4  22.3   -1     4.14 0.28
[1] "Jarque-Bera Test for normality(Test1)"
[1] "Covid19"
  obs skewness kurtosis df statistic       pvalue
1 113 1.567133 4.434016  2  55.93514 7.142065e-13
[1] "N50"
  obs  skewness kurtosis df statistic pvalue
1 113 -1.013786 7.268644  2  105.1482      0
[1] "Lagrange Multiplier (LM) test for ARCH"
[1] "Covid19"

        Engle's LM ARCH Test

data:  as.vector(wti)
statistic = 100.3, lag = 10, p-value < 2.2e-16
alternative hypothesis: ARCH effects of order 10 are present

[1] "N50"

        Engle's LM ARCH Test

data:  as.vector(wti)
statistic = 11.553, lag = 10, p-value = 0.3161
alternative hypothesis: ARCH effects of order 10 are present
```

```
[1] "Covid19"

        Box-Ljung test

data:  resid(fit1)
X-squared = 93.877, df = 19, p-value = 6.771e-12

[1] "N50"

        Box-Ljung test

data:  resid(fit1)
X-squared = 34.069, df = 19, p-value = 0.01803
```

For both covid and N50 we obtained mean values significantly different from zero and Standard deviation larger than mean. The skewness and kurtosis statistics determine whether returns are normally distributed. Asymmetry of the probability distribution of a random variable about its mean is measured by skewness value. Positive skewness for Covid indicates skewed right data and negative skeweness of N50 indicates the opposite. For both timeseries the kurtosis values surpass three, indicating the leptokurtic distribution which shows heavy tails, giving a sign of outliers. The result obtained from Jarque and Bera (1987) test on all the return series are significant to reject normality null hypothesis. Any deviation from normal distribution with an expected value 0 for both skewness and excess kurtosis rises the JB statistic. According to Lagrange multiplier test Covid series exhibit an ARCH behavior and, therefore, estimation of a GARCH model is appropriate for modeling. Conforming to the results of Ljung-Box test Q (10) results, there is strong evidence for serial correlations at 1% significant level for residuals in Covid and N50. Phillips-Perron(PP) test, Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test and Augmented Dickey-Fuller(ADF) test are the unit root tests we applied to confirm if the timeseries is stationary. Unit root is a characteristic of a time series that makes it non-stationary. If the absolute value of the test statistic is greater than the critical value, you can declare statistical significance and reject the null hypothesis. The test scores clearly reveals the stationarity of all return series at 1% significance level.

	Covid	**N50**
ADF	5.5479***	-6.8215***
PP	4.264***	-9.2846***
KPSS	0.5343***	0.1246

Cointegration tests are used to find out if there is a long run correlation between 2 or more time series. Cointegration tests identify scenarios where two or more non-stationary time series are integrated together in a way that they cannot deviate from equilibrium in the long term. The most popular cointegration tests including Engle-Granger, the Johansen Test, and the Phillips-Ouliaris test are conducted to find out cointegration. The Engle-Granger test is conducted with and without trend.

Engle-Granger test:

	Without Trend				With trend					
		1pct	5pct	10pct	test statistics		1pct	5pct	10pct	test statistics
Covid~N50	tau1	-2.58	-1.95	-1.62	3.6908	tau3	-3.99	-3.43	-3.13	1.4287
						phi2	6.22	4.75	4.07	10.5334
						phi3	8.43	6.49	5.47	9.6372

Johansen Test:

```
######################
# Johansen-Procedure #
######################

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.3041748 0.1478121

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 | 17.75  6.50  8.18 11.65
r = 0  | 58.01 15.66 17.95 23.52
```

Phillips-Ouliaris test

