Intraday price discovery in Indian stock index futures market: new evidence from neural network approach

Saurabh Kumar
Information Technology and Systems Area, FPM-22, Indian Institute of Management (IIM), Lucknow, Prabandh Nagar, Lucknow, PIN 226013, Uttar Pradesh, India
Email: saurabh@iiml.ac.in

Sarveshwar Kumar Inani*
Finance and Accounting Area, FPM-5, FPM Hostel, Indian Institute of Management (IIM), Lucknow, PIN 226013, Uttar Pradesh, India
Email: sarveshwarinani@gmail.com
Email: fpm13006@iiml.ac.in
*Corresponding author

Abstract: Using neural network approach, this study revisits the price discovery relationship between spot and futures prices of S&P CNX Nifty Index of India. This study uses minute-by-minute data of 167 trading days (from January 2015 to August 2015). Empirical results reveal that there exists a bi-directional lead-lag relationship between spot and futures markets, but the causality from the spot market to the futures market is more dominant as compared to that of running from the futures market to the spot market. Root mean squared error of the results indicates that the incorporation of spot returns in modelling futures returns improves the results by 62.37%. Whereas, the inclusion of futures returns in modelling spot returns improves the model by only 44.60%. Price discovery dynamics has significant implications for market participants, regulators, and exchanges. These findings could be helpful for portfolio management, market structure design, and the implementation of arbitrage and hedging strategies.

Keywords: price discovery; futures market; intraday data; India; market microstructure, neural network.


Biographical notes: Saurabh Kumar is currently an FPM student of IT and systems area in IIM Lucknow. Prior to joining IIM, he has served at Tata Consultancy Services (TCS), Bangalore. He received his BTech in Computer Science and Engineering from SATRA University, Tamil Nadu. His research interest includes data mining, soft computing techniques, privacy and security issues in big data and cloud computing. He has published a research paper in The Journal of Prediction Markets.
1 Introduction

Markets have two important functions—liquidity and price discovery; and these functions are important for asset pricing [O’Hara, (2003), p.1335]. Price discovery is the process through which financial markets reach the efficient equilibrium price. Price discovery is the efficient and timely incorporation of the information implicit in investor trading into market [Lehmann, (2002), p. 259]. As per efficient market hypothesis, spot and futures prices of any underlying security should move simultaneously without any lead or lag. However, due to some market frictions such as liquidity, lot size, or transaction costs (bid-ask spread and impact cost) lead-lag relation between spot and futures markets arises. The analysis of such lead-lag relationship is termed as price discovery in the market microstructure literature.

Ideally, spot and futures prices should be cointegrated with (1, –1) cointegrating relationship, if not then costless arbitrage will bring them back to the long-run equilibrium of identical prices. Price discovery process shows which market (spot or futures) moves first to incorporate information as compared to another market; and also points out the level of efficiency in the markets. Price discovery has significant implications for market participants, regulators, and exchanges. First of all, price discovery concept is related to market efficiency. Second, an investigation of price discovery can help in designing market structure. Finally, hedging and arbitrage strategies must consider the lead-lag relation between spot and futures markets.

Indian derivatives market for index futures has witnessed phenomenal success in the last decade. India is the second fastest emerging economies in the world following China. National Stock Exchange of India (NSE) is the leading and prominent derivatives market in India. Besides, NSE is ranked fourth largest derivatives market in the world, following CME group; Intercontinental exchange; and Eurex exchanges respectively, as per annual volume survey of Futures Industry Association (FIA) in the year 2014. FIA survey also states that a total of 1.88 billion contracts were traded in NSE during the year 2014. Indian equity market and its market microstructure have certain distinguished characteristics, which make it an ideal candidate for sole analysis. Hence, it stimulates the interest for a separate study of price discovery process of S&P CNX Nifty index spot and futures contracts.

The S&P CNX Nifty-fifty (hereafter Nifty) is a well-diversified index of fifty stocks, which account for 13 sectors of the Indian economy. The primary purpose of trading in Nifty is to offer a risk-management tool for investors, through hedging and
diversification, against any market movements. Nifty also works as a benchmark for portfolios. The Nifty index represents about 66.17% of the free float market capitalisation of the stocks listed on NSE as on March 31, 2015. The total traded value of Nifty index constituents for the last six months ending March 2015 is approximately 46.22% of the traded value of all stocks on the NSE. Impact cost of the Nifty 50 for a portfolio size of Rs.50 lakhs is 0.06% for the month March 2015. Nifty is professionally maintained and is considered an ideal medium for derivatives trading. NSE introduced futures for Nifty on June 12, 2000. Trading in NSE occurs on all days of the week except Saturdays, Sundays and declared holidays. Both spot and futures segments have identical trading hours in same exchange (NSE) with similar trading practices through the screen-based trading system. Futures contracts are cash settled on last Thursday of every month. If last Thursday is a holiday, then contracts are settled on the day before that holiday. Short sales are not allowed in the spot market. However, it is very common in the futures market.

Different methodologies have been employed in literature to investigate price discovery process in index futures contracts. These methods could be divided into two broad categories – conventional techniques and common factor methods. Conventional techniques consist of correlation technique (Finnerty and Park, 1987), three-stage least square regression technique (Kawaller et al., 1987), cross-correlation technique (Herbst et al., 1987), cointegration and error correction models (Wahab and Lashgari, 1993), and GARCH models (Abhyankar, 1995). Whereas, common factor methods are built upon error correction models. In the recent literature, the majority of the studies have used common factor methods. These methods are component share method (Gonzalo and Granger, 1995), information share method (Hasbrouck, 1995), and modified information share method (Lien and Shrestha, 2009). Some prominent studies in the recent literature have used these common factor methods (Chen and Gau, 2010; Fricke and Menkhoff, 2011; Frijns et al., 2015a; 2015b; Mizrach and Neely, 2008; Narayan et al., 2014). In addition to these methods, Gong et al. (2016) have employed thermal optical path method to examine the price discovery. However, no study has ever used the neural network approach to investigate price discovery process to the best of our knowledge. Therefore, it motivates us to reexamine the lead-lag relationship or price discovery process of S&P CNX Nifty index spot and futures with the help of the neural network methodology.

Neural network can be defined as a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs (Caudill, 1987). To put it simply, Neural networks are trainable algorithms which are analogous to the functioning of the human brain. Neural networks have self-learning ability from the available data, and they are well-known for forecasting. Traders can make better trading strategies with the aid of neural networks as compared to the conventional technical analysis. Neural networks are used for detecting nonlinear relationships, patterns, and associations. The success of neural networks depends on inputs used in analysis and the configuration of neural networks. For instance, to predict the day ahead index futures price with better accuracy, what inputs are required in the model: only historical futures prices or a combination of futures prices with spot prices? Neural networks ensure that the nonlinearity associated with the data is handled effectively. Neural networks models are generally used for the estimation or approximation of unknown quantities which depend on considerably a large number of inputs. Neural networks are used widely for forecasting purpose mainly because of the fact that the neural networks are universal approximators and can approximate a large
class of functions with a high degree of accuracy over other nonlinear models (Chen et al., 2003; Zhang and Qi, 2005). Neural networks are generally used in technical analysis and high-frequency trading, but we have not come across any study, to the best of our knowledge, which investigates the price discovery process using the neural network approach.

The aim of this study is to re-investigate the price discovery process between S&P CNX Nifty index spot and futures with the help of one of the machine learning techniques – neural network. Neural network ensures that the nonlinearity associated with the data is handled effectively. This paper contributes to the price discovery literature in several ways. First and foremost, this study is the first attempt to apply neural network approach to examine the price discovery process. The application of neural network model in analysing the lead-lag relation between spot and futures prices extends the price discovery literature by incorporating nonlinear relationships, which has not been incorporated in the extant literature. Secondly, this study employs intraday one-minute high-frequency data, which is very rare in the context of an emerging market like India. Since price discovery is associated with market efficiency, it has significant implications for market participants, regulators, and exchanges. Hence, the research in price discovery needs constant attention from researchers and practitioners across the globe. This study provides new evidence regarding price discovery in the Indian equity market. These results may help in drawing fresh insights regarding the efficiency of futures markets in emerging economies. Moreover, the results of this study indicate that spot returns lead futures returns. Therefore, the spot market in India is more efficient than the futures market.

The remainder of the paper is organised as follows: Section 2 describes the theoretical background; Section 3 discusses existing literature; Section 4 explains data description; Section 5 presents the neural network methodology used in the study; Section 6 deals with the findings and discussion of the results; and finally, Section 7 concludes the study with limitations and future scope.

2 Theoretical background

There are different theoretical arguments for the lead-lag relationship between spot and futures markets. The trading cost hypothesis argues that the market with lower transaction cost will react faster to new information. Since futures market has lower transaction cost vis-à-vis spot market, any information is expected to be revealed first in the futures market. This explanation is in line with the leverage hypothesis as well. Leverage hypothesis suggests that lower margin requirements in futures market enable it to dominate the price discovery process. Therefore, if futures market is unable to dominate the price discovery process, it does not imply that futures market is inefficient; it might be having higher transaction costs or trading restrictions on market participants which impede the futures market in leading the price discovery process.

The cost of trading in index futures is much lesser as compared to that of trading in the portfolio of index stocks. As a result, index futures lead the price discovery process (Fleming et al., 1996). Index futures are traded continuously as a single product rather than a portfolio of stocks. Hence, its prices keep getting updated with the arrival of new information in the market. On the other hand, index spot price is the weighted average of last prices recorded for its components. Hence, index spot price level does not update that
quickly. Continuous and synchronous trading of all component stocks is hardly possible. Hence, index spot prices lag actual developments in the market. Therefore, Stoll and Whaley (1990) opine that, non-synchronous trading in components of the index is responsible for the dominant role of index futures in price discovery process. The index consists of some thinly traded stocks in which trading are very infrequent. Hence, index spot price is affected by stale pricing issue. Therefore, index spot responds slowly to new information in the market as compared to index futures.

On the contrary, speculators are more likely to trade in futures markets due to lower costs and high leverage available in the futures market. It may cause a shift in noise traders from spot to the futures market, which will ultimately reduce the information asymmetry in the spot market and enhances price discovery, and market efficiency. Therefore, spot market may lead the futures market as well. Because of these contradictory arguments mentioned above, there might also be a bi-directional relationship in price discovery process.

3 Literature review

Price discovery in index futures and spot has been studied extensively. However, a substantial part of price discovery literature deals with developed countries’ markets only, whereas the studies regarding emerging and developing markets are very few. Despite the use of different methodologies to examine the lead-lag between index spot and futures, the majority of the findings suggest that index futures leads index spot. The issue of price discovery has received a huge attention in the last decade, and several studies have examined price discovery process in index futures, single stock futures, commodities, bonds, treasuries, foreign exchange, and credit default swaps markets (Chen and Gau, 2010; Fricke and Menkhoff, 2011; Frijns et al., 2015a; Lee and Zeng, 2011; Mizrahi and Neely, 2008; Ryu, 2015; Schultz and Swieringa, 2013; Sehgal et al., 2015; Shastri et al., 2008; Shrestha, 2014).

A recent study by Gong et al. (2016) examine the lead-lag relationship between China Securities Index 300 (CSI 300) of China, Hang Seng Index (HSI) of Hong Kong, Standard and Poor 500 Index (S&P 500) of USA, and their associated futures with the help of a non-parametric approach– thermal optimal path (TOP) method. The results reveal that futures leads spot in well-established index futures markets (HSI and S&P 500), whereas in the case of a developing market, index spot (CSI 300) leads futures. These results are in line with findings of similar previous studies in developed markets, which contend that futures leads spot (Booth et al., 1999; Brooks et al., 2001; Chan, 1992; De Jong and Donders, 1998; Kang et al., 2006; Kavallier et al., 1987; Pizzi et al., 1998; Roope and Zurbruegg, 2002; Tse, 1995). But the results of Gong et al. (2016) contradict with the results of Hou and Li (2012) in the case of CSI 300 index of China. Hou and Li (2012) reveal that CSI 300 index futures play a dominant role in the price discovery process. But findings of Yang et al. (2012) for the same CSI 300 index are in line with Gong et al. (2016) that spot leads futures in case of developing markets. This discussion reveals that price discovery relationship is dynamic and it may vary over different time-horizons and/or across the various countries. Hence, studying price discovery process with recent data in an emerging country like India may give new insights and evidence for price discovery in index futures.
Intraday price discovery in Indian stock index futures market

