How crude oil crisis impact on stock index and find out correlation between crude oil and stock index and develop machine learning or statistical model for predict the same.

Crude oil data collected from Multi Commodity Exchange of India Limited. OIL prices are measured in Indian rupees per barrel. 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 10-07-2006 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

The high demand for Crude oil in developing countries has resulted in increased oil prices and policy makers have shifted their interest to evaluate the correlation between Crude oil prices and Stock market. Since India emerged as an important part in Oil market by replacing Japan to become the third largest Oil Importer, understanding the correlation in the context of India gained importance more than ever before. The price of oil has sunk to levels not seen since 2002 as demand for crude collapses amid the coronavirus pandemic. Energy stocks constituted around 12.5 percent of Nifty50. As an economy that depends on oil imports to meet 80% of its oil needs, it has come as a welcome relief for India’s already struggling economy, which also affected the stock market. Listed companies which reply on crude oil for their production and transportation will benefit from the drop in crude oil prices. Subsequently, this will improve their standing in the stock 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.

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

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

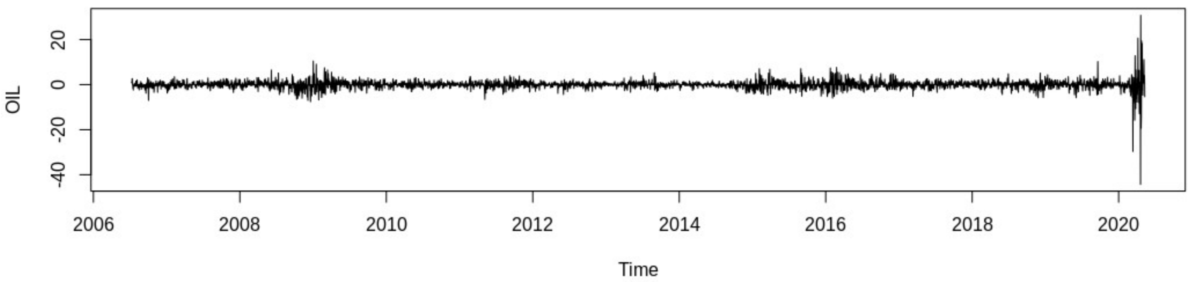
$$h_{it} = c_{it} + \sum_{j=1}^{3} \alpha_{ij} \varepsilon_{jt-1}^{2} + \sum_{j=1}^{3} \beta_{ij} h_{jt-1} + d_i \varepsilon_{it-1}^{2} I(\varepsilon_{it-1}), \tag{3}$$

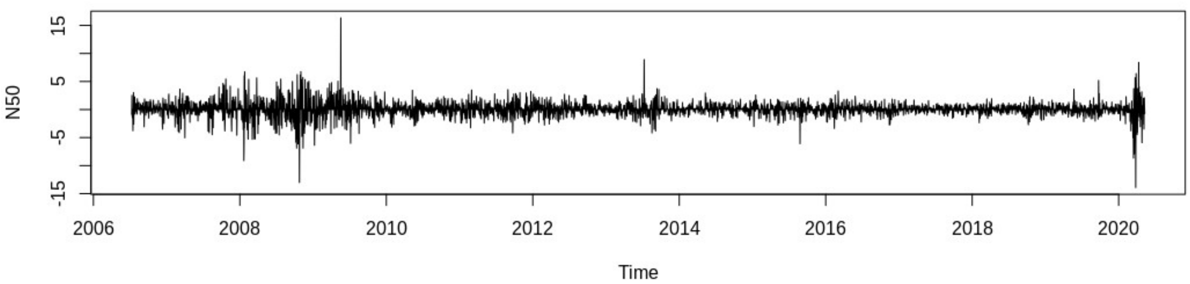
Ri,t : return of the series

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 Oil and N50 during the sample period taken into consideration is plotted below.





Implementation and Results

Descriptive statistics

```
[1] "OIL"
   vars    n  mean   sd median trimmed  mad    min   max range  skew kurtosis   se
X1    1 5051 -0.03 1.83  -0.04   -0.01 0.71 -44.31 30.76 75.07 -2.36   113.14 0.03
[1] "N50"
   vars    n mean   sd median trimmed  mad   min   max range  skew kurtosis   se
X1    1 5051 0.03 1.19   0.03    0.05 0.47 -13.9 16.33 30.24 -0.28    20.61 0.02
[1] "Jarque-Bera Test for normality(Test1)"
[1] "OIL"
   obs  skewness kurtosis df statistic pvalue
1 5051 -2.357718 116.1884  2   2700991      0
[1] "N50"
   obs   skewness kurtosis df statistic pvalue
1 5051 -0.2762003 23.61989  2  89546.83      0
[1] "Lagrange Multiplier (LM) test for ARCH"
[1] "OIL"

        Engle's LM ARCH Test

data:  as.vector(wti)
statistic = 905.13, 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 = 408.75, lag = 10, p-value < 2.2e-16
alternative hypothesis: ARCH effects of order 10 are present

        Box-Ljung test

data:  resid(fit1)
X-squared = 242.79, df = 19, p-value < 2.2e-16

[1] "N50"

        Box-Ljung test

data:  resid(fit1)
X-squared = 85.612, df = 19, p-value = 1.968e-10
```

For both Crude oil and N50 we obtained mean values significantly different from zero and Standard deviation larger than mean. Asymmetry of the probability distribution of a random variable about its mean is measured by skewness value. Negative skewness for all series indicates the negatively skewed or skewed left data. It means the longer tail on the left side. Value less than -1 indicates high skewness for Crude oil. For both timeseries the kurtosis values surpass three, indicating the leptokurtic distribution which shows heavy tails, giving a sign of large 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 all return 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 OIL 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. As per the ADF and PP test scores the return serieses are stationary at 1% significance level.

	Crude OIL	**N50**
ADF	-52.2151***	-49.4316***
PP	-68.2874***	-70.5177***
KPSS	0.0513	0.0563

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
OIL~N50	tau1	-2.58	-1.95	-1.62	-52.4038	tau3	-3.96	-3.41	-3.12	-52.3935
						phi2	6.09	4.68	4.03	915.0254
						phi3	8.27	6.25	5.34	1372.537

Johansen Test:

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

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.3527995 0.3264751

Values of teststatistic and critical values of test:

            test 10pct  5pct  1pct
r <= 1 | 1995.52  6.50  8.18 11.65
r = 0  | 4192.33 15.66 17.95 23.52
```

Phillips-Ouliaris test:

```
        Phillips-Ouliaris Cointegration Test

data:  dat
Phillips-Ouliaris demeaned = -5258.4, Truncation lag parameter = 168, p-value = 0.01
```

Engle-Granger method showing the stationarity of the residuals in ADF test indicates that timeseries is cointegrated. 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 (4192.33 > 23.52), we have solid confirmation of cointegration. A small p-value (typically 0.05) and high test statistics gives all indications to reject the null hypothesis in Phillips-Ouliaris test

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 N50, OIL 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.

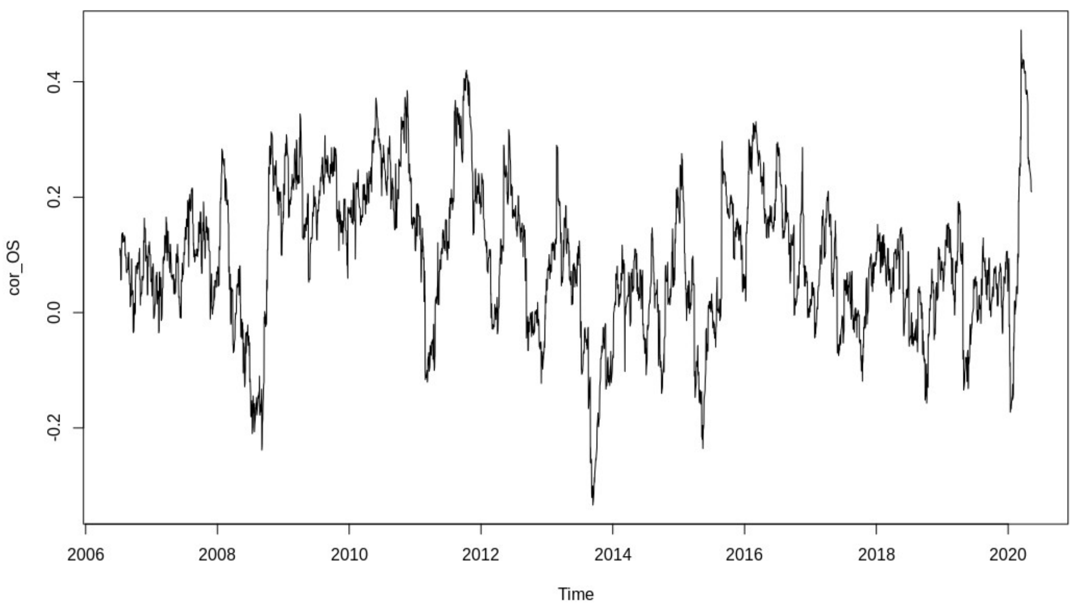
If the leverage term (Gamma coefficient) is significant and positive, it indicates the presence of an asymmetric behavior. Positive and significant gamma coefficients in Asymmetric DCC GJR-GARCH model and Asymmetric CCC GJR-GARCH models for OIL and N50 in AR-GARCH and VARMA-GARCH indicate that the effect on volatility by negative shocks is more intense than positive shocks. Information asymmetry occuring between investors in the trade leads to this leverage effect. 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 outsail the Constant Conditional Correlation model.

DCC-VARMA-GARCH model with leverage effect estimates non constant Conditional correlation during July 2006 to April 2020 and it is plotted below. Study reveals that correlation coefficients are time-varying in pair-wise estimations. The profit of diversifying and financing in the Indian stock and oil commodity market is suggested by correlations. From the CCC models we obtained indistinguishable values for constant conditional correlations. CCC models give a positive correlation between N50 and oil indicates that they move in same directions. In DCC models we observed the correlation going below -0.2. When it comes to investing it doesn’t mean that oil should be avoided. This help investors to include a mix of investments with negative correlation to stock market to reduce portfolio risk.