```
        Phillips-Ouliaris Cointegration Test

data:  dat
Phillips-Ouliaris demeaned = 4.4619, Truncation lag parameter = 3, p-value = 0.15
```

Engle-Granger method shows the stationarity of the residuals in ADF test without trend. In Johansen cointegration test the first hypothesis,r=0, tests whether there is cointegration or not. Since the test statistic beats the 1% level significantly (58.01 > 23.52), we have confirmation of cointegration. But in Phillips-Ouliaris test we could’t obtain p-value less than 0.05. Overall since we cannot trust the results produced by small sample size, we don’t have enough evidence to prove long run correlation between Covid and N50.

EMPIRICAL RESULTS

We employ the MGARCH approach using AR-GARCH and VARMA-GARCH models and evaluate the results estimated where order of AR, order of GARCH terms and order of ARCH terms in AR-GARCH are 1,1,1 respectively and the order of VAR, order of MA, order of GARCH terms and order of ARCH terms in VARMA-GARCH are 1,1,1,1 respectively. We observe positive and significant coefficients in mean model equations of the Covid, N50 with and without asymmetry for all the models. The Arch (alpha) terms gives indications for short-term volatility persistence and volatility clustering in the the estimated coefficients for GARCH. All variables provides solid proof of short-term persistence in volatility. Alpha catches the effect of innovations on future volatility. (alpha+beta) -indicates how long the effect persist (or how fast it decays over time). In each of the models alpha+beta is high, so the effect decays slowly. As per coefficients estimated we observe that long term volatility persistence (beta) is greater than the short term volatility persistence in the case of all variables in all the models.

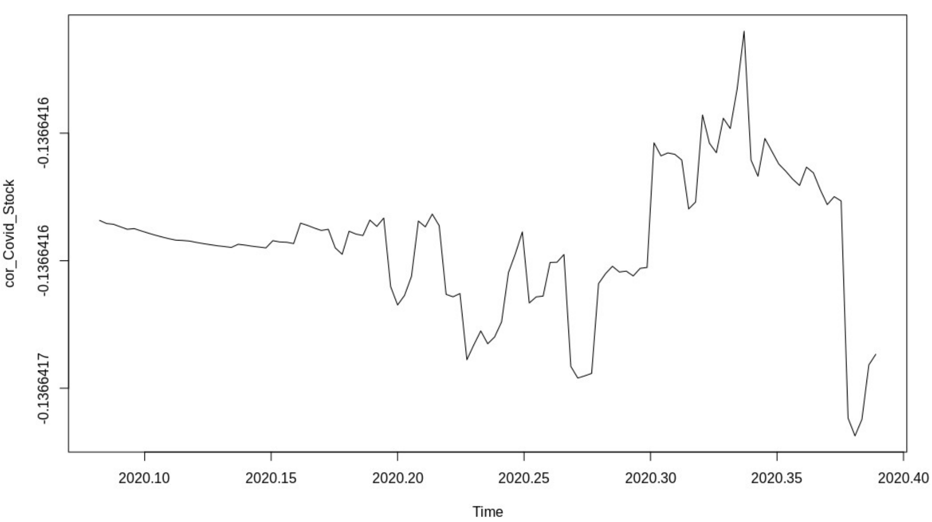
