ON THE RELATIONSHIP OF IMPLIED, REALIZED AND HISTORICAL VOLATILITY: EVIDENCE FROM NSE EQUITY INDEX OPTIONS

Puja PADHI¹, Imlak SHAIKH²

¹Department of Humanities and Social Sciences, Indian Institute of Technology Bombay, Mumbai 400 076, India
²Department of Management, Birla Institute of Technology and Science, Pilani, Rajasthan, 333 031, India
E-mails: ¹pujapadhi@iitb.ac.in; ²imlak786@gmail.com (corresponding author)

Received 14 December 2011; accepted 04 April 2013

Abstract. This study examines the information content of implied volatility, using the options of the underlying S&P CNX Nifty index. In this study, implied, historical and realized volatilities are calculated using non-overlapping monthly at-the-money samples. The study covers the period from introduction of options on the derivative segment of NSE, June 2001 to May 2011. The results reveal that call and put implied volatility of S&P CNX Nifty index option does contain information about future realized return volatility. This study accounts for the problem of error-in-variable and controls for it by using the instrumental variable technique. In the 2SLS estimation, the Hausman H-statistic shows that call implied volatility is measured with error. Hence, 2SLS coefficients are more consistent than the OLS estimates. Results of this study might prove to be helpful to the volatility traders in volatility forecasting and option pricing.

Keywords: call implied volatility, put implied volatility, realized volatility, historical volatility, measurement errors, two stage least squares, instrumental variable.

JEL Classification: G13, G14, C53.

Introduction

Implied volatility contains enough information such that the traders can have enough information about realized volatility. This issue has gained popularity in the last two decades among the practitioners, analysts and academics. Implied volatility is obtained by inverting the Black-Scholes (BS) option pricing model for the given market price of the option. It is the volatility that is obtained when input into the BS option pricing model yields the price of the option. Implied volatility is the function of time to expiration, spot price, exercise price, market price of an option, risk-free rate of interest, and dividend. A very limited literature is available on the implied volatility dynamics. There is a lack
of studies on emerging markets like India. A report on the Indian Securities Market\(^1\) indicates that Indian derivatives market has reached its maturity level. The National Stock Exchange (NSE) of India has introduced a number of derivative products in the F&O segment like currency derivatives, interest rate derivatives, and recently, it also constructed the implied volatility index (India VIX). Further, NSE has proposed to launch some volatility products (i.e., futures and options on India VIX). This study validates the initiative of NSE.

This work is motivated by the two important issues. First, there is a lack of studies in the context of a developing economy. Second, there is still a lack of understanding due to contradictory results of previous studies (Day, Lewis 1992; Canina, Figlewski 1993; Lamoureux, Lastrapes 1993; Jorion 1995; Jackwerth, Rubinstein 1996; Chance 2003; Koopman \textit{et al.} 2005; Filis 2009). This study also revisits the information content of implied volatility on the S&P CNX Nifty index options, but this study in particular, is significantly different from the earlier studies in several ways. The novel aspects of the study are: (i) This study aims to examine the emerging market nation – India’s stock market volatility; (ii) Forecasting period: Gwilym and Buckle (1999) suggest that the best method of forecasting may depend on the length of the forecasting period. This is one of the comprehensive studies that take into account a longer time series unlike the previous studies. In this study, 120 non-overlapping samples of implied volatility is considered, similar to Christensen and Prabhala (1998) who had taken 139-month data of the OEX options market. However, Panda \textit{et al.} (2008) studied only 42 months for the initial period of introduction of options on NSE F&O’s segment. During this period the market was in a nascent stage and less trading volume in options was observed; (iii) Recent data. One more motivating factor of this study is that it deals with more recent data and highest trading volume in option trading; (iv) The strongest point of the study is the distinction between realized and historical volatility, which has not been ventured in any of the previous studies. They simply considered lagged realized volatility as historical volatility, while distinguishing these two may result in more accurate measure of \textit{ex-post} volatility as the problem of multicollinearity can be controlled in multiple regressions; (v) This study complements the study of Panda \textit{et al.} (2008), but Panda \textit{et al.} used only three measures, viz., call implied volatility, put implied volatility and realized volatility, while this study employs some more variables for more precise estimation. Call and put implied volatility are separately calculated. Ultimately, the results show less measurement errors in the measures of \textit{ex-post} and \textit{ex-ante} volatilities; and (vi) The method (2SLS) used in this work is a novel approach as compared to the

---

\(^1\) For more details see Indian Securities Market: a Review (2012), National Stock Exchange of India Limited, vol. XV:

i. About 1,753 FIIs are trading on the NSE.

ii. India stood at the second position after China among the emerging markets with $1,015,370 mn market capitalization.

iii. India is one of the most successful developing countries with regard to being a vivacious market for exchange-traded derivatives.

iv. India stood at third position among the top five exchanges of various derivative contracts.

v. In addition, NSE holds second rank among the top five exchanges with regard to the number of stock index option contracts traded in 2011.
methods (e.g., ARCH and GARCH) used in previous studies (see Engle 2001; Lehar et al. 2002). This is because sometimes emerging markets suffer from liquidity problem, and poor liquidity results in Error-in-Variable (EIV) problem. Thus, the study controls for EIV problem using 2SLS method that provides consistent estimates. In addition, its contribution to the literature is two-fold. First, it establishes a relationship between ex-ante and ex-post volatility in the emerging options market. Second, it extends the literature of market efficiency based on options market using BS options pricing model.

1. Literature review

The innovative work of Black and Scholes (1973) and Merton (1973) in the area of option pricing has made it possible to study implied volatility. Recently, this option pricing model has become popular among the academicians, practitioners and policy makers. According to BS option pricing model, if the market is efficient implied volatility appears to be an unbiased and efficient predictor of future ex-post return volatility. Implied volatility should subsume the information contained in all other variables used to explain future realized volatility.

Whether option prices impound some information about the future volatility or not has been discussed widely for the last three decades. There is still some inconclusive evidence present in the available literature that makes this topic more contentious. There are some academicians and practitioners like (Latané, Rendleman 1976; Chiras, Manaster 1978; Beckers 1981; Day, Lewis 1992; Jorion 1995; Christensen, Prabhala 1998; Hansen 2001; Christensen, Hansen 2002; Szakmary et al. 2003; Corrado, Miller 2005; Panda et al. 2008; Li, Yang 2009; Shaikh, Padhi 2013a, 2014a) are in the favor of implied volatility as the best predictor of future realized return volatility. On the other hand, some scholars are quite suspicious about market efficiency and the predictive power of implied volatility. Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), Gwilym and Buckle (1999), and Filis (2009) present mixed conclusions on the information content of option prices and the predictive power of implied volatility. However, some scholars like Jackwerth and Rubinstein (1996), Chance (2003), and Koopman et al. (2005) strongly oppose the information content of implied volatility. They show that there is no correlation between implied and realized volatility. They also show that historical volatility outperforms future realized volatility, and is the best predictor of future realized volatility.

The volatility traders invest in the call and put option to trade implied volatility. Hence, correct valuation of derivatives is of crucial importance for practitioners in any financial market (e.g. see Shaikh, Padhi 2014c). Lehar et al. (2002) empirically tested the performance of option pricing model, using BS, GARCH and Stochastic Volatility (SV) model. They found that a more complex volatility model outperforms the simpler implied volatility approach. Thus, they conclude that the selection of pricing model depends upon the pricing performance. The second strand of the literature (e.g. Wong, Tu 2009; Chung et al. 2011; Shaikh, Padhi 2013b,c; 2014a,b; 2015) