Price Movements in Futures and Spot Markets: Evidence from the S&P CNX Nifty Index

Kailash Chandra Pradhan
Joint Director
National Institute of Labour Economic Research & Development (NILERD)
Sector A-7, Narela, Delhi-110040, India
kailasheco@gmail.com

Abstract. This paper examines the price discovery and causality between spot and futures markets. It forecasts spot prices using the NIFTY futures markets. Vector Error Correction Model (VECM), Impulse Response Function analysis, and Variance Decomposition analysis are used to examine the price discovery process between spot and futures prices. This paper compares the forecast ability of futures prices on spot prices using Auto Regressive Integrated Moving Average (ARIMA) and VEC model. The results find that there exists a bi-directional causality between Nifty spot and futures markets and the spot markets disseminate new information stronger than futures prices. The forecast performance of VEC model is better than ARIMA model on post-sample periods. Because, VEC model incorporates the importance of taking into account the long-run relationship between the futures and the spot prices in forecasting future spot prices.

Keywords: price discovery; causality; forecasting; India.
1. INTRODUCTION

Price discovery process in the futures markets helps to achieve the market efficiency in the stock markets and also the futures markets minimising risk through hedging. In this paper, the price linkage between Indian stock index futures and its underlying index is examined. Price discovery functions depend on whether new information is reflected first in the futures markets or cash markets. In this process, both markets achieve a unique and common unobservable price, which is the efficient price. In perfect efficient markets, profitable arbitrage should not exist, as price adjusts simultaneously and fully to incoming information. And, new information disseminating into the market should be immediately reflected in cash and futures prices by triggering trading activity in one or all of the markets simultaneously. Therefore, nobody can make any profit in the long run.

The issue of price discovery process between cash and futures markets has been discussed and debated extensively in the literature. Studies such as Kawaller (1987), Harris (1989), Stoll and Whaley (1990), Chan (1992), Teppo and Vessa (1995), Arshanpalli and Doukesh (1997), Alphones (2000), Lafuente (2002), Tenmozhi (2002), Kavussanos and Nomikos (2003), So and Tse (2004), Bhatia (2007), Theissen, E. (2011) supported that the futures markets play an important role in the price discovery process by transferring new information faster than the cash market. Because futures markets are different from cash markets in terms of lower cost of transaction, capital required and other aspects. Chan and Kaloyi (1991), Tang, et al (1992), Turkington and Walse (1999), Zou and Pinfold (2001) and Raju and Karande (2003) showed that the bi-directional causality exists between both markets and price discovery takes place in both futures and cash market. Wahab and Lashgari (1995) and Mukherjee and Mishra (2006) showed that spot markets disseminate price information to futures markets. Wahab and Lashgari (1995) observed that though there is a lower transaction costs in the futures market but the spot market is more responsive to shocks in the futures market than to shocks in its own. Abhyankar (1995) found in his study that futures lead cash by an hour on average. More interestingly, he showed that lower transaction costs in the London cash market after the Big Bang have dampened the lead of futures, whereas short sale restrictions in the cash market have increased this lead. Mukherjee and Mishra (2006) observed the role of the futures market in the matter of price discovery tends to weaken and sometime disappear after the release of major firm-specific announcements.

Besides, this paper also examines whether the existence of a causal relationship between spot and futures prices can lead to more accurate predictions of future spot prices. Ghosh (1993), Wahab & Lashgari (1995), Tse (1995), Teppo et al (1995), Brooks, et al (2001) and Kavussanos and Nomikos (2003) observed the prices of financial futures contracts can be interpreted as forecasts of the spot rates, which will be applied at the final delivery date of that contracts. Futures prices play an essential role as a predictor of spot prices, because both the markets are interrelated. They also showed that the error correction model (ECM) performs better than other forecasting models like random walk, auto regressive integrated moving average (ARIMA) and vector auto regression (VAR) model. The present paper examined the prices discovery process between spot and futures markets and it also examined the forecasting performances of futures market to forecast the spot prices using the latest available data in National Stock Exchange (NSE), India.

After the brief introduction and identifying the objective of the paper in the section one, the rest of the chapter is structured as follows. Section two explains the methodology and data information. Section three offers empirical results and discussions of price discovery process between the two markets and it is also evaluates the forecast performance of the estimated model. Finally, section five presents the conclusion of the paper.