In the context of India, Pati and Rajib (2011) study price dynamics of S&P CNX Nifty futures and its underlying spot index. They employ vector error correction model and Granger causality on five-minute transaction prices, and the results show that Nifty futures prices lead spot prices. These results are consistent with previous Indian studies (Pati and Padhan, 2009; Swaroop Debasish, 2009). However, some recent studies of individual stock futures have found the dominating role of spot market over futures market in India (Aggarwal and Thomas, 2011; Jain and Biswal, 2015; Kumar and Chaturvedula, 2013; Kumar and Tse, 2009). Some other studies, across the world, have also established that index spot leads index futures (Ivanov et al., 2013; Judge and Reancharoen, 2014; Yang et al., 2012). Moreover, there are some studies which could not determine a conclusive unidirectional relationship but bidirectional feedback with a stronger causality from futures to spot as compared to causality from spot to futures returns (Chan, 1992; De Jong and Nijman, 1997; Pizzi et al., 1998).

Granger causality, cointegration, and common factor methods are based on the assumption of a linear relationship between spot and futures prices Moshiri and Foroutan, (2006). Whereas, the relationship between spot and futures is found to be nonlinear in some studies. Alzahrani et al. (2014) examine the linear and nonlinear causality between spot and futures oil prices using wavelet transformed prices and find a bi-directional causality. Moreover, Trippi and DeSieno, (1992) suggest that a specific neural network based day trading system for S&P 500 index futures contract outperforms passive investments in the index. Furthermore, Gregoriou et al. (2014) apply nonlinear mean reverting unit root on spot and futures prices in European Union emissions trading scheme to test for the efficiency of these markets. This analysis indicates that the relationship between spot and futures could be captured in a better way with the assumption of a nonlinear relationship. Therefore, involving neural network may help in providing improved insights regarding price discovery in index spot and futures as compared to the results based upon linear models. Neural networks find its applications in various domains such as currency prediction (Refenes et al., 1993), stock market prediction (Grudnitski and Osburn, 1993), credit ratings and approvals (Huang et al., 2004), bankruptcy prediction (Atiya, 2001; Kumar and Ravi, 2007; Wilson and Sharda, 1994), and debt risk assessment. However, neural networks have not been used earlier for analysing the price discovery in the equity futures market. This study would be the first attempt, to the best of our knowledge, to understand the nuances of the intraday price discovery in the Indian context.

Although lead-lag relationship or price discovery in Nifty has been studied extensively, there are still inconsistencies in the findings. Therefore, application of neural network methods in the investigation of price discovery dynamics may provide improved insights on the relationship between spot and futures. The majority of the existing literature is based on models with an assumption of a linear relationship between spot and futures returns. If this assumption is not satisfied, the results might be misleading. Overall, the primary objective of this paper is to use neural network methodology in examining the price discovery dynamics between spot and futures of S&P CNX Nifty index to capture the nonlinear relationship between spot and futures returns.
4 Data description

One of the most important considerations while designing any neural network configuration is deciding the frequency and size of the data to be used in the study. The frequency of the data should be determined by the forecast horizon. If the forecast horizon is very short, then high-frequency data would be preferred; whereas for the longer horizon, daily, weekly, or monthly data would be preferred. In the present study, we want to examine the intraday price discovery of S&P CNX Nifty index spot and futures prices, which is a short-horizon for analysis. Hence, we would be using minute-by-minute closing price data for S&P CNX Nifty index spot and futures prices.

Another equally important consideration in neural network configuration is choosing the length of the data. Neural network generally performs better with more data points. However, this is not necessarily the case with a financial time-series because economic conditions changes in a country over the time and such changes could affect the financial time-series. Therefore, the use of irrelevant (very old) information might affect the prediction results of the neural network negatively. Hence, we would use the latest time-series data (January 1, 2015 to August 31, 2015) while examining the price discovery process in the Indian equity futures market.

