Trends and Determinants of Volatility: A Study of Soybean Futures Contracts

Saroj Joshi* and Ritu Sapra**

ABSTRACT

This study examines the trends and determinants of volatility in the context of the Indian futures market by taking soybean futures contracts traded on NCDEX. The sample consists of daily data on closing price, trading volume and open interest from Jan 3, 2005 to Dec 31, 2019. ARMA-GARCH model is being estimated for empirical analysis. The study finds that return distribution exhibits thick tails, time-varying volatility and volatility persistence. The GARCH effects are greater than the ARCH effects, which indicate that volatility is more sensitive to its own lagged values than recent news. The study finds a positive relationship between trading volume and volatility, whereas a negative relationship is observed between open interest and volatility. It was also observed that the inclusion of trading volume and open interest in the GARCH model reduces volatility persistence. The study concludes that trading volume and open interest are two important determinants of volatility.

Keywords: Soybean futures contracts; Trading volume; Open interest; Volatility; GARCH.

1.0 Introduction

Price volatility is fundamental to the theory and practice of asset allocation, asset pricing and risk management. Understanding and forecasting volatility has become an

*Corresponding author; Research Scholar, Department of Commerce, Delhi School of Economics, University of Delhi, Delhi, India (E-mail: saroj.joshi@srcc.du.ac.in)

**Professor, Department of Commerce, Delhi School of Economics, University of Delhi, Delhi, India (E-mail: sapra.ritu@gmail.com)
active area of research in finance. Several models are available to measure and forecast volatility, starting from standard deviation to complex models like ARCH (Autoregressive Conditional Heteroskedasticity) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) class of models. Investigating the behaviour of volatility is one of the important issues in financial markets. Some of the important determinants of volatility are trading volume and open interest. Thus, there is a need to study the effects of trading volume and open interest, which may help in explaining the behaviour of price volatility of futures contract. Volatility-volume relationships are important as they can form the basis of trading strategies, and this also has implications for market efficiency. Thus, to improve the understanding of the market microstructure, the relationship between volatility and volume has received substantial attention in financial research. The majority of empirical studies in financial literature support a positive relationship between trading volume and price volatility. Two of the leading hypotheses to explain this relationship are the Mixture of Distribution Hypothesis (MDH) and Sequential Information Arrival Hypothesis (SIAH).

Besides trading volume, open interest is an important measure of trading activity. It has generally been used as a proxy for hedging demand (Chen et al., 1999), hedgers’ opinions (Kamara, 1993), market depth (Bessembinder & Seguin, 1993) and difference in traders’ opinions (Bessembinder et al., 1996). In this study, we are using open interest as a proxy for market depth. Bessembinder & Seguin (1993) suggested that including open interest with trading volume in analyzing volatility may provide a better understanding of the price effect of trading activity generated by informed versus uninformed traders or speculators versus hedgers. Open interest, being a proxy for market depth, is expected to mitigate volatility.

Despite the importance of the relationship between trading activity (trading volume and open interest) and volatility, there is a paucity of research, especially in the emerging commodity futures market. Thus, in order to fill this research gap, this study seeks to empirically investigate the relationship between price volatility, trading volume and open interest in the Indian commodity futures market. The results of the present study will be useful for market participants, clearing houses and regulators. Studying the relationship between price volatility and measures of trading activity also improves the understanding of the market microstructure of the futures market.

The rest of the paper is organized as follows: Section 2 briefly reviews related studies. Section 3 deals with the conceptual framework. The objectives and hypotheses of the study are presented in section 4. In section 5 data and methodology are discussed. In section 6 empirical results and their analysis are reported. The last section deals with the summary and conclusions of the study.
2.0 Review of Literature

Lamoureux & Lastrapes (1990) first applied the GARCH methodology to examine the validity of MDH in equity markets. They took data on 20 actively traded stocks from the Chicago Board Options Exchange (CBOE). Daily trading volume was used as a measure of information flow based on Clark (1973). They found that when trading volume was included in the variance equation, the GARCH effect no longer appeared. The empirical results of this study strongly supported the MDH framework.

Bessembinder & Seguin (1993) explained the relationship between price volatility, trading volume, and open interest for eight futures market in the U.S. from May 1982 to March 1990. In addition to the trading volume, they also used open interest as a measure of trading activity, which was used as a proxy for market depth. They used a different modelling approach based on the time series method rather than examining single contracts. They decomposed volume and open interest into expected and unexpected components. Consistent with the earlier studies, they also found a strong positive relationship between price volatility and trading volume.

It was found that unexpected volume shocks influence price volatility more than expected volume shocks. In addition, they also found that expected open interest shocks have a significant negative effect on price volatility, which indicated that increased market depth (larger expected open interest) mitigates price volatility. Concerning the asymmetric relation between price volatility and volume, it was concluded that positive volume shocks have a larger effect on volatility than negative shocks. Similar results were found by Ragunathan & Peker, (1997) for the Sydney futures exchange and by Chan et al., (2004) for the Chinese Futures Exchange.