If the leverage term (Gamma coefficient) is significant and positive, it indicates the presence of an asymmetric behavior. But we didnt observe any significant gamma terms indicating that timeseries doesn’t show any asymmetric behaviour. From DCC models we obtain positive and statistically significant DCC(A) term and DCC(B) term at the 1% significance level Mean reversion in time-varying conditional correlations is indicated by the DCC(A) and DCC(B) values with sum lower than 1. The AIC ,BIC ,SBC and Hannan-Quinn information criterion for both with and without applying asymmetry show that the Dynamic Conditional Correlation model outperform the Constant Conditional Correlation model for VARMA-GARCH but CCC is outperforms DCC in AR-GARCH. Overall the DCC-VARMA-GARCH without asymmetry is the best model.

DCC-VARMA-GARCH model without leverage effect estimates non constant Conditional correlation during sample period and it is plotted below. Study reveals that correlation coefficients are time-varying in pair-wise estimations. From all the CCC models we obtained indistinguishable values for constant conditional correlations. Negative correlation between N50 and Covid indicates that they move in opposite directions. In the initial stages due to reduced imports from China the prices of available good rise and lack of supply and demand can cause investors to be cautious and not invest further or withdraw from the Indian markets in the foreseeable future. The Indian stock market after recovering its losses on March 2, again ended on a negative with recent cases of coronavirus being reported in India.

AR (1)-multivariate GARCH(1,1) models.

AR(1)-GARCH(1,1)	Without Asymmetry		With Asymmetry	
	DCC	CCC	DCC	CCC
AR(1) Mean Equation				
Covid-constant	0.97843	0.99977	82.97347***	83.078691***
Covid-ar1	1***	1.00000***	0.98896***	0.988956***
N50-constant	-0.23669	-0.23669	-0.28108	-0.281084
N50-ar1	0.20174	0.20174	0.16961	0.169612
GARCH variance equation				
[Covid].omega	12.3721	12.28182	11.48149	11.466383
[Covid].alpha1	0.45102***	0.45101***	0.56656***	0.56656***
[Covid].beta1	0.54798***	0.54799***	0.53177***	0.531763***
[Covid].gamma1			-0.19865	-0.198647