We would use minute-by-minute high-frequency data for a sample period of 167 trading days ranging from January 1, 2015 to August 31, 2015. Minute-by-minute closing price data for S&P CNX Nifty index spot and futures has been generated from tick-by-tick transaction data provided by National Stock Exchange of India (NSE). Since data has been obtained directly from the stock exchange, data authenticity is quite high as compared to other data sources. To make a continuous futures price series, the nearest month futures contract has been considered by rolling it to next near month series on one day before the expiry date. Spot and futures prices have been matched to make an equally spaced one-minute synchronised time-series. Normal trading hours for spot and futures trading in NSE are from 09:15 AM to 03:30 PM. Continuously compounded returns for spot and futures have been computed by taking logarithmic differences such as:

\[
\text{Return}_t = \ln \frac{\text{Price}_{t}}{\text{Price}_{t-1}}
\]

Consistent with the literature, first five minutes of each trading day have been discarded to avoid overnight returns, and noisy trades in opening time of market Guo et al. (2013). Subsequently, time-series analysis has been conducted on continuous trading session only i.e. from 09:21 AM to 03:30 PM, which makes the sample observations 61,790 (167 trading days with 370 one-minute time intervals) closing prices returns. Among the above 61,790 observations, training has been done with 61,420 observations (166 days), and then the model is used for forecasting of 25 observations for the very next trading day (last trading day of the sample). Nifty is a very liquid security, and the probability of non-trading in Nifty is zero. There are no outliers found in data evaluation process, and the graph of prices show that there is no structural break in data in this duration. The graphical plot of spot and futures prices has been shown in Figure 1.

The descriptive statistics of the returns have been shown in Table 1. Table 1 indicates that mean returns for spot market is –2.29, whereas it is –1.15 for futures market that is slightly higher as compared to the futures market. Average higher returns in futures market may, therefore, attract more trading. The standard deviation in futures is 428.76 which are also higher as compared to 398.09 of the spot market. Besides, the kurtosis for
futures return series is also greater than that of spot market, which suggests that futures market is more volatile and having fat tails. The Jarque-Bera test statistic indicates that both series are not distributed normally.

**Figure 1** Spot and futures prices for entire sample

![Spot and futures prices for entire sample](image)

**Table 1** Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Spot returns</th>
<th>Futures returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-2.29</td>
<td>-1.15</td>
</tr>
<tr>
<td>Median</td>
<td>5.91</td>
<td>0.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>5,178.40</td>
<td>6,113.46</td>
</tr>
<tr>
<td>Minimum</td>
<td>-5,614.39</td>
<td>-6,220.03</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>398.09</td>
<td>428.76</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.37</td>
<td>-0.21</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>11.88</td>
<td>12.27</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>204,334</td>
<td>221,696</td>
</tr>
<tr>
<td>Probability</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: Mean, median, maximum, minimum, and standard deviation values have been multiplied by 1,000,000 for better comparison of both markets

## 5 Neural network methodology

This study has used neural network models for investigating the price discovery relationship between spot and futures prices of Indian equity index S&P CNX Nifty. We have used neural networks in this study because of its ability to capture the nonlinear relationships, patterns, and associations in the data. The conventional techniques like correlation technique (Finnerty and Park, 1987), three-stage least square regression technique (Kawaller et al., 1987), cross-correlation technique (Herbst et al., 1987), cointegration and error correction models (Wahab and Lashgari, 1993), and GARCH models (Abhyankar, 1995) are based on the assumption of linear relationship in the data.
Such conventional techniques are unable to capture the nonlinearity associated with the time-series data. Besides, neural networks are universal approximators and can approximate a large class of functions with a high degree of accuracy over any other nonlinear models (Chen et al., 2003; Zhang and Qi, 2005). Hence, we have used neural networks for investigating the price discovery process to uncover the lead-lag relationship between spot and futures prices. To understand the lead-lag relationship, this study has used four different neural network models:

a  spot-spot model (SSM)

b  spot-futures model (SFM)

c  futures-futures model (FFM)

d  futures-spot model (FSM).

Neural network models are generally used for forecasting of unknown quantities (output) which depend on considerably a large number of inputs. It is a type of statistical learning model that simulates the structure of biological neural networks. This model ensures that the nonlinearity associated with the data is handled and captured effectively. Neural network consists of an input layer, hidden layer(s) and an output layer. Neural network is defined by the configuration of the network along with the training algorithm used to build the neural network. The configuration of the neural network is characterised by a number of nodes, connections among various nodes and the weights associated with each connection of the nodes Lee and Chen, (2002). The training algorithm used to build a neural network can be feed forward neural network or back propagation neural network. Neural network for SSM has been shown in Figure 2. Neural network for remaining three models have not been shown to conserve the space.

Figure 2  Neural network for SSM (see online version for colours)
This study has used feed-forward neural networks with a single hidden layer and lagged returns of spot and futures market as inputs for forecasting. Since we are dealing with time-series data consisting of lags, feed-forward neural network has been used in this study. All the four models repeat themselves 20 times, with random initial weights every time. The results of all these 20 models are then averaged for computing the forecasted values, and accuracy of the model. The configuration of neural network is selected based upon trial and error method from a large number of parameters. These parameters are some hidden layers, the number of nodes in each hidden layer, etc. Thus, the final configuration of neural network is selected based on the least error values, and the maximum accuracy obtained for each model. The selected lag length to be used as input nodes in input layer is eight, which is obtained by Schwartz information criterion (SIC) in a vector auto regression (VAR) framework. The fitted model of neural network is called an NNAR \((p, P)\) model which is analogous to an ARIMA \((p, 0, 0) (P, 0, 0)\) model but it also incorporates nonlinear functions\(^2\). Here \(p\) is a number of non-seasonal lags and \(P\) is a number of seasonal lags used as inputs (Hyndman et al., 2015).

The data have been normalised in the range of \([-1, 1\]} by using min-max normalisation technique as shown below in equation (1) as proposed by Han et al. (2011)

\[
x_i' = \frac{x_i - A_{\text{min}}}{A_{\text{max}} - A_{\text{min}}} A'_{\text{max}} - A'_{\text{min}} + A_{\text{min}}
\]

Here \(A_{\text{max}}\) and \(A_{\text{min}}\) are the original maximum and minimum values of an attribute \(A\) (here attribute implies the series of spot or futures returns). This normalisation technique linearly transforms the original value \(x_i\) to \(x_i'\) in the new range of \(A_{\text{min}}'\) to \(A_{\text{max}}'\). For this study, \(A_{\text{min}}' = 1\) and \(A_{\text{max}}' = 1\). This is done to ensure that the normalised values occur in the range of \([-1, 1\]. Normalisation scales all the inputs in the same range, and thus helps in speeding up the learning phase of prediction algorithms. The four modes used in the study are SSM, SFM, FFM and FSM as shown in Figure 3. A brief description of these models is given below.

**Figure 3** Four neural network models: (a) SSM (b) SFM (c) FFM (d) FSM (see online version for colours)
5.1 Spot-spot model

This model uses returns of spot market as input nodes and then forecasts one step ahead, i.e., next minute, returns of the spot market. The number of optimum lags is eight, obtained as per SIC, which implies a total of eight nodes would be used as inputs in the input layer of SSM. There is only one hidden layer consisting of four nodes (which is half of the number of nodes in the input layer), and an output layer consists of only one node, which forecasts one step ahead returns of the spot market. The model configuration consists of overall 41 links connecting input layer, hidden layer and output layer. The dataset has been split into training and testing dataset. Training dataset consists of the data from January 1, 2015 to August 30, 2015, which corresponds to a total of 61,420 records in the training period. The model has been tested on the first 25 records (out-of-sample forecast) of spot returns of August 31, 2015. Figure 4(a) reveals the trend of last 100 records of the data in the training dataset and forecasts next 25 records of the test-set using SSM.

Figure 4 Forecasting by all four models (a) SSM (b) SFM (c) FFM (d) FSM (see online version for colours)

5.2 Spot-futures model

This model uses returns of spot market as well as returns of futures market as the input nodes and then forecasts the returns of the spot market. The lags obtained used in SFM model is eight, according to the SIC. The configuration of neural network for SFM model
is 16 nodes (eight for spot + eight for futures) in the input layer, eight nodes in the hidden layer and one output node. Similar to SSM, SFM also forecasts the returns on the spot market. The training and testing data have been split in the same way as in SSM. Training dataset consists of observations from returns on spot as well as returns on futures for the duration from January 1, 2015 to August 30, 2015, which corresponds to a total of 122,840 records in training period and the model has been tested on first 25 records (out-of-sample forecast) of spot returns of August 31, 2015. Figure 4(b) reveals the trend of last 100 records of the data in the training dataset and forecasts next 25 records of the test-set using SFM.

