Investigating long-run cointegration and lead lag relationship between spot and future markets of energy commodities

Sunil Bhardwaj and Sameer Gupta

Abstract
The study tries to investigate the cointegration and causality relationship between the spot and future prices of crude oil and natural gas traded on Multi Commodity Exchange (MCX) of India. The study has used daily closing spot and future prices of the commodities under study from the period 2005 to 2020. The data is converted into stationary time series which is a mandatory condition for time series analysis. The results of Johansen Co-integration test reveals that there is long-run cointegration between the spot and future prices of both the commodities. Appropriate lag length was selected by using Schwarz information Criterion (SIC) lag length criteria. The estimates of VEC Granger Causality Test/Block Exogeneity Wald Test show a bi-directional causality relationship in the prices of crude oil and natural gas. The spot and future markets of energy commodities are equally efficient in adjusting the new information in their equilibrium prices which reveals the absence of any lead-lag relationship between the markets.

Keywords: Cointegration, causality, equilibrium price, energy commodities

Introduction
An extensive research investigating causality relationship in spot and future markets of energy commodities has been carried out by several researches and is well documented. Price discovery in energy commodities is a critical research area in energy economics because of greater importance of crude oil and natural gas in global economy as well as high price volatility in these commodities. Future markets perform two important function of price discovery and risk hedging. The studies done by Tse et al., 2006; Tse and Xiang 2005 found future markets as a locus of innovation which leads the spot markets in price discovery process. On the other hand Chiou-Wei et al. 2019 were empirically convinced that spot markets leads in price discovery process or Bekiros and Diks 2008; Kaufmann and Ullman 2009 [3, 6, 14, 15] noticed a feedback relationship across spot and future markets of crude oil and natural gas. The present study has done an extensive review of literature. However, we found no study which has analyzed the cointegration and lead-lag relationship across two markets of energy commodities in the long run in recent times.

Literature review
Raju and Shirodkar (2020) [18] have investigated the lead-lag relationship between the spot and futures markets of energy sector stocks on which single stock futures (SSFs) are available and found that future markets leads in price discovery process. Zavadska et al. (2018) [21] analyzed the behavior and importance of crude oil in the global economy with a more focus on investigating the lead-lag relationship between the spot and future markets. The researchers have done an extensive review of literature on the dynamic relationship between the spot and future markets of crude oil, volatility in these markets as well as on the efficiency of price discovery mechanism. The study found no conclusive argument indicating which market dominates the price discovery process in crude oil especially during the period of uncertainty. Singh and Sharma (2017) [19] has investigated the cointegration and causality between in Indian gold and crude oil spot market and has found that gold prices granger cause price discovery in crude oil prices. Aggarwal et al. (2014) [1] examined the price discovery and risk management in six agriculture commodities viz. pepper, soya oil, castor seed, sugar, wheat, rubber and two non-agriculture commodities i.e. crude oil and gold over a period from 2003 to 2014.
The test estimated of the study revealed that the price discovery takes place in future markets of these commodities. The study also highlighted certain bottleneck for improving risk management efficiency of future markets like high settlement cost, unreliability of warehouse receipts, few delivery centers etc. Chhatwal and Puri (2014) [10] examined the causality relationship between the spot and future prices of crude oil over a period from May 2005 to December 2012. The period of the study was divided into three phases i.e. pre crisis period from (May 2005 to August 2008, crisis period from (September 2008 to December 2010) and post crisis period (January 2011 to December 2012). The spot and future prices of the crude oil were analyzed using Johansen cointegration test and granger causality test. The test estimates of the study predicted a unidirectional relationship between the spot and future prices during pre-crisis and post crisis period whereas during crisis period bidirectional relationship between the two prices was estimated. Mahalik et al. (2014) [17] used vector error correction model (VECM) and bivariate exponential GARCH model (EGARCH) to study price discovery in future spot and future indices of Multi-Commodity Exchange. The VECM estimates that the price discovery takes place in future market (LAGRIFP, LENERGYFP and LCOMDEXFP) in agriculture, energy and aggregate commodity indices. In case of metal index no relationship is found between the spot (LMETALSP) and future index (LMETALFP), Kumar and Pandey (2013) [16] examined the log run as well as short run market in Indian future market of seven non-agriculture commodities (gold, silver, copper, zinc, aluminum, natural gas and crude oil) and four agriculture commodities (caster seed, gaur seed, soybean and corn) using different models of asset pricing like ECM Model and ECM-GARCH model. The study concluded that in the long run spot and future prices of the commodities are cointegrated but in near month future contracts the cointegration between the spot and future prices is not evident due to low trading volume. The study found that thinly trade contracts are unable to forecast future spot price which means price discovery takes place in future market only when liquidity of contract is high. Agnihotri and Sharma (2011) [18] analyzed the relationship between the future and spot prices zinc, natural gas, chana and zeera by employing correlation, regression model and standard deviation. The tests predicted a clear delinking between the spot and future prices of the commodities taken for the study. The study clearly mentions that the degree of convergence of spot prices towards the future prices is insignificant. Therefore, it is concluded that the future markets are unable to hedge the risk arising due the volatility in the spot market. Zhang and Wei (2010) [22] investigated the price discovery, cointegration and causality between the gold market and the crude oil market. The correlation results revealed that the gold market and crude oil market are positively correlated with each other. The study noticed long run equilibrium between the two market and a causality running from the crude oil price to the gold price which means an inter-commodity lead-lag relationship. As far as the impact on the global economy is concerned, a higher impact of crude oil on the global economy in comparison to gold is estimated both by permanent transitory (PT) model and by the information share (IS) model. The study done by Kaufmann and Ullman (2009) [14, 15] outlined a lead-lag relationship in the crude oil sector in international markets and show that both futures and spot markets contribute in the process of price innovation in the oil sector, as well as that market participants work together rather than independently in the formation of prices. Bekiros and Diks (2008) [5, 6] examined the linear and non-linear causality between the future and spot prices of crude oil contracts traded in New York Mercantile Exchange (NMEX) of having maturity between 1 to 4 month over two time horizons, one October 1991 to October 1999 and two November 1999 to October 2007. The study has employed Diks and Panchenko’s non-parametric test for non-linear causality after controlling cointegration besides granger causality test. GARCH BEKK model was employed to test non-linear non-causality for conditional heteroscedasticity. The empirical results of the study revealed bi-directional causal relationship between the spot and future prices of crude oil. It also suggested that there is no consistency of lead-lag relationship which can change over time if non-linear effect is accounted for. Ballinger et al. (2004) [19] examined the price discovery in the spot ad future markets of crude oil in USA. The study noticed the existence of long run cointegration between the spot and future prices of the analyzed commodity. The information transmission takes place form future to the spot market and gets adjusted in the future spot price. The study also highlighted that there is increased volatility in the spot market after the introduction of crude oil future trading. Hammoudeh et al. (2003) [13] tested for causality on three major energy commodities traded inside and outside the U.S. from 1986 to 2001. Their results identified a bi-directional causality in the heating oil and gasoline markets; in the case of crude oil, the futures market leads on price innovation. Asche and Guittormsen (2002) [3] analyzed the lead-lag relationship between the spot and future prices of gas oil. The study highlighted the limitations of bivariate models and Eagle Granger framework and has used multivariate model to test the relationship. It was concluded that future price lead spot price. Also future contracts with longer expiration lead the future contracts with shorter expiration. The study revealed the fact that the future markets provide efficient hedging opportunities to the price risk in the spot markets.

**Hypothesizes**

- **H1:** The spot and future prices of the commodities are not cointegrated in the long run.
- **H2:** No causality relationship exists between spot and future prices of energy commodities.
- **H3:** No lead-lag relation exists in the spot and future markets of energy commodities.

**Research Methodology**

**Data**

The study has used daily closing spot and future prices of crude oil and natural gas for the time 2005 to 2020. The data was first converted into log series. ADF and PP Test unveiled that the data is stationary of the first order 1(1). The log value of spot and future prices of energy commodities has been used.