[N50].omega	0.21709	0.21709	0.18177	0.181765
[N50].alpha1	0.21062***	0.21062***	0.132	0.132
[N50].beta1	0.7883***	0.78838***	0.80261***	0.802614***
[N50].gamma1			0.12877	0.128772
[Joint]dcca1	0		0	
[Joint]dccb1	0.9178		0.9159	
[Joint]C1		-0.1013311		-0.07617145
Akaike	16.728	16.69	16.751	16.714
Bayes	17.042	16.956	17.113	17.028
Shibata	16.705	16.673	16.721	16.691
Hannan-Quinn	16.85	16.798	16.898	16.841

Conditional Correlation between Covid and N50 obtained from VARMA-GARCH-DCC without asymmetry



VARMA (1,1)–multivariate GARCH(1,1) models.

VARMA(1,1)-GARCH(1,1) DCC Without Asymmetry		
Residuals cov-matrix	Covid	**N50**
Covid	48232.1075	-105.9549
N50	-105.9549	8.6543
VAR coefficient matrix	Covid	N50
Covid	1.0483041	-48.9565304
N50	7.30228E-05	-0.0622884
VMA coefficient matrix	Covid	N50
Covid	0.67690511	-51.6832
N50	0.00178	-0.0841
Constant	Covid	N50
	-3.504045	-0.25904

GARCH variance equation	
[Covid].omega	11.88107
[Covid].alpha1	0.43631***
[Covid].beta1	0.56269***
[N50].omega	0.21743
[N50].alpha1	0.21291***
[N50].beta1	0.78609***
[Joint]dcca1	0
[Joint]dccb1	0.91761
Akaike	16.651
Bayes	17.013
Shibata	16.621
Hannan-Quinn	16.798
Log L	-925.7873

VARMA(1,1)-GARCH(1,1) DCC With Asymmetry		
Residuals cov-matrix	Covid	**N50**
Covid	48232.1075	-105.9549
N50	-105.9549	8.6543
VAR coefficient matrix	Covid	N50
Covid	1.0483041	-48.9565304
N50	7.30228E-05	-0.0622884
VMA coefficient matrix	Covid	N50
Covid	0.67690511	-51.6832
N50	0.00178	-0.0841
Constant	Covid	N50
	-3.504045	-0.25904

GARCH variance equation	
[Covid].omega	12.117925
[Covid].alpha1	0.559833***
[Covid].beta1	0.543439***
[Covid].gamma1	0.208544
[N50].omega	0.18164
[N50].alpha1	0.133787
[N50].beta1	0.799092***
[N50].gamma1	0.132243
[Joint]dcca1	0
[Joint]dccb1	0.914818
Akaike	16.652
Bayes	17.062
Shibata	16.614
Hannan-Quinn	16.819
Log L	-923.8411

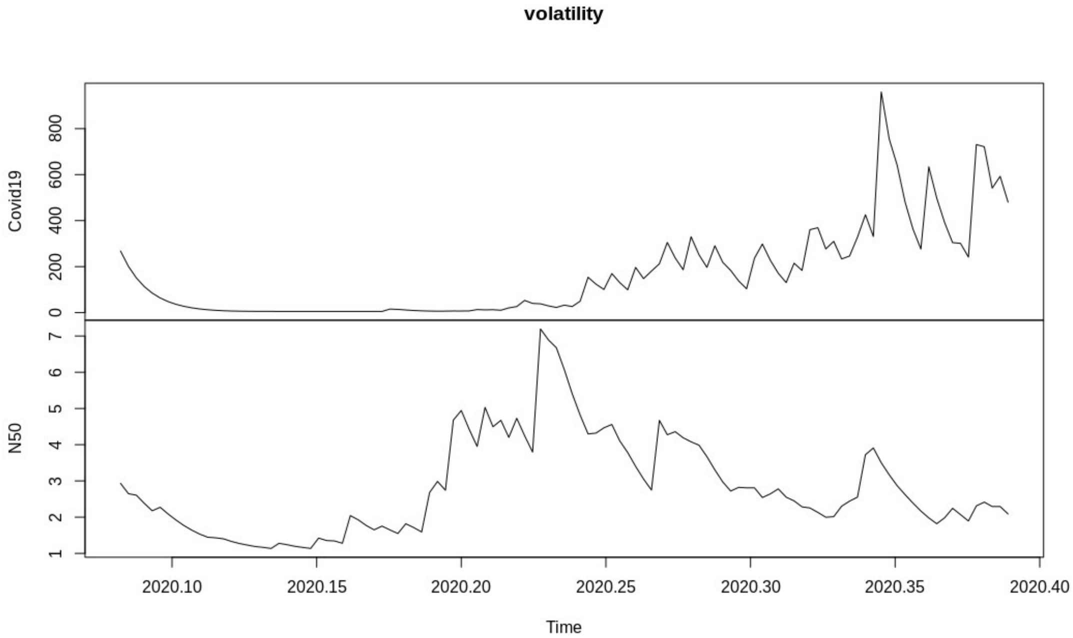
VARMA(1,1)-GARCH(1,1) CCC Without Asymmetry		
Residuals cov-matrix	**Covid**	**N50**
Covid	48232.1075	-105.9549
N50	-105.9549	8.6543
VAR coefficient matrix	Covid	N50