2. METHODOLOGY AND DATA

Johansen’s (1988) Vector Error Correction Model (VECM) was employed to examine the causal relationship between spot and futures prices. The following steps are followed to estimate Johansen’s Vector Error Correction Model (VECM).
Step 1: Augmented Dickey-Fuller (ADF), Dickey-Fuller Generalised Least Square (DF-GLS) and Phillips-Perron (PP) tests are conducted to examine the stationary of the data series.

Step 2: If the series are integrated in an identical order, then Johansen Multivariate Maximum likelihood cointegration test is used to investigate the long-run relationship between spot and futures prices and it is presented below.

\[
\Delta X_t = \sum_{i=1}^{p-1} \Gamma_i X_{t-i} + \Pi X_{t-1} + \varepsilon_t;
\]

\[
\varepsilon_t = \begin{pmatrix} \varepsilon_{S,t} \\ \varepsilon_{F,t} \end{pmatrix} \approx N(0, \Sigma). \tag{1}
\]

Where \(X_t = (S_t, F_t)\) is the vector of spot and futures prices, each being I (1) such that the first differenced series are I (0); \(\Delta\) denotes the first difference operator; \(\Gamma_i\) and \(\Pi\) are 2×2 coefficient matrices measuring the short- and long-run adjustment of the system to change in \(X_{it}\) and \(\varepsilon_t\) is 2×1 vector of white noise error terms.

Step 3: The test results are quite sensitive to the lag length. Hence, the lag length \(P\) is selected on the basis of multivariate generalizations of Akaike’s information criteria (AIC) and Schwarz’s criteria (SC).

Step 4: The likelihood ratio tests are employed to identify the co-integration between the two series. The first statistic \(\lambda_{\text{trace}}\) tests whether the number of cointegrating vectors is zero or one.

In general, if \(r\) cointegrating vector is correct. The following test statistics can be constructed as:

\[
\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^{n} \ln \left(1 - \hat{\lambda}_i \right). \tag{2}
\]

Where, \(n\) is the number of separate series to be examined, \(T\) is the number of usable observations and \((\hat{\lambda}_i)\) are the estimated eigenvalues (also called characteristic roots) obtained from the \((i+1) \times (i+1)\) ‘cointegrating matrix.’

The test statistic \(\lambda_{\text{trace}}\) tests whether the number of distinct cointegrating vectors is less than or equal to \(r\). Johansen and Jueselins (1990) provide the critical values of these statistics. The rank of \(\Pi\) may be tested using the \(\lambda_{\text{trace}}\). If rank \((\Pi) = 1\), then there is single cointegrating vector and \(\Pi\) can be factored as \(\Pi = \alpha \beta\), where \(\alpha\) and \(\beta\) are 2×1 vectors. Using this factorisation \(\beta\) represents the vector of cointegrating parameters and \(\alpha\) is the vector of error correction coefficients measuring the speed of convergence to the long-run steady state.

Step 5: If spot and futures prices are cointegrated, then causality must exist at least in one direction (Granger, 1986). To test the causality, the following vector error correction model (VECM) is estimated by using ordinary least square (OLS) in each equation.

\[
\Delta S_t = a_{S,0} + a_{S,i} \Delta F_{t-i} + \alpha_{S} Z_{t-1} + \varepsilon_{S,t} \tag{3}
\]

\[
\Delta F_t = a_{F,0} + a_{F,i} \Delta S_{t-i} + \alpha_{F} Z_{t-1} + \varepsilon_{F,t} \tag{4}
\]

where \(a_{S,0}, a_{F,0}\) are intercept terms; \(a_{S,i}, a_{F,i}\) are the short-run coefficients and \(Z_{t-1} = \beta^\prime X_{t-1}\) is the error correction term from equation (1).

In terms of the vector error correction model (VECM) of equation (3) & (4), F, Granger Causes S, if some of the \(b_{S,i}\) coefficients, \(i = 1, 2, \ldots, p-1\) are not zero and \(\alpha\), the error correction coefficient in the equation for spot prices, is significant at conventional levels. Similarly, S, Granger causes F, if some of the \(a_{F,i}\) coefficients, \(i = 1, 2, \ldots, p-1\) are not zero and \(\alpha\) is significant at the conventional levels. These hypotheses can be tested by using either t-tests or F-tests on the joint significance of the lagged estimated coefficients. If both S, and F, Granger cause each other, then there is a feedback relationship between the two markets. Therefore, the error correction coefficients, \(\alpha\) and \(\alpha\) serve two purposes. They are (i) to identify the direction of causality between spot and futures prices and (ii) to measure the speed with which deviations from the long-run relationship are corrected by changes in the spot and futures prices.