**Econometric tools**

Simple statistical techniques like mean, standard deviation, skewness, kurtosis and Jarque bera are employed with advanced econometric tools like Augmented Dickey Fuller Test (ADF), Phillip-Perron Test (PP), Johansen Co-
integration Test, VEC Granger Causality Test to make the study empirically more authentic. To examine long run co-integration between the spot and future prices of commodities Johansen Co-integration Test has been employed. We have used both maximum Eigen value test and trace value test. In order to investigate short run co-integration, direction of causality and lead lag relationship between the spot and future prices of commodities VEC Granger Causality Test/Block Exogeneity Wald Test was employed.

Table 1: Descriptive statistics of energy commodities

<table>
<thead>
<tr>
<th>Commodities</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>J-Bera</th>
<th>Prob.**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude oil</td>
<td>Future</td>
<td>3.572</td>
<td>0.124</td>
<td>0.110</td>
<td>2.281</td>
<td>90.476</td>
</tr>
<tr>
<td></td>
<td>Spot</td>
<td>3.566</td>
<td>0.128</td>
<td>0.104</td>
<td>2.245</td>
<td>98.195</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>Future</td>
<td>2.336</td>
<td>0.126</td>
<td>0.555</td>
<td>3.579</td>
<td>221.925</td>
</tr>
<tr>
<td></td>
<td>Spot</td>
<td>2.331</td>
<td>0.129</td>
<td>0.553</td>
<td>3.519</td>
<td>211.610</td>
</tr>
</tbody>
</table>

Note: Significant at: *0.01 and **0.05 levels

Fig 1: Histogram of the spot and future price

Histograms of the return series of spot and future prices of the crude oil and natural gas has been plotted to have a visual observation of the skewness, kurtosis and normality. From the FIGURE 1, it has been visualized that the data of return series of both the crude oil and natural gas is a flattened curve with long right tail and is not normal. Thus predicting some non-linear dynamics of the dataset.

Estimating stationarity
Augmented Dickey-Fuller (ADF) test
ADF test is considered one of the authentic tests for estimating stationary in time series data due to their ability to incorporate general ARIMA (p, q) with uncertain orders. The hypotheses of the test are:
Null hypothesis (H₀) is δ=0 i.e. the series has a unit root
Alternate hypothesis (H₁) δ<0, i.e. the series do not has a unit root

\[ \Delta Y_t = \mu + \delta Y_{t-1} + \sum_{i=1}^{n} i \alpha_i \Delta Y_{t-i} + \varepsilon_t \]

Where \( \mu \) is a constant or drift term, \( t \) denotes the coefficient on a time trend, \( \varepsilon_t \) denotes error term (white noise). \( n \) denotes the largest lag used and \( \Delta \) is first difference operator.

The value of test statistics (t-stat.) can be calculated as
\[ DF_T = \frac{\delta}{SE(\delta)} \]

Where \( DF_T \) represents the test statistics.

### Table 2: Augmented Dickey-Fuller Test of Energy Commodities

<table>
<thead>
<tr>
<th>Commodities</th>
<th>ADF Test</th>
<th>t-statistics</th>
<th>Critical Value</th>
<th>P-value**</th>
<th>Co-efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude oil Future</td>
<td>Level</td>
<td>-2.080</td>
<td>-2.862</td>
<td>0.252</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>First Difference</td>
<td>-59.922</td>
<td>-2.862</td>
<td>0.000</td>
<td>-0.966</td>
</tr>
<tr>
<td>Crude oil Spot</td>
<td>Level</td>
<td>-2.131</td>
<td>-2.862</td>
<td>0.232</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>First Difference</td>
<td>-67.903</td>
<td>-2.862</td>
<td>0.000</td>
<td>-1.091</td>
</tr>
<tr>
<td>Natural Gas Future</td>
<td>Level</td>
<td>-2.462</td>
<td>-2.862</td>
<td>0.125</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>First Difference</td>
<td>-57.673</td>
<td>-2.862</td>
<td>0.000</td>
<td>-0.990</td>
</tr>
<tr>
<td>Natural Gas Spot</td>
<td>Level</td>
<td>-2.584</td>
<td>-2.862</td>
<td>0.096</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>First Difference</td>
<td>-65.089</td>
<td>-2.862</td>
<td>0.000</td>
<td>-1.111</td>
</tr>
</tbody>
</table>