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

AR(1)-GARCH(1,1)	Without Asymmetry		With Asymmetry	
	DCC	CCC	DCC	CCC
AR(1) Mean Equation				
OIL-constant	0.017849	0.017852	-0.010634	-0.01063
OIL-ar1	0.209250***	0.209251***	0.209797***	0.209798***
N50-constant	0.058008***	0.057949***	0.034587***	0.034606***
N50-ar1	0.038042***	0.037991***	0.048192***	0.048352***
GARCH variance equation				
[OIL].omega	0.013059***	0.013058***	0.014869***	0.014868***
[OIL].alpha1	0.067994***	0.067993***	0.033096***	0.033094***
[OIL].beta1	0.931006***	0.931007***	0.932965***	0.932966***
[OIL].gamma1			0.063563***	0.063564***
[N50].omega	0.006765***	0.006779***	0.007630***	0.00764***
[N50].alpha1	0.059666***	0.059683***	0.019490***	0.019501***
[N50].beta1	0.938151***	0.938126***	0.938813***	0.938755***
[N50].gamma1			0.077291***	0.077366***
[Joint]dcca1	0.015103		0.969087	
[Joint]dccb1	0.972784		0.017146	
[Joint]C1		0.10964		0.1003821
Akaike	6.0239	6.039	5.9892	6.0048
Bayes	6.0407	6.0532	6.0085	6.0216
Shibata	6.0239	6.039	5.9891	6.0047
Hannan-Quinn	6.0298	6.044	5.996	6.0106
Log L	-15200.31	-15240.47	-15110.63	-15152.02

Conditional Correlation between Crude oil and N50 obtained from VARMA-GARCH-DCC with asymmetry



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

VARMA(1,1)-GARCH(1,1) DCC Without Asymmetry		
Residuals cov-matrix	**Crude oil**	**N50**
Crude oil	3.3155769	0.3133206
N50	0.3133206	1.4216702
VAR coefficient matrix	Crude oil	N50
Crude oil	-0.4070167	0.2979102
N50	0.05804937	-0.0418952
VMA coefficient matrix	Crude oil	N50
Crude oil	-0.4732208	0.30221218
N50	0.07114545	-0.05509632
Constant	Crude oil	N50
	-0.05937907	0.03719171

GARCH variance equation	
[Crude oil].omega	0.013027***
[Crude oil].alpha1	0.068154***
[Crude oil].beta1	0.930846***