The vector error correction model (VECM) equation (3) & (4) provides a framework for
valid inference in the presence of I (1) variable. Moreover, the Johansen (1988) procedure provides more efficient estimates of the cointegrating relationship than the Engel and Granger (1987) estimator (Gonzalo, 1994). Also, Johansen (1988) tests are shown to be fairly robust to presence of non-normality (Cheung and Lai, 1993) and heteroscedasticity disturbances (Lee and Tse, 1996).

THE FORECASTING MODELS

The prices of financial futures contracts can be interpreted as forecasts of the spot rates, which will be applied at the final delivery date of that contract. This study compares the forecasting ability of futures prices on spot prices with two major forecasting techniques namely auto regressive integrated moving average (ARIMA) and vector error correction (VEC) model.

Cointegration and vector error correction model

Johansen’s cointegration and vector error correction model are explained in the first sub-section of this section. The forecasting of the vector error correction model (VECM) for the spot and futures prices can be expressed as

\[ \Delta S_t = a_{S,0} + \sum_{i=1}^{p-1} a_{S,i} \Delta S_{t-i} + \sum_{i=1}^{q-1} b_{S,i} \Delta F_{t-i} + \alpha_S Z_{t-i} + e_{S,t}. \]  

(5)

An ARIMA model

In order to form a benchmark for comparison to the vector error correction (VECM) models previously, an auto regressive integrated moving average (ARIMA) model is estimated (with \( S_t \) as the dependent variable since prediction of the spot series is the modeling motivation). An ARIMA \((p, d, q)\) model is a univariate time series modeling technique, where \( p \) denotes the number of autoregressive terms, \( d \) the number of integrated order and \( q \) the number of moving average terms which is based on Box-Jenkins methodology (Box-Jenkins, 1970). The ARIMA model is expressed as

\[ S_t = a_{S,0} + \sum_{i=1}^{p-1} a_{S,i} S_{t-i} + \sum_{i=1}^{q-1} b_{S,i} F_{t-i} + \epsilon_{S,t}. \]  

(6)

Again the Akaike’s information criteria (AIC) and Schwarz’s criteria (SC) is utilized for selecting lags of the model.

Then, the constructed models or techniques are examined on the basis of whether each significantly “outperforms” the forecasting ability of the futures price. Performance of the model is measured by the validity of its estimate on the basis of its forecasting power tests such as: root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and Theil’s inequality coefficient (U-statistic).

All the required data information for the study has been collected from the National Stock Exchange (NSE), India website. The main data set for the study consists of the daily closing values of the S&P CNX Nifty index futures and spot Nifty index, which are considered from June 12, 2000 to January 28, 2016 for near month futures contracts and it consists 3892 observations. In-sample analysis is carried out for the period June 12, 2000 to November 30, 2015 with 3851 observations and remaining observations (41) from December 1, 2015 to January 28, 2016 are considered to evaluate the out-of-sample forecasting performance of the model. The study has taken \( S_t \) and \( F_t \) as natural logged spot and futures prices respectively. The near month futures have been analysed as they are mostly heavily traded.

3. RESULTS AND DISCUSSIONS

The stationary of the spot and futures prices series are tested using the augmented Dickey Fuller (ADF) tests, Dickey-Fuller test statistic using a generalized least squares (DF-GLS) and Phillips Perron (PP) tests. The optimal lag numbers of each series are tested by using the Akaike’s Information Criteria (AIC) and Schwarz Criteria (SC). According to Akaike’s Information Criteria (AIC) and Schwarz Criteria (SC), four lags for the DF and PP tests and maximum 8 lags for the DF-GLS test have been selected for both Nifty spot and futures
prices series. In the table 1, the results reject the presence of a unit root in both series because the test statistic is significant at 1% level. The findings concluded that both spot and futures prices are non-stationary at levels and stationary at first difference. In the table 2, Johansen’s cointegration test is performed for Nifty Index spot and Nifty futures prices. The test finds that one cointegration relationship exists between spot and futures markets and there is long relation between them. Thus Johansen tests for cointegration justify the use of a vector error correction model (VECM) for showing short run dynamics.

To assess the optimal lag length, Stata varsoc command is used with a maximum lag length of four. In the table 3, most of criteria support a lag of length four. Therefore four

<table>
<thead>
<tr>
<th>Constraint</th>
<th>ADF</th>
<th>DF-GLS</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Levels</td>
<td>Difference</td>
<td>Levels</td>
</tr>
<tr>
<td>ln(spot price)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.692</td>
<td>-30.822*</td>
<td>0.827</td>
</tr>
<tr>
<td>Intercept and trend</td>
<td>-2.097</td>
<td>-30.818*</td>
<td>-1.684</td>
</tr>
</tbody>
</table>

| ln(futures price)   |          |           |          |           |        |            |
| Intercept           | -0.704   | -31.047*  | 0.783    | -9.148*  | -0.698 | -60.558*   |
| Intercept and trend | -2.148   | -31.043*  | -1.703   | -14.403* | -2.148 | -60.550*   |

Note: * denotes 1% level of significance.