5.3 Futures-futures model

This model uses returns of futures market as the input nodes and then forecasts the return of futures market. The lags used for FFM model is also eight. In this model, input layer consists of eight nodes; hidden layer consists of four nodes and output layer gives the forecast value of the one step ahead futures returns. The training and testing datasets have been split in the similar fashion to that of SSM. Figure 4(c) shows the trend of last 100 records of the data in the training dataset and then forecasts next 25 records of the test-set using FFM.

5.4 Futures-spot model

This model uses returns of spot and futures markets as the input nodes and then forecasts the one step ahead returns of the futures market. The lags used in FSM model are eight, and the configuration of neural network for FSM is 16:8:1. It means, there are 16 nodes (eight for spot + eight for futures) in the input layer, eight nodes in the hidden layer, and one output node. This model forecasts the returns on the futures market. The training and testing data have been split in the same manner as that of SFM. Figure 4(d) reveals the trend of last 100 records of the data in the training dataset and then forecasts next 25 records of the test-set using FSM.

5.5 Comparative analysis of all models

This study has used root mean square error (RMSE) and mean absolute error (MAE) for measuring the forecasting accuracy of all models. The values for RMSE and MAE for all the four models, viz. SSM, SFM, FFM, and FSM are shown in Table 2.

| Panel A: investigation of causality from futures returns to spot returns |
|-------------------------------------------------|-----------------|-----------------|-----------------|-----------------|
| Model 1 (SSM)                                   | Model 2 (SFM)   | Improvement (%) |
| Training                                       | Testing        | Training        | Testing        | Training        | Testing        |
| RMSE 0.0367                                    | 0.3269         | 0.0365          | 0.1811         | 0.54            | 44.60          |
| MAE 0.0255                                     | 0.2260         | 0.0254          | 0.1572         | 0.39            | 30.44          |

Note: RMSE stands for root mean squared error, and MAE stands for mean absolute error.
Table 2  Comparative analysis of all models (continued)

<table>
<thead>
<tr>
<th></th>
<th>Model 3 (FFM)</th>
<th></th>
<th>Model 4 (FSM)</th>
<th></th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Testing</td>
<td>Training</td>
<td>Testing</td>
<td>Training</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0347</td>
<td>0.3405</td>
<td>0.0345</td>
<td>0.1281</td>
<td>0.57</td>
</tr>
<tr>
<td>MAE</td>
<td>0.0241</td>
<td>0.2558</td>
<td>0.0239</td>
<td>0.0932</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Note: RMSE stands for root mean squared error, and MAE stands for mean absolute error.

6 Discussion of results

Overall, the results of this study indicate that there exists a bi-directional lead-lag relationship, but the causality from spot returns to futures returns is more dominant as compared to that of running from futures returns to spot returns. Panel A of Table 2 shows the results of investigation of causality from futures returns to spot returns. It reveals that when both spot and futures returns are used for modelling spot returns rather than using only spot returns as input nodes, the RMSE for training sample has reduced from 0.0367 to 0.0365 which is a minor improvement of just 0.54% in the training of the model. Similarly, MAE for training sample has reduced from 0.0255 to 0.0254 which is again an improvement of just 0.39% in the training of the model. However, when we look at the RMSE and MAE for a testing sample for both the models (SSM and SFM), the similar interpretation could be inferred but with major improvements. When both spot and futures returns are used for forecasting spot returns rather than using only spot returns, the RMSE for the testing sample has reduced from 0.3269 to 0.1811, which is a huge improvement of 44.60% in the testing of the model. Similarly, MAE for the testing sample has reduced from 0.2260 to 0.1572 which is also big improvement of 30.44% in the testing of the model.