Covid	1.0483041	-48.9565304
N50	7.3E-05	-0.0622884
VMA coefficient matrix	Covid	N50
Covid	0.67690511	-51.6832
N50	0.00178	-0.0841
Constant	Covid	N50
	-3.504045	-0.25904

GARCH variance equation	
[Covid].omega	79.788252
[Covid].alpha1	0.336831***
[Covid].beta1	0.662169***
[N50].omega	0.22131
[N50].alpha1	0.206437***
[N50].beta1	0.792563***
[Joint]C1	-0.02875452
Akaike	17.285
Bayes	17.454
Shibata	17.277
Hannan-Quinn	17.353
Log L	-969.5777

VARMA(1,1)-GARCH(1,1) CCC With Asymmetry		
Residuals cov-matrix	**Covid**	**N50**
Covid	48232.1075	-105.9549
N50	-105.9549	8.6543
VAR coefficient matrix	Covid	N50
Covid	1.0483041	-48.9565304
N50	7.3E-05	-0.0622884
VMA coefficient matrix	Covid	N50
Covid	0.67690511	-51.6832
N50	0.00178	-0.0841
Constant	Covid	N50
	-3.504045	-0.25904

GARCH variance equation	
[Covid].omega	84.571089
[Covid].alpha1	0.352634***
[Covid].beta1	0.660787***
[Covid].gamma1	-0.028842
[N50].omega	0.182471
[N50].alpha1	0.132907
[N50].beta1	0.803442***
[N50].gamma1	0.125301
[Joint]C1	-0.02389112
Akaike	17.307
Bayes	17.524
Shibata	17.296
Hannan-Quinn	17.395
Log L	-968.8521

volatility of Covid ,N50 from VARMA-GARCH-DCC model



We also forecast the correlation that might occur in the next 30 days(starting from may 22 2020 as we considered data till may 22 2020) between Covid and N50.

[1] "Forecast of correlation between Covid19 and N50 for next 30 days"

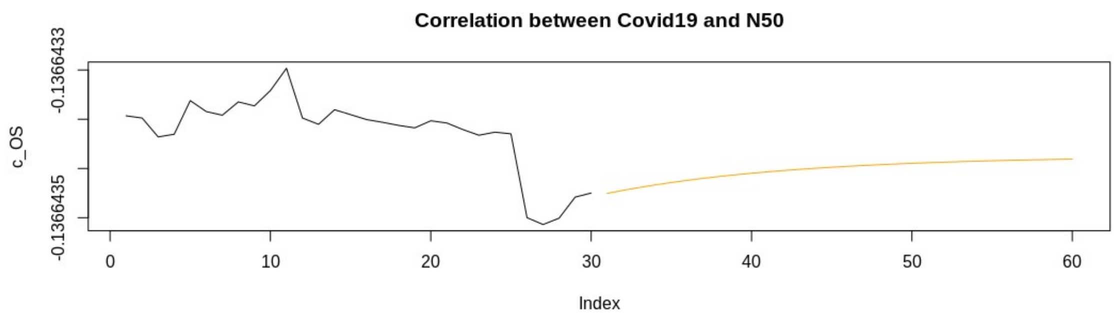
[1] -0.1366435 -0.1366435 -0.1366435 -0.1366435 -0.1366435 -0.1366435 -0.1366435 -0.1366435

[9] -0.1366435 -0.1366435 -0.1366435 -0.1366435 -0.1366435 -0.1366434 -0.1366434 -0.1366434

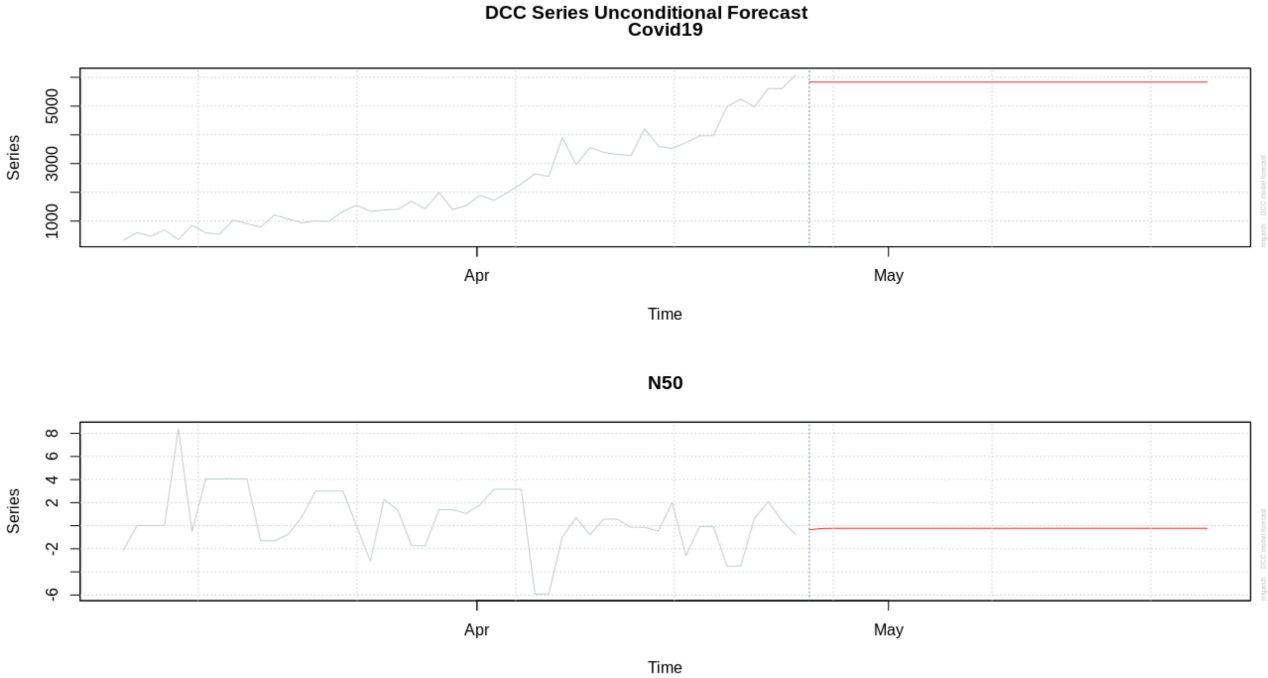
[17] -0.1366434 -0.1366434 -0.1366434 -0.1366434 -0.1366434 -0.1366434 -0.1366434 -0.1366434

[25] -0.1366434 -0.1366434 -0.1366434 -0.1366434 -0.1366434 -0.1366434

Correlation forecast between Covid and N50



DCC volatility(sigma) unconditional forecast



This forcast predicts the increased volatility to be experienced in stock market and Covid cases