Pati, (2006) examined the volatility dynamics and investigated the Samuelson Maturity Hypothesis in the Indian futures market. ARMA-GARCH and ARMA-EGARCH models were estimated for empirical analysis. The study did not support the Samuelson hypothesis in the Indian futures market. With regard to the relationship between volume and volatility, the study indicated a clear acceptance of MDH. The study concluded that volume and open interest are important determinants of volatility; and the volume effect was stronger than the open interest effect. But GARCH effect was not completely vanished after the inclusion of volume and open interest in the variance equation.

Mahajan & Singh (2009) examined the relationship between return, volume and volatility in the Indian stock market. They used daily data of SENSEX from October 1996 to March 2006. This study has used GARCH (1, 1) and EGARCH (1, 1) models for analysis. The study found a positive and significant relationship between volume and volatility. The study documented a small decline in volatility persistence using the
GARCH (1, 1) model with trading volume. Further, it was found that ARCH and GARCH effects remained significant even after the inclusion of trading volume in the GARCH model, which highlights inefficiency in the market. Results of the EGARCH (1, 1) model indicated the presence of leverage effect.

Gupta & Rajib, (2012) examined the relationship between price volatility and time to expiry, as well as trading volume and open interest in the Indian commodity futures market using eight commodities traded on NCDEX. The methodology used consisted of GARCH, EGARCH and TGARCH models. Trading volume, open interest and time to maturity were three variables that were incorporated in the conditional variance equation. They found that the coefficient of time to maturity was not significant for majority of the contracts traded. Thus, it was concluded that Samuelson’s hypothesis does not hold true for most of the contracts considered. The study also found that volatility is moreover dependant on trading volume compared to open interest or time to maturity.

Joarder & Mukherjee (2018) empirically investigated the impact of time to maturity and volume on various oil and oilseed contracts on NCDEX using regression analysis. They found that volume is an important determinant of volatility than time to maturity of the contract.

3.0 Conceptual Framework

This section describes various concepts related to the study.

3.1 Stylized facts of financial returns

Fat-tail distributions, volatility persistence, and leverage effects are the well documented stylized facts about the financial returns (Mandelbrot, 1963; Fama, 1965; Black, 1976). Each of these stylized facts is explained below:

- **Fat tails:** A distribution is said to have fat tails, if it has high probability of extreme values than observed in a normal distribution. In the case of a fat tail distribution, the value of kurtosis is higher than 3.
- **Volatility persistence:** It means large changes (of either sign) are followed by large changes (of either sign), whereas small changes (of either sign) are followed by small changes (of either sign).
- **Leverage effect:** There is an asymmetric impact of good news (positive shocks) and bad news (negative shocks) on conditional volatility. Volatility increases more by a negative shock than by a positive shock of the same magnitude; this is known as the leverage effect.
3.2 Volatility

Volatility is defined as the variability of actual returns from the expected return. Higher volatility is associated with higher risk. The most common measure to calculate volatility is standard deviation, which is calculated as follows:

\[
\sigma = \sqrt{\frac{\sum_{t=1}^{n}(r_t - \bar{r})^2}{n}} \quad \text{...(1)}
\]

Where, \(r_t\) is return at time \(t\), \(\bar{r}\) is the average return and, \(n\) is the number of observations. Standard deviation is the most popular method of computing volatility. However, its main drawback is that it gives equal weight to all the residual terms. It also fails to capture various stylized facts like fat tails, volatility clustering and leverage effect.

3.3 Volatility persistence

Volatility persistence means large changes (of either sign) are followed by large changes (of either sign), whereas small changes (of either sign) are followed by small changes (of either sign). It is also well documented (Bollerslev, 1986) that fat tails of a distribution and volatility persistence are intimately linked, as distribution will have fat tails when there is volatility persistence. Persistence has a fundamental effect on the behaviour of a series. The presence of high volatility persistence in a series means that the lagged or past effects continue to affect the future values of the series. Volatility persistence in any market is indicative of market inefficiency and market irregularities. “The Efficient Markets Hypothesis (EMH) asserts that prices fully reflect available information in an efficient market”. This implies that investors can expect to earn merely risk-adjusted return from an investment as prices move instantaneously and randomly to any new information (Fama, 1970). Fama (1970) says that when markets are efficient, asset prices reflect an asset's true value. Thus, efficient markets are capable of arriving at the true value of an asset. Volatility persistence arises because markets are incomplete.

3.4 Relationship between trading volume and volatility

The majority of empirical studies in financial literature support a positive relationship between trading volume and volatility. Two of the leading hypotheses to explain this relationship are MDH and SIAH.

3.5 Mixture of distribution hypothesis (MDH)

The MDH states that there is a positive contemporaneous relationship between price volatility and trading volume (Clark, 1973; Epps & Epps, 1976; Tauchen & Pitts, 1983). MDH states that the volume-volatility relationship is on account of a joint dependence on a common mixing variable or event, i.e., the rate of information arrival in
the market. As the rate of information arrival is unobservable, trading volume is considered as a proxy for the rate of information arrival in the market. Any information flow in the market will affect both volume and volatility contemporaneously, and therefore, volume and volatility are expected to be positively related. The MDH suggests only a contemporaneous relationship between volume and volatility. Thus, under MDH past volatility data cannot be used to forecast volume or vice-versa since these variables change contemporaneously in response to the arrival of new information.

3.6 Sequential information arrival hypothesis (SIAH)

SIAH is another popular hypothesis that explains the relationship between trading volume and price volatility (Copeland, 1976; Morse, 1980; Jennings et al., 1981). According to this model, there is gradual dissemination of information in the market, which means a series of intermediate equilibriums exist before the final equilibrium is reached. Therefore, informed traders who get the information first take their positions and adjust their portfolios. This results in a shift in demand and supply and a series of intermediate equilibriums. Final equilibrium is restored in the market when the information is fully absorbed by all traders, informed as well as uninformed. All traders do not receive the information simultaneously, and as and when information is received, they revise their expectations accordingly. This sequential arrival of information to the market causes movements in both trading volume and price, with both increasing during periods of information shocks.

Thus, where MDH suggests only a contemporaneous relationship between trading volume and price volatility, the SIAH implies a dynamic relationship between these variables whereby lagged values of volatility may have the ability to predict the current trading volume and vice-versa (Mahajan & Singh, 2009).

3.7 Market depth

While the relationship between trading volume and volatility has been examined frequently, there are a few studies which have incorporated market depth in the analysis of trading volume and volatility even though market depth may be fundamentally related to trading activity and price movements (Bessembinder & Seguin (1992, 1993)).

3.8 Relationship between market depth and volatility

Generally, speculators in the market are ‘day traders’ who do not hold open positions overnight. Thus, open interest primarily reflects the uninformed trading or hedging activity in the market. Open interest, being a proxy of market depth or hedging activity, is expected to mitigate return volatility (Bessembinder & Seguin; 1993, 1996,
Chen et al., 1999). As open interest reflects futures traders’ current willingness to risk their capital in the futures position, it is a good proxy for market depth. A high level of open interest could help create market conditions that would reduce pressure from prices when trading provides new information. The relationship between volatility and open interest is expected to be negative. That is, deeper markets have relatively less price volatility.

4.0 Objectives and Hypotheses

This section describes the objectives and hypotheses of the study.

4.1 Objectives of the study

The main objective of the study is to analyze trends and to examine some of the determinants of volatility in the context of the Indian commodity futures market. In specific terms, this study is being done to assess the following aspects relating to the Indian commodity futures market:

- To examine the nature of volatility in the Indian commodity futures market.
- To measure the generalized volatility persistence (GARCH effect) in the Indian commodity futures market.
- To study the determinants of volatility in the Indian commodity futures market.
- To examine the nature of relationship between volatility and trading volume.
- To examine the nature of relationship between volatility and open interest.
- To examine whether volatility persistence is diminished or reduced when trading volume and open interest are incorporated as explanatory variables in the GARCH equation.

4.2 Hypotheses of the study

In order to achieve the objectives as mentioned above, the following null hypotheses have been framed for empirical testing:

To examine the first objective, we will test the following hypotheses:

$H_{01}$: The volatility is time-invariant (i.e. homogeneous) in soybean futures contracts.

$H_{02}$: There are no thick tails in the return distribution of soybean futures contracts.

$H_{03}$: There is no volatility clustering in the return distribution of soybean futures contracts.

To examine the second objective, we will test the following hypothesis:

$H_{04}$: GARCH effects are not present in the volatility of returns of soybean futures contracts.
To examine the third objective, we will test the following hypothesis:

H$_{05}$: Trading volume and open interest are not the determinants of volatility in the case of soybean futures contracts.

To examine the fourth objective, we will test the following hypothesis:

H$_{06}$: There is no relationship between volatility and trading volume in soybean futures contracts.

To examine the fifth objective, we will test the following hypothesis:

H$_{07}$: There is no relationship between volatility and open interest in soybean futures contracts.

To examine the sixth objective, we will test the following hypothesis:

H$_{08}$: There is no reduction in volatility persistence after incorporating trading volume and open interest in the GARCH equation in soybean futures contracts.

5.0 Data and Methodology

This section describes the data used for analysis and the various steps forming part of the methodology undertaken for examining the data.

5.1 Data and data source

The data consist of daily closing prices, trading volume and open interest for the soybean futures contracts traded at NCDEX from January 3, 2005 to December 31, 2019. All the data has been collected from Bloomberg.

The present study is restricted to soybean futures contracts traded on NCDEX due to constraints of time and resources. The reason for choosing soybean is that it is one of the largest oilseeds grown in India. India is also one of the top five soybean producers in the world. Moreover, the purpose was to choose a commodity that has been traded for an extended period on the national exchange with a good trading volume. The soybean futures contract is one such contract that has been actively traded in the National Commodity Derivatives Exchange (NCDEX) for a very long period of time.

5.2 Variables studied

- Daily futures returns: Daily continuous compound percentage return is obtained by taking natural logarithmic first difference of the closing prices multiplied by 100 i.e.
  \[ r_t = 100 \times \ln \left( \frac{P_t}{P_{t-1}} \right) \]
- Trading volume: It refers to the amount of futures contract traded in a given period of time during a trading day.
- Open interest: Open Interest is described as the total number of contracts that are not closed or yet to be squared off as on a particular day. It is reported as the number of contracts that remain outstanding at the end of a trading day. It is used as a proxy for market depth.

  Trading volume and open interest are used as measures of trading activity.

5.3 Methodology

5.3.1 Construction of return series

We have used daily closing prices to construct the return series. The daily continuous compound percentage returns \( r_t \) are defined as the first differences of the natural logarithmic price levels multiplied by 100.

\[
    r_t = 100 \times \ln \left( \frac{P_t}{P_{t-1}} \right)
\]

Where, \( P_t \) is the closing price on date \( t \) and \( P_{t-1} \) is the closing price on date \( t-1 \).

5.4 Preliminary analysis

Firstly, a number of descriptive statistics like mean, median, standard deviation, skewness and kurtosis are calculated to study various characteristics of the variables under study. We use kurtosis to examine the thick tails of a distribution. Jarque-Bera test statistic is used to test the normality of various time series (i.e. return, volume and open interest series) under consideration. The study also uses simple graphical analysis to visually examine the characteristics of the return series. Augmented Dickey-Fuller (ADF) test is used for testing the stationarity of various time series under consideration. Ljung-Box Statistics is used to check the presence of autocorrelation in the squared returns series. The presence of significant autocorrelation in squared returns series at various lags provides evidence for volatility persistence/clustering in the data.

5.5 Building volatility model

Building a volatility model for an asset return consists of the following four steps:

- Specify a mean equation: We have used the Box-Jenkins methodology (ARIMA modelling) to decide the mean equation. In this study, mean returns are modelled as an ARMA \((p, q)\) process.
- Testing for ARCH effects: To check the presence of ARCH effects in the data, the ARCH-LM test is used.
- Specify the volatility model if ARCH effects are statistically significant and jointly estimate the mean and variance equations.
- Check the fitted model carefully and refine it, if necessary.
5.6 ARCH/GARCH models

Financial research has shown much evidence that futures return exhibits skewness, leptokurtosis, time-varying volatility and volatility clustering, which have been well documented. Time series models that can take into account the changing variance and can model the conditional variance of error term are found to be more suitable for handling the financial data over the linear time series models. One of the prominent tools to capture the above-stylized facts is the ARCH and GARCH class of models (Brooks, 2008; Pati, 2006). This study captures the financial time series characteristics by employing Bollerslev’s (1986) GARCH model. In this study, we are using the GARCH (1, 1) model as it is found to be the most parsimonious model in literature and fits reasonably well with many economic time series.

5.7 ARCH model

ARCH (q) model proposed by Engle (1982) is given by:

\[ r_t = \mu_{t-1} + \varepsilon_t \]  
\[ \varepsilon_t | \psi_{t-1} \sim N(0, \sigma_t^2) \]
\[ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \]

Where, \( r_t \) is the return at time \( t \), \( \mu_{t-1} \) is the conditional mean, \( \psi_{t-1} \) is the information set that contains all relevant information at time \( t-1 \), \( \varepsilon_t \) is the error term of the mean equation, which is distributed with 0 mean and conditional variance which equals \( \sigma_t^2 \), \( \alpha_0 \) is the intercept of variance equation, \( q \) is the number of lag terms, \( \alpha_i \) is the coefficient of past squared residuals and \( \varepsilon_{t-i}^2 \) is the past squared residual.

To ensure a positive conditional variance, there are some restrictions on the conditional variance parameters. These are: \( \alpha_0 > 0 \) and \( \alpha_i \geq 0 \). To have a stationary ARCH model, the coefficient \( \alpha_i \) must be less than 1, otherwise the conditional variance (\( \sigma_t^2 \)) will continue to increase over time, eventually exploding. The problem with the ARCH (q) type model is estimating the number of lags (q) to be included in the variance equation.

5.8 GARCH model

Bollerslev (1986) extended the ARCH process to more general class of processes, GARCH (Generalized Autoregressive Conditional Heteroskedastic), which allow for a more flexible lag structure. In the ARCH (q) process, the conditional variance is specified as a linear function of lagged squared residuals only. In contrast, the GARCH (p, q) process allows lagged conditional variances to enter as well. For \( p=0 \), the GARCH process reduces to the ARCH (q) process (Bollerslev 1986).

The GARCH (p, q) model developed by Bollerslev (1986) is explained by:
\[ r_t = \mu_{t-1} + \epsilon_t \]  
\[ \epsilon_t \mid \psi_{t-1} \sim N(0, \sigma_t^2) \]  
\[ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \]  

The conditional variance equation specification in the above equation is a linear function of lagged squared residuals and lagged residual conditional variance.

To ensure positive conditional variance, there are some restrictions on the conditional variance parameters. These are: \( \alpha_0 > 0, \alpha_i \geq 0 \) and \( \beta_j \geq 0 \). To have a stationary GARCH model, the sum of \( \alpha_i \) and \( \beta_j \) must be less than 1 (\( \alpha_i + \beta_j < 1 \)).

5.9 Economic interpretation

The parameter \( \alpha_i \)'s capture the ARCH effect. It is interpreted as ‘news coefficient’ that measures the impact of recent news on volatility. The parameter \( \beta_j \)'s capture the GARCH effect. \( \beta_i \) is the ‘persistent’ coefficient that measures the impact of past volatility on current volatility. The sum of ARCH coefficient (\( \alpha_i \)) and GARCH coefficient (\( \beta_i \)) indicates the degree of persistence in volatility. In a GARCH framework, a high degree of persistence is said to exist if the sum of ARCH and GARCH parameters is close to unity.

5.10 Contemporaneous relationship between conditional volatility, trading volume and open interest

The relationship between conditional volatility, trading volume and open interest are studied by modifying the GARCH equation. For this, GARCH (1, 1) model is augmented by including trading volume and open interest as exogenous explanatory variables in the conditional variance equation. The augmented GARCH (1, 1) model is given by:

\[ r_t = \mu_{t-1} + \epsilon_t \]  
\[ \epsilon_t \mid \psi_{t-1} \sim N(0, \sigma_t^2) \]  
\[ \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma_1 VOL_t + \gamma_2 OINT_t \]  

Where, VOL is the trading volume at time \( t \), OINT is the open interest at time \( t \), \( \gamma_1 \) and \( \gamma_2 \) are the coefficients of trading volume and open interest, respectively. We expect a significant positive relationship between volatility and trading volume. On the other hand, the relationship between volatility and open interest is expected to be negative. Thus, the expected sign of \( \gamma_1 \) is positive and the expected sign of \( \gamma_2 \) is negative.

6.0 Results and Analysis

This section provides the details of various results obtained and their analysis.
The descriptive statistics of soybean futures returns, trading volume and open interest are presented in Table 1.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Return</th>
<th>Trading Volume</th>
<th>Open Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.029307</td>
<td>55544.93</td>
<td>100065.6</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>44790</td>
<td>87670</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>1.306091</td>
<td>42296.35</td>
<td>53573.33</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.7828</td>
<td>1.594097</td>
<td>1.136082</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>13.84722</td>
<td>6.36183</td>
<td>4.282732</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>20884.68</td>
<td>3732.484</td>
<td>1183.763</td>
</tr>
<tr>
<td>(P-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

Source: Authors' own computation

From Table 1, it can be seen that the mean daily return is 0.029307\% with a standard deviation of 1.306091\%. We can say that the return series is highly volatile as the value of standard deviation is very high compared to the value of mean return. The value of skewness is -0.7828, which indicates that distribution of the return series is negatively skewed or has a longer tail to the left. The value of kurtosis is 13.84722, which suggests that the return series exhibits fat tails and has excess peakedness at the mean than a normal distribution. The negative value of skewness and kurtosis of more than 3 indicates that the return series is not normally distributed but is negatively skewed and leptokurtic. The non-normality of the return series can also be confirmed by the Jarque-Bera test, where the null hypothesis is that the given series is normally distributed. Here the Jarque-Bera statistic is highly statistically significant and hence, we can reject the null hypothesis and say that the return series is not normally distributed. After examining the descriptive statistics, we visually examined the graph of soybean futures returns. Figure 1 exhibits the return series of soybean futures contracts for the sample period.

Figure 1 clearly shows that the mean returns are constant over a period of time and the variances fluctuate over time around some “normal” level. We can also see that large changes (of either sign) are followed by large changes (of either sign), whereas small changes (of either sign) are followed by small changes (of either sign), clearly indicating the volatility clustering in the return series. It can also be seen that the soybean futures contracts are experiencing excessive volatility from time to time, but volatility will eventually settle down to its long-run level. Thus, we can say that the soybean futures returns series displays mean reversion of returns, time-varying volatility and volatility clustering.
6.1 Testing for stationarity

Stationarity was tested using ADF test with intercept and ADF test with intercept and trend at level. The lag length was automatically selected using Akaike Information Criteria (AIC), such that the best fit model has minimum AIC. The null hypothesis of ADF test is that the given series contains unit root. If a series contains unit root it means that it is a non stationary series.

Decision Rule: If the computed absolute value of the tau statistic (|τ|) exceeds the Mackinnon critical tau value, we reject the null hypothesis that the unit root is present. In this case the time series is stationary.

Table 2: ADF Test (At Level)

<table>
<thead>
<tr>
<th>Series</th>
<th>With intercept</th>
<th>With both intercept and trend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test statistic (τ)</td>
<td>Test statistic (τ)</td>
</tr>
<tr>
<td>Return</td>
<td>-60.46082*</td>
<td>-60.45442*</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Trading Volume</td>
<td>-4.870251*</td>
<td>-4.983592*</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Open Interest</td>
<td>-4.118930*</td>
<td>-4.132216*</td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0056)</td>
</tr>
</tbody>
</table>

Source: Authors’ own computation

* Significant at 1% level. In the case of the ADF test with intercept, the critical value at 1% level of significance is -3.43. In the case of the ADF test with both intercept and trend, the critical value is -3.96 at 1% level of significance. Values in parentheses are p values.
From Table 2, we can see that all the series under consideration are stationary at 1% level of significance and can be used for further analysis.

6.2 Ljung-box test- autocorrelation of squared returns

In order to check the presence of volatility persistence in the return series, we conducted the Ljung-Box test on the squared return series. We computed the autocorrelation of squared return series at various lags. The results are presented in Table 3.

Null hypothesis: no significant autocorrelation is present in the squared return series.

**Table 3: Ljung-Box Test: Autocorrelation of Squared Returns**

<table>
<thead>
<tr>
<th>Lag</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.123</td>
<td>.123</td>
<td>62.778</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>.129</td>
<td>.116</td>
<td>132.644</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>.155</td>
<td>.130</td>
<td>232.695</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>.126</td>
<td>.086</td>
<td>298.661</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>.167</td>
<td>.122</td>
<td>415.506</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>.106</td>
<td>.046</td>
<td>462.806</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>.103</td>
<td>.041</td>
<td>507.107</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>.106</td>
<td>.040</td>
<td>553.800</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>.090</td>
<td>.027</td>
<td>587.584</td>
<td>0.000</td>
</tr>
<tr>
<td>10</td>
<td>.094</td>
<td>.030</td>
<td>624.717</td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>.108</td>
<td>.049</td>
<td>673.166</td>
<td>0.000</td>
</tr>
<tr>
<td>12</td>
<td>.094</td>
<td>.034</td>
<td>710.417</td>
<td>0.000</td>
</tr>
<tr>
<td>13</td>
<td>.063</td>
<td>-.001</td>
<td>727.211</td>
<td>0.000</td>
</tr>
<tr>
<td>14</td>
<td>.133</td>
<td>.077</td>
<td>801.012</td>
<td>0.000</td>
</tr>
<tr>
<td>15</td>
<td>.061</td>
<td>-.006</td>
<td>816.488</td>
<td>0.000</td>
</tr>
<tr>
<td>16</td>
<td>.080</td>
<td>.018</td>
<td>843.427</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Source: Authors’ own computation

It is clearly visible from Table 3 that Q statistic values at different lags are highly significant at 1% level of significance. Thus, we can reject the null hypothesis of no significant autocorrelation in the squared return series and accept the alternate hypothesis.
that significant autocorrelation is present in the squared return series. Thus, LB statistic provided evidence for the presence of volatility persistence in soybean futures.

6.3 Building volatility model

In this section, the results of various steps which have been followed to model the volatility in the return series are discussed.

6.3.1 Specifying the mean equation

To specify the mean equation, we followed the method suggested by Box & Jenkins (1976). The very first step of the BJ method involves a visual inspection of the ACF and PACF plots of the return series. Significant spikes in the ACF plot indicate the number of moving average (MA) terms that might be included in the model. Similarly, significant spikes in the PACF plot indicate the number of autoregressive (AR) terms that might be included in the model. Thus, to decide the mean equation, ACF and PACF plots of return series are made which are shown by Figure 2 and Figure 3, respectively.

Figure 2: ACF Plot of Soybean Futures Return
It can be seen from the plot of ACF (Figure 2) that the series has significant spike at lag 1. Thus, the first MA term might be included in the mean equation. Similarly, with the help of PACF plot (Figure 3) the first AR term might be included in the mean equation, as the PACF has significant spike at lag 1. Thus, on the basis of ACF and PACF the mean equation might have first AR term and first MA term. But this is not a foolproof method; it is only a trial and error method. Therefore, after having a rough idea about the mean equation, we estimated various plausible models and then on the basis of AIC & SC criteria, the best model was chosen. The best model is the one in which all the parameters are significant and which has minimum values for AIC & SC. On the basis of ACF and PACF graphs the mean equation can have one AR term and one MA term. But, when we tested this model (ARMA (1, 1)) with the least squares method, MA term was not coming out to be significant. Then, after testing the various models we came to the conclusion that AR (1) is the best model to describe the data generating process in case of soybean return series. The final decision was made by using the principle of parsimony while applying AIC and SC criteria. Thus, in this case the mean equation that needs to be estimated is as follows:
\[ r_t = \varphi_0 + \varphi_1 r_{t-1} + \varepsilon_t \quad \text{...(9)} \]

The results obtained after estimating the mean equation, with the help of least squares method are given in Table 4.

**Table 4: Results of AR (1) Model: Soybean Futures Contracts**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.033310</td>
<td>1.569463</td>
<td>0.1166</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.065575</td>
<td>4.243325</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Source: Authors’ own computation

We can see from Table 4 that the coefficient of AR (1) is significant at 1% level of significance.

6.3.2 Testing ARCH effect

After deciding the mean equation, we need to check the residuals of the equation for the presence of ARCH effects. Before we estimate GARCH model, it is customary and essential to check for the presence of ARCH effect in the data. This is done by employing Engel’s ARCH-LM test on the estimated squared residuals obtained from the estimating mean equation. The results are reported in Table 5.

Null Hypothesis: ARCH Effect is not present

**Table 5: ARCH-LM Test**

<table>
<thead>
<tr>
<th>F-statistic</th>
<th>Prob. F(1,4168)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.801481</td>
<td>0.0030</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ own computation

We can see from Table 5 that the p value of the test statistic is significant at 1% level of significance. Thus, we can reject the null hypothesis of no ARCH effect is present in the Soybean futures return and accept the alternate hypothesis that ARCH effect is present in the Soybean futures return. Presence of significant LM test statistic indicates the evidence of ARCH effects in the data. So we can further proceed to employ time-varying models of ARCH and GARCH to explain the return volatility over time.