Note: Significant at: *0.01 and **0.05 levels

### Phillips Perron (PP) Test

It is another most widely used estimate in time series analysis to check the Stationarity in the data set. The test was given by Phillips and Perron in 1988 for checking unit root in a time series data and can be expressed as

\[ \Delta Y_t = \delta Y_{t-1} + \epsilon_t \]

### Johansen’s cointegration test

The test was given by Professor Soren Johansen who is well known for his contribution to the theory of cointegration. It is used to analyses the long run cointegration in time series data. The Johansen’s Cointegration Test has following hypothesis.

Null Hypothesis (R=0), there is no long run cointegration

Alternate Hypothesis (R=1), there is long run cointegration

\[ \Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{k-1} \Gamma \Delta Y_{1,i} + B X_t + \epsilon_t \]

Where \( \Delta Y_t = Y_t - Y_{t-1}, \epsilon_t \) denotes error term or white noise, \( T_t \) and \( \Pi \) represents the co-efficient matrix. The lag length \( k \) can be selected by following VAR lag length criteria.

The study has used vector auto regression (VAR) Schwarz Information Criterion (SIC) developed by statistician Gideon Schwarz (Profilidis and Botzoris, 2019) for selecting the lag length as

\[ SIC(p) = \ln(\sum) + \frac{K^2p\ln T}{T} \]

Where \( T \) is the sample size, \( K \) represents the number of variables and \( \sum(p) \) is the residual matrix without a degree of freedom correction from a VAR (p) model.

Johansen’s Cointegration Test was carried out on both the energy commodities viz crude oil and natural gas at lag 6 and 4 respectively. As presented in table 4, Johansen λ trace (trace statistics) and λ max (maximal eigen value) test has predicted that the p-value of the test is less than 5%. The null hypothesis of non-cointegration (R=0) has been rejected at 0.05 level of significance for both the commodities viz crude oil and natural gas. Hence, the null hypothesis \( H_0 \), is rejected. The study predicted a long run cointegration between the spot and future prices in energy commodities which is similar to the results found by various researchers viz Chhatwal and Puri (2014) [10], Bekiros and Diks (2008).
VEC granger causality test/ block exogeneity Wald test
The study employed Granger Causality Test not only to estimate causality relationship but also to investigate the direction of causality between the spot and future prices of commodities taken for the estimations. The log value of spot and future prices of energy commodities have been used and not the first difference series for running the test because the VEC granger causality test itself convert the series into first difference.

The test is based on following two basic assumptions
a) Cause happen before effect
b) Cause has some important information related to effect.

Hypothesis of the test are
Null Hypothesis (H0): non-existence of causal relationship
Alternate Hypothesis (H1): existence of causal relationship

When the two variables X and Y are integrated of order 1(1) following regression model can be used to test their causal relationship (Gupta, 2016).

\[ Y_t = \alpha + \sum_{i=1}^{r} a_t y_{t-i} + \sum_{j=1}^{q} b_j x_{t-j} + \epsilon_t \]

Where X Granger Causes Y, if the null hypothesis H0: \( b_1 = b_2 = \ldots = b_q = 0 \) is rejected and the alternate hypothesis H1: at least one \( b_j \neq 0, j=1, 2, 3 \ldots q \).

\[ X_t = \beta + \sum_{i=1}^{r} c_i X_{t-i} + \sum_{j=1}^{q} d_j Y_{t-j} + \eta_t \]

Where Y Granger Causes X, if H0: \( d_1 = d_2 = \ldots = d_q = 0 \) is rejected and the alternate hypothesis Ha: At least on \( d_j \neq 0, j = 1, 2, 3 \ldots s \). Following are the hypothesis of VEC Granger Causality Test:

Null Hypothesis (H0)
- H0: Fp does not granger cause Sp
- H0: Sp does not granger cause Fp

Alternate Hypothesis (H)
- H1: Fp does granger cause Sp
- H1: Sp does granger cause Fp

From the table 5, study concludes that the VEC granger causality test has estimated the existence of short run causal relationship between the spot and future prices between the spot and future prices of crude oil and natural gas. Hence, null hypothesis H2 is rejected. The estimates of the VEC Granger Causality Test have suggested that in short run there is a bi-directional causal relationship between the spot and future prices of crude oil and natural gas.