[N50].omega	0.006727***
[N50].alpha1	0.059450***
[N50].beta1	0.938363***
[Joint]dcca1	0.015208
[Joint]dccb1	0.972747
Akaike	6.0198
Bayes	6.0392
Shibata	6.0198
Hannan-Quinn	6.0266
Log L	-15188.1

VARMA(1,1)-GARCH(1,1) DCC With Asymmetry		
Residuals cov-matrix	**Crude oil**	**N50**
Crude oil	3.3155769	0.3133206
N50	0.3133206	1.4216702
VAR coefficient matrix	Crude oil	N50
Crude oil	-0.40701669	0.2979102
N50	0.05804937	-0.0418952
VMA coefficient matrix	Crude oil	N50
Crude oil	-0.4732208	0.30221218
N50	0.07114545	-0.05509632
Constant	Crude oil	N50
	-0.05937907	0.03719171

GARCH variance equation	
[Crude oil].omega	0.01489***
[Crude oil].alpha1	0.032921***
[Crude oil].beta1	0.933027***
[Crude oil].gamma1	0.063297***
[N50].omega	0.007664***
[N50].alpha1	0.017758***
[N50].beta1	0.938786***
[N50].gamma1	0.081436***
[Joint]dcca1	0.017395
[Joint]dccb1	0.968879
Akaike	5.9844
Bayes	6.0064
Shibata	5.9844
Hannan-Quinn	5.9921
Log L	-15096.59

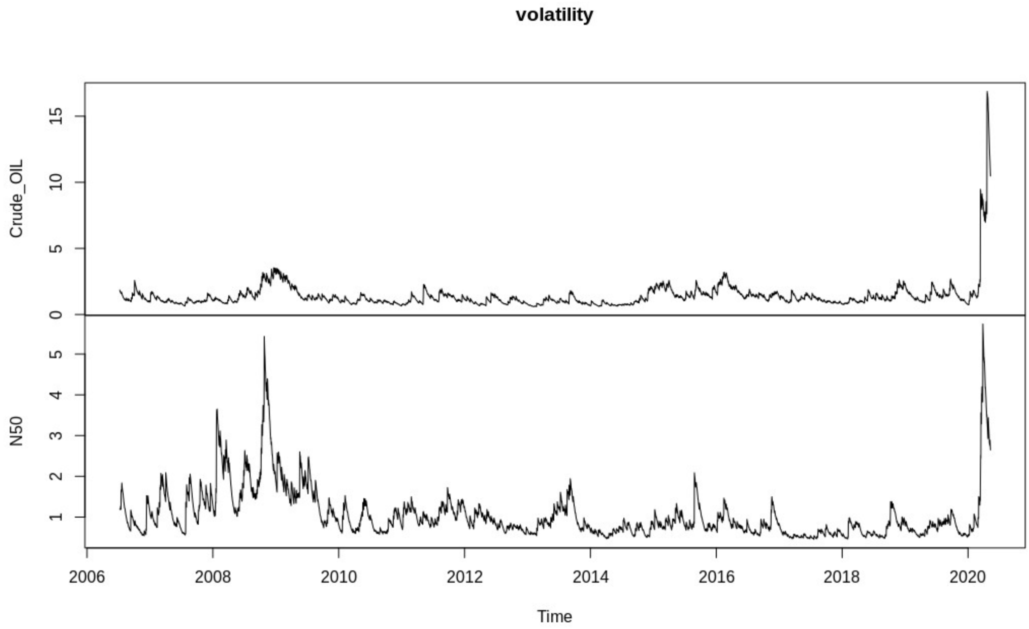
VARMA(1,1)-GARCH(1,1) CCC Without Asymmetry		
Residuals cov-matrix	**Crude oil**	**N50**
Crude oil	3.3155769	0.3133206
N50	0.3133206	1.4216702
VAR coefficient matrix	Crude oil	N50
Crude oil	-0.407	0.2979
N50	0.058	-0.0419
VMA coefficient matrix	Crude oil	N50
Crude oil	-0.4732	0.3022
N50	0.0711	-0.0551
Constant	Crude oil	N50
	-0.05937907	0.03719171

GARCH variance equation	
[Crude oil].omega	0.013063***
[Crude oil].alpha1	0.065695***
[Crude oil].beta1	0.933305***
[N50].omega	0.006682***
[N50].alpha1	0.058999***
[N50].beta1	0.938825***
[Joint]C1	0.107979
Akaike	6.0654
Bayes	6.0745
Shibata	6.0654
Hannan-Quinn	6.0686
Log L	-15096.59

VARMA(1,1)-GARCH(1,1) CCC With Asymmetry		
Residuals cov-matrix	**Crude oil**	**N50**
Crude oil	3.3155769	0.3133206
N50	0.3133206	1.4216702
VAR coefficient matrix	Crude oil	N50
Crude oil	-0.407	0.2979
N50	0.058	-0.0419
VMA coefficient matrix	Crude oil	N50
Crude oil	-0.4732	0.3022
N50	0.0711	-0.0551