6.3.3 Modelling time-varying volatility of soybean futures

GARCH model: Existence of significant ARCH effects in residuals of the fitted AR (1) model took us to the next step of modelling. We used GARCH (1, 1) model to
estimate the conditional volatility present in the soybean futures returns. GARCH model requires joint estimation of both mean and variance equation. Here, the mean equation is specified as AR (1). The following GARCH model is estimated to investigate volatility in soybean futures contracts.

\[ r_t = \varphi_0 + \varphi_1 r_{t-1} + \varepsilon_t \]  
\[ \varepsilon_t | \psi_{t-1} \sim N(0, \sigma^2_t) \]  
\[ \sigma^2_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} + \beta_1 \sigma^2_{t-1} \]

Where,
- \( r_t \) is the return at time \( t \).
- \( \varepsilon_t \) is the error term of the mean equation.

The parameter \( \alpha_1 \) captures the ARCH effect whereas \( \beta_1 \) captures the GARCH effect. The estimated results of the GARCH (1, 1) model are presented in Table 6.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>(( \alpha_0 )) = 0.017178</td>
<td>0.0000</td>
</tr>
<tr>
<td>RESID(-1)^2</td>
<td>(( \alpha_1 )) = 0.055098</td>
<td>0.0000</td>
</tr>
<tr>
<td>GARCH(-1)</td>
<td>(( \beta_1 )) = 0.935720</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>( \alpha_1 + \beta_1 = 0.990818 )</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ own computation

We can see from Table 6 that the coefficients of all three parameters (\( \alpha_0, \alpha_1 \) and \( \beta_1 \)) in the conditional variance equation are positive and highly significant. This satisfies the non-negativity constraint of the GARCH model. The sum of \( \alpha_1 \) and \( \beta_1 \) is also less than 1, which satisfies the stationarity condition of the GARCH model. The significance of the parameter \( \alpha_1 \) indicates that recent news has an impact on volatility.

Similarly, the significant \( \beta_1 \) parameter indicates that there is an impact of past volatility on current volatility. The sum of ARCH and GARCH coefficients (\( \alpha_1 + \beta_1 \)) is 0.990818. This shows a very high degree of volatility persistence in soybean futures returns. It is also observed that the estimated \( \beta_1 \) coefficient is larger than the estimated \( \alpha_1 \) coefficient, which indicates that volatility is more sensitive to its own lagged values than it is to recent news (or surprises or innovations) in the market. In other words, it can be said that the lagged effect of volatility is stronger than recent news in case of soybean futures contracts.
6.3.4 Testing the validity of the model

After testing the GARCH model, we need to make sure that the estimated model is valid for the given data. In other words, is the estimated model well specified to capture the ARCH effects present in the data? For this purpose, we tested the GARCH model for the presence of heteroscedasticity in the residuals by using ARCH-LM test. The results of the ARCH-LM test on the residuals of the GARCH model are presented in Table 7.

Null hypothesis: ARCH effect is not present

Table 7: ARCH-LM Test - On the GARCH Model

<table>
<thead>
<tr>
<th></th>
<th>F-statistic</th>
<th>Prob. F(1, 4168)</th>
<th>0.8264</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs*R-squared</td>
<td>0.048107</td>
<td>Prob. Chi-square(1)</td>
<td>0.8264</td>
</tr>
</tbody>
</table>

Source: Authors’ own computation

On the basis of the results presented in Table 7, we can observe that p values are not significant at 5% level of significance. So, we cannot reject the null hypothesis of no ARCH effects are present. Thus, we accept the null hypothesis that ARCH effects are not present in the data. Therefore, we can conclude that once we model the volatility using the GARCH model, there are no further ARCH effects left in the data. The estimated GARCH model adequately captures the volatility persistence in returns.

6.4 Contemporaneous relationship between volatility, trading volume and open interest

The relationship between volatility, trading volume and open interest is studied by modifying the GARCH equation. To examine the impact of trading volume and open interest on the volatility, GARCH (1, 1) model is augmented by including trading volume and open interest, as exogenous explanatory variables in the conditional variance equation. The augmented GARCH (1, 1) model is given by:

\[
\begin{align*}
    r_t &= \varphi_0 + \varphi_1 r_{t-1} + \varepsilon_t \\
    \varepsilon_t | \psi_{t-1} &\sim N(0, \sigma_t^2) \\
    \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma_1 VOLT + \gamma_2 OINT_t
\end{align*}
\]

...(12)  

The results of the augmented GARCH (1, 1) model are summarized in Table 8.

It can be seen from Table 8 that the values of coefficients \( \alpha_0 \), \( \alpha_1 \) and \( \beta_1 \) in the conditional variance equation are positive and highly significant. The coefficient of trading volume \( (\gamma_1) \) is coming out to be positive and highly significant. It shows that there is a significant positive contemporaneous relationship between the trading volume and volatility. Thus, we can say that as trading volume increases, we can expect that volatility...
will also increase and vice versa. These results are consistent with the findings of Joarder and Mukherjee (2018) - that trading volume has a positive relationship with volatility in the case of oil and oilseeds traded in NCDEX.