Furthermore, panel B of Table 2 shows the results for investigation of causality from spot returns to futures returns. When both spot and futures returns are used for modelling futures returns rather than using only futures returns as the input nodes, the RMSE for training sample has reduced from 0.0347 to 0.0345, which is a very minor improvement of only 0.57% in the training of the model. Similarly, MAE for training sample has reduced from 0.0241 to 0.0239 which is also very small improvement of 0.82% in the training of the model. However, when we look at the RMSE and MAE for a testing sample for both the models in panel B (FFM and FSM), the similar interpretation could be inferred but with major improvements. When both spot and futures returns are used for forecasting futures returns rather than using only futures returns, the RMSE for the testing sample has reduced from 0.3405 to 0.1281, which is a huge improvement of 62.37% in the testing of the model. Similarly, MAE for the testing sample has reduced from 0.2558 to 0.0932 which is also big improvement of 63.56% in the testing of the model.
Intraday price discovery in Indian stock index futures market

It can also be observed that model 4 (FSM) emerges as the best model out of all the models, based on the values of the error terms (RMSE and MAE) in Table 2. The RMSE and MAE values of model 4 (FSM) for the testing sample are 0.1281 and 0.0932 respectively, which are the lowest amongst the error terms of all the four models. Thus, it could be stated that the spot market is more dominant in the price discovery process vis-à-vis futures market. Overall, it can be inferred that improvements in the models occur bi-directionally. However, the improvement in the model is better when spot and futures returns are used to model futures returns rather than spot returns. These results imply that the spot market is more agile in impounding new information vis-à-vis futures market. The plausible reasons might be the lower impact cost in the spot market as suggested by Kumar and Chaturvedula (2013).

7 Conclusions

The current study confirms the bi-directional lead-lag relationship between spot and futures prices of S&P CNX Nifty index, traded on NSE. However, the price discovery from spot to futures is more dominant as compared to that of futures to spot. The possible reasons for leading role of spot market could be attributed to the lower impact cost in the spot market as suggested by Kumar and Chaturvedula (2013). These results are in line with the findings of many previous studies regarding price discovery between spot and futures prices of Nifty index (Kumar and Chaturvedula, 2013; Kumar and Tse, 2009). However, our results are not consistent with the findings of some of the previous studies in the Indian context (Inani, 2017; Pati and Rajib, 2011; Swaroop Debasis, 2009).

The major contribution of this paper is that it is the first attempt to analyse the price discovery process of S&P CNX Nifty index spot and futures with the help of the neural network on high-frequency data of one minute, which is very rare for emerging markets. Price discovery has significant implications for market participants, regulators, and exchanges. First, price discovery concept is related to market efficiency. Second, investigation of price discovery can help in designing market structure.

The major limitation of the study is the small sample period of eight months only, which could be extended in future studies. Second, this paper has examined only Indian market. Hence, the results could not be generalised to other emerging countries. Therefore, future studies may include other emerging markets (such as BRICS) and conduct further analysis to compare and understand the phenomenon of price discovery. Third, this paper has not included the time-variation in price discovery analysis (Avino et al., 2015), which could be a further dimension for research. Fourth, this paper has only identified the direction of price discovery, but not determinants of price discovery. Some recent studies have identified the determinants of price discovery (Frijns et al., 2015a, 2015b), which could be a fruitful domain for future research. An obvious extension of this study could be an analysis of price discovery process using per second data points. Finally, on methodology front, neural networks tend to be slower than other methods mainly because of the fact that it computes the weight of each node and connection separately, thereby increasing the execution time and time complexity. This limitation can be eliminated with the aid of high-performance parallel computing environment.
References


Intraday price discovery in Indian stock index futures market


Intraday price discovery in Indian stock index futures market


Notes
2 NNAR and ARIMA stand for neural network auto-regressive and auto-regressive integrated moving average respectively.
3 Each connection from the input layer to hidden layer corresponds to $(8 + 1) 4 = 36$; here, 8 is the number of nodes in the input layer, and 1 is the error term associated with each node in hidden layer. Similarly, each connection from hidden layer to output layer corresponds to $(4 + 1) 1 = 5$; here 4 is the number of nodes in the hidden layer, and 1 is the error term associated with the node in the output layer. Therefore, total connections in neural network configuration are $36 + 5 = 41$. 