<table>
<thead>
<tr>
<th>Maximum rank</th>
<th>Parms</th>
<th>LL</th>
<th>Eigenvalue</th>
<th>Trace statistic</th>
<th>5% critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14</td>
<td>28249.63</td>
<td>–</td>
<td>194.8009</td>
<td>15.41</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>28346.79</td>
<td>0.04875</td>
<td>0.4827*</td>
<td>3.76</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>28347.03</td>
<td>0.00012</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * denotes 1% level of significance.

<table>
<thead>
<tr>
<th>lag</th>
<th>LL</th>
<th>LR</th>
<th>FPE</th>
<th>AIC</th>
<th>HQIC</th>
<th>SBIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>11583.20</td>
<td>8.90E-06</td>
<td>-5.96</td>
<td>-5.96</td>
<td>-5.95</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>28146.60</td>
<td>33127</td>
<td>1.80E-09</td>
<td>-14.48</td>
<td>-14.47</td>
<td>-14.47</td>
</tr>
<tr>
<td>2</td>
<td>28311.00</td>
<td>328.90</td>
<td>1.60E-09</td>
<td>-14.56</td>
<td>-14.55</td>
<td>-14.54</td>
</tr>
<tr>
<td>3</td>
<td>28339.10</td>
<td>56.16</td>
<td>1.60E-09</td>
<td>-14.57</td>
<td>-14.56*</td>
<td>-14.54</td>
</tr>
<tr>
<td>4</td>
<td>28347</td>
<td>15.89*</td>
<td>1.60E-09*</td>
<td>-14.57*</td>
<td>-14.56</td>
<td>-14.55*</td>
</tr>
</tbody>
</table>

Table 1. Unit Root Tests

Table 2. Johansen tests for cointegration

Table 3. Selection–order criteria
lags have taken to test the vector error correction model (VECM).

In the table 4, the VECM estimation results have shown that both spot and futures price series are adjusting to the previous period’s deviation from long-run equilibrium. But the futures price series have a greater speed of adjustment to the previous period’s deviation from than the spot price series. Because it is noticed that $\alpha_f$ is 0.3279 which is greater than $\alpha_s$ (0.1645). This finding is suggested that the delivery date of each contract the futures price has to adjust itself to the prevailing spot price. The results find that there is causality from spot to futures at the first lag periods i.e., spot market leads the futures market and the significance level is 5 percent. And, the futures market leads the spot markets at

**Table 4. Tests for Vector Error Correction Model**

<table>
<thead>
<tr>
<th>Variables</th>
<th>$\Delta S_t$</th>
<th>Std. Err.</th>
<th>$\Delta F_t$</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{i=s,f}$</td>
<td>0.0002*</td>
<td>0.0003</td>
<td>-0.0001</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\Delta S_{t-1}$</td>
<td>-0.0020</td>
<td>0.1019</td>
<td>0.2325**</td>
<td>0.1070</td>
</tr>
<tr>
<td>$\Delta S_{t-2}$</td>
<td>-0.1043</td>
<td>0.1010</td>
<td>0.0196</td>
<td>0.1060</td>
</tr>
<tr>
<td>$\Delta S_{t-3}$</td>
<td>-0.1410</td>
<td>0.0921</td>
<td>-0.0888</td>
<td>0.0966</td>
</tr>
<tr>
<td>$\Delta F_{t-1}$</td>
<td>0.0833</td>
<td>0.0979</td>
<td>-0.1737***</td>
<td>0.1027</td>
</tr>
<tr>
<td>$\Delta F_{t-2}$</td>
<td>0.0557</td>
<td>0.0974</td>
<td>-0.0589</td>
<td>0.1023</td>
</tr>
<tr>
<td>$\Delta F_{t-3}$</td>
<td>0.1485***</td>
<td>0.0888</td>
<td>0.1012</td>
<td>0.0932</td>
</tr>
<tr>
<td>$Z_{t-1}$</td>
<td>0.1645*</td>
<td>0.0642</td>
<td>0.3279*</td>
<td>0.0674</td>
</tr>
</tbody>
</table>

Note: *, ** and *** denotes 1%, 5% and 10% level of significance respectively.

**Figure 1. Graphs for the stability condition**
the third lag periods and the significance level is 10 percent. Here, the results also show that there is bi-directional causality between spot and futures markets. Thus, the price discovery process takes place in both spot and futures market. The findings from the different results concluded that the price lead of spot market is stronger than futures market.

To assess the validity of VECM, stability of the model is tested. The varstable command examines the dynamic stability of the system. In the figure 1, none of the eigen values is even close to one. The test concludes that the system is stable.

Further, impulse response functions and variance decomposition of the VECM are used to get a more detailed insight on the causal relationship between spot and futures prices. The diagonal panels in Figure 2 show the effects of shocks to each change of market price on future values of its own change. In case of futures prices, the shock is reflecting increased initial periods and then it is declined. Spot price is increasing of its own shocks. The off-diagonal panels (bottom-left and top-right) show the effects of a growth shock in one market price on the path of growth in the other. In the bottom-left panel, it shows that a one-standard-deviation shock in change of spot prices raises the change of the futures prices and the impact of spot prices on futures prices is very high. In the top-right panel, it

<table>
<thead>
<tr>
<th>Forecast performance</th>
<th>VECM</th>
<th>ARIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.00004</td>
<td>0.00115</td>
</tr>
<tr>
<td>MAE</td>
<td>0.00194</td>
<td>0.03818</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.00022</td>
<td>0.00428</td>
</tr>
<tr>
<td>U-stat</td>
<td>0.0005</td>
<td>0.0027</td>
</tr>
</tbody>
</table>

Figure 2. Impulse response functions
shows that the estimated effects of a shock to futures prices on spot prices. The impact of futures prices on spot prices has increased in the initial periods and after that the impact of futures prices is declined. Figure 3 show the variance decompositions between spot and futures prices. It shows that spot prices explain most (98%) of the variation in growth of spot prices and its impact on futures prices is greater. The futures prices explain only 17 percent of the variations in growth of its own and its impact on spot prices is very low. The impulse response function and variance decomposition analysis shows that the impact of spot markets is higher than futures market and the shocks of spot prices seem to have large effect relative to futures prices. The findings conclude that most of variations of prices in spot and futures prices take place due to the spot prices.

This paper considers two models of predicting the spot price series such as: ARIMA model and VECM to compare the forecasting performance. These two model specifications are estimated recursively during the out-of-sample period and generate forecasts of the spot prices up to one steps (trading days) ahead. Then, these forecast values are compared to the actual prices on the basis of standard statistical criteria of root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and Theil’s inequality coefficient (U-statistic). In the table 5, the results have shown that the reduction in the RMSE, MAE, MAPE and U-statistic is achieved by the VECM over the ARIMA model in forecasting spot price. The results conclude that the forecast performance of VECM is better than ARIMA model because VECM takes the lead-lag relationship between the spot and futures markets rather than simply using information contained in the univariate spot series alone.

4. CONCLUSION
This paper examined the price discovery, causality and forecasting in the S&P CNX Nifty futures prices. The findings from unit root tests
have shown that Nifty spot Index and Nifty futures Index are not stationary at their levels. But they are stationary at their first difference. The cointegration test results have shown that there is a long run relationship between spot and futures prices. Therefore, a vector error correction model (VECM) is used to examine the short-run dynamics and price movements in the two markets. The Johansen’s vector error correction model (VECM) results found that there is a bi-directional causality between spot and futures markets and the lead of the spot market on the futures market is more pronounced. Spot prices tend to discover new information more rapidly than futures prices. The impulse response analysis and variance decomposition analysis has shown that spot prices tend to discover new information more rapidly than futures prices. Finally the results find that the information and cointegrating relationship between spot and futures prices can be used to generate more accurate forecast of the prices.

This paper has shown that there is feedback relationship, but the spot lead was stronger than the futures index lead. The leading role of futures market weakens around the firm-specific announcements (Mukherjee and Mishra, 2006). In the futures market, the payoffs and risk that buyer and seller face are considerably more difficult than those seen on the equity market. Therefore, spot market lead is stronger than futures market. Also, the findings suggest that vector error correction model (VECM) performs well on a post-sample basis against the univariate auto regressive integrated moving average (ARIMA) model. The results show clearly that it is important to take into account the long-run relationship between the futures and the spot prices in forecasting future spot prices. The market participants can be benefited by taking the VECM to forecast the spot futures price index and it will help to design more efficient investment and speculative trading strategies.

REFERENCES