Table 8: Parameter Estimates of GARCH (1, 1) Model with Volume and Open Interest

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>$(\alpha_0) = 0.714694$</td>
<td>0.0000</td>
</tr>
<tr>
<td>RESID(-1)^2</td>
<td>$(\alpha_1) = 0.058535$</td>
<td>0.0000</td>
</tr>
<tr>
<td>GARCH(-1)</td>
<td>$(\beta_1) = 0.120001$</td>
<td>0.0000</td>
</tr>
<tr>
<td>VOL</td>
<td>$(\gamma_1) = 1.98E-05$</td>
<td>0.0000</td>
</tr>
<tr>
<td>OINT</td>
<td>$(\gamma_2) = -4.85E-06$</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>$\alpha_1 + \beta_1 = 0.178536$</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ own computation

The coefficient of open interest $(\gamma_2)$ is coming out to be negative and highly significant. It shows that there is a significant negative relationship between volatility and open interest. Thus, we can say that as open interest increases, volatility is expected to go down and we can expect a more stable market.

Now the sum of ARCH and GARCH coefficients $(\alpha_1 + \beta_1)$ is 0.178536. We can see that the after including the trading volume and open interest in the variance equation, the degree of volatility persistence has reduced from 0.990818 to 0.178536. From the above results it can be clearly seen that the inclusion of trading volume and open interest has led to a significant decline in the volatility persistence. It means that a large part of volatility persistence is absorbed by the trading volume and open interest. This clearly shows the importance of trading volume and open interest as key parameters influencing the volatility in case of soybean futures contracts. We also notice from Table 8 that even after inclusion of trading volume and open interest in the GARCH model, ARCH and GARCH effect remain significant. This shows the possibility that there can be other variables besides trading volume and open interest which contribute to the ARCH/GARCH effects in returns, in case of soybean futures.

Our results are consistent with earlier empirical findings of Gupta and Rajib (2012) and Pati (2006) that trading volume and open interest are significant determinants of volatility and their inclusion in the variance equation leads to a reduction in GARCH effects.
7.0 Results of the Hypotheses Tested

\( H_{01} \): The volatility is time invariant (i.e. homogeneous) in soybean futures contracts. We reject the null hypothesis and accept the alternate hypothesis that volatility in heterogeneous.

\( H_{02} \): There are no thick tails in the return distribution of soybean futures contracts. We reject the null hypothesis and accept the alternate hypothesis that thick tails are present in the soybean returns distribution.

\( H_{03} \): There is no volatility clustering in the return distribution of soybean futures contracts. We reject the null hypothesis and accept the alternate hypothesis that there is volatility clustering in the return distribution of soybean futures contracts.

\( H_{04} \): GARCH effects are not present in the volatility of returns of soybean futures contracts. We reject the null hypothesis and accept the alternate hypothesis that GARCH effects are present.

\( H_{05} \): Trading volume and open interest are not the determinants of volatility for soybean futures contracts. We reject the null hypothesis and accept the alternate hypothesis that trading volume and open interest are the determinants of volatility in case of soybean futures contracts.

\( H_{06} \): There is no relationship between volatility and trading volume in soybean futures contracts. We reject the null hypothesis and accept the alternate hypothesis that there is a relationship between volatility and trading volume.

\( H_{07} \): There is no relationship between volatility and open interest in soybean futures contracts. We reject the null hypothesis and accept the alternate hypothesis that there is a relationship between volatility and open interest.

\( H_{08} \): There is no reduction in volatility persistence after incorporating trading volume and open interest in the GARCH equation, in soybean futures contracts. We reject the null hypothesis and accept the alternate hypothesis that there is a reduction in volatility persistence after incorporating trading volume and open interest in the GARCH equation.
8.0 Summary and Conclusion

This study was conducted to examine the trends and determinants of volatility in the context of the Indian futures market by taking soybean futures contracts traded on NCDEX. We have used daily data on closing price, trading volume and open interest from January 3, 2005 to December 31, 2019. The methodology used in the study includes descriptive statistics, graphical analysis, ADF test, Ljung Box test, ARCH-LM test and GARCH model.

The study found that the return distribution of soybean futures contracts exhibits thick tails, time-varying volatility and volatility persistence. It was also found that ARCH and GARCH effects are present in soybean futures contracts. It was also observed that the estimated GARCH effects are greater than the estimated ARCH effects, which indicates that volatility is more sensitive to its own lagged values than it is to recent news (or surprises or innovations) in the market.

The study also found evidence that trading volume and open interest are important determinants of volatility in the case of soybean futures contracts. It was also observed that there is a positive relationship between volatility and trading volume, whereas there is a negative relationship between volatility and open interest in the case of soybean futures contracts. This negative relationship between volatility and open interest provides evidence that deeper markets tend to have lesser volatility. Thus, an increase in open interest leads to a more stable market. It was also observed that an inclusion of trading volume and open interest in the GARCH model leads to a reduction in volatility persistence from 0.990818 to 0.178536. It shows that the trading volume and open interest absorb a large part of volatility persistence.

Lastly, it was also seen that inclusion of trading volume and open interest in the GARCH model, resulted in a reduction in volatility persistence. But the GARCH effects did not completely vanish. This indicates that there are other factors also; which are responsible for the volatility persistence in the case of soybean futures contracts traded in the Indian commodity futures market.

9.0 Limitations and Scope for Further Research

The present study is confined to only Soybean futures contracts traded on NCDEX. Moreover, only two variables, trading volume and open interest are taken as determinants of price volatility. Further, this study has not explored the leverage effect in the commodity futures market.
In the future, this study can be extended by examining more contracts and taking more variables as determinants of price volatility. The present work can also be extended to study the leverage effects in the Indian commodity futures market.

References