Lead-Lag relationship
According to table 5, the price discovery takes place simultaneously in the spot as well as future markets and both the markets are predicted to be equally efficient in incorporating the new information. Hence, null hypothesis H3, is accepted for crude oil and natural gas as no lead-lag relationship is predicted between the spot and future markets.

Conclusion
The study has investigated long run cointegration and lead-lag relationship between the spot and futures market of energy commodities. The empirical estimates of the study predict a long-run cointegration between both the markets which means that they cannot depart much from the equilibrium price in the long run. Cointegrated series can be integrated in the same order, or they should have a similar identifiable trend that can define a correlation between them. The VEC granger causality test suggests the presence of bi-directional causality between the spot and future markets of crude oil and natural gas. A feedback relationship also suggests the absence of any lead-lag relationship between the two markets. Thus, the study concludes that both the markets simultaneously react to the market innovations as well as shocks and are equally efficient in adjusting them in their respective equilibrium prices. The study reinforces the argument of previous studies augmenting a feedback relationship between the spot

Table 4: Johansen’s cointegration test of energy commodities

<table>
<thead>
<tr>
<th>Commodities</th>
<th>Lags</th>
<th>Ho: R</th>
<th>Trace Statistics</th>
<th>Max-Eigen statistics</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity</td>
<td></td>
<td></td>
<td>( \lambda ) trace</td>
<td>Prob.**</td>
<td>( \lambda ) max</td>
</tr>
<tr>
<td>Crude oil Future</td>
<td>6</td>
<td>0</td>
<td>62.014</td>
<td>0.000</td>
<td>59.184</td>
</tr>
<tr>
<td>Crude oil Spot</td>
<td>1</td>
<td>5.248</td>
<td>0.458</td>
<td>6.470</td>
<td>0.503</td>
</tr>
<tr>
<td>Natural Gas Future</td>
<td>4</td>
<td>1</td>
<td>145.117</td>
<td>0.000</td>
<td>168.225</td>
</tr>
<tr>
<td>Natural Gas Spot</td>
<td></td>
<td>1</td>
<td>9.804</td>
<td>0.254</td>
<td>7.827</td>
</tr>
</tbody>
</table>

Note: Significant at: *0.01 and **0.05 level

Table 5: VEC Granger Causality Test of Energy Commodities

<table>
<thead>
<tr>
<th>Commodity</th>
<th>Null Hypothesis (H0)</th>
<th>Chi-sq</th>
<th>P-Value**</th>
<th>Lags</th>
<th>Direction</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude oil</td>
<td>Fp does not granger cause Sp</td>
<td>4525.544</td>
<td>0.000</td>
<td>6</td>
<td>Bi-directional</td>
<td>F \longleftrightarrow S</td>
</tr>
<tr>
<td></td>
<td>Sp does not granger cause Fp</td>
<td>20.455</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural Gas</td>
<td>Fp does not granger cause Sp</td>
<td>1608.775</td>
<td>0.000</td>
<td>4</td>
<td>Bi-directional</td>
<td>F \longleftrightarrow S</td>
</tr>
<tr>
<td></td>
<td>Sp does not granger cause Fp</td>
<td>13.855</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Significant at: *0.01 and **0.05 levels, Sp-spot returns; Fp-future returns; \( \longleftrightarrow \) represents bi-directional causality.

~ 143 ~
and future markets of crude oil and natural gas.

References
1. Aggarwal N, Jain S, Thomas S. Do futures markets help in price discovery and risk management for commodities in India; c2014.