Constant	Crude oil	N50
	-0.05937907	0.03719171

GARCH variance equation	
[Crude oil].omega	0.016004***
[Crude oil].alpha1	0.033375***
[Crude oil].beta1	0.935004***
[Crude oil].gamma1	0.058059***
[N50].omega	0.007538***
[N50].alpha1	0.020171***
[N50].beta1	0.939363***
[N50].gamma1	0.074761***
[Joint]C1	0.098202***
Akaike	6.0306
Bayes	6.0423
Shibata	6.0306
Hannan-Quinn	6.0347
Log L	-15221.34

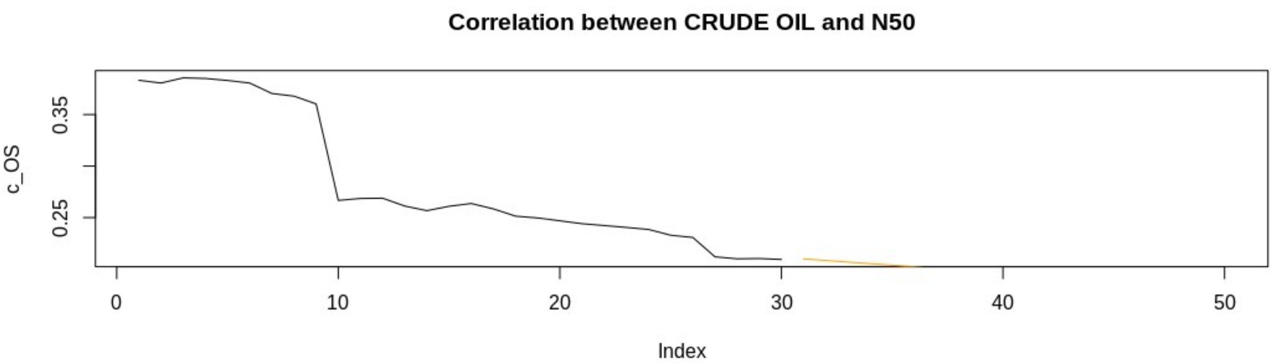
volatility of OIL ,N50 from VARMA-GARCH-DCC model



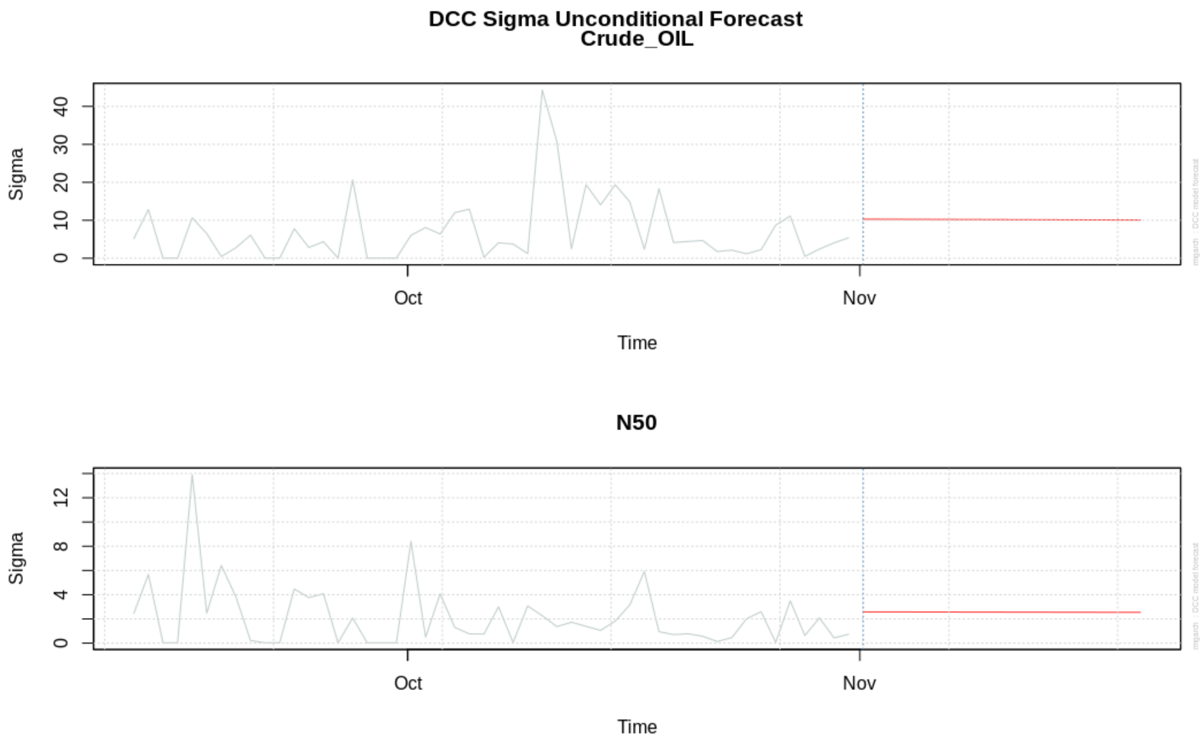
We find that among symmetric models the DCC model out- performs the Constant CC model in evaluating dynamic correlations by investigating different types of MGARCH models. We also forecast the correlation that might occur in the next 20 days(starting from may 22 2020 as we considered data till may 22 2020) between OIL and N50.

```
[1] "Forecast of correlation between Oil and N50 for 20 days"
 [1] 0.2099989 0.2084284 0.2068795 0.2053519 0.2038452 0.2023592 0.2008936 0.1994481 0.1980225
[10] 0.1966164 0.1952296 0.1938618 0.1925129 0.1911824 0.1898702 0.1885760 0.1872996 0.1860407
[19] 0.1847991 0.1835745
```

Correlation forecast between OIL and N50



DCC volatility(sigma) unconditional forecast



In addition to this, optimal hedging ratios based on risk measures of Crude oil in Stock market is also evaluated. Values of optimal hedging ratio for each risk measure at different probabilities of interest is given below.

	1%	5%
StD	0.09343387	0.09343387
VaR	0.15028346	0.20968973
EL	19.99994991	19.99994991
ELD	0.15293471	0.15293471
ES	0.14117580	0.12174687
SDR	0.15059393	0.14401324
EVaR	0.14150920	0.14342451
DEVaR	0.16891025	0.15895861
ENT	0.35383576	0.35383576
DENT	0.35137253	0.35137253
ML	0.41691029	0.41691029

This is the ratio that tells us the percentage of asset we should hedge. Though expected loss is showing higher value for all other factors Oil shows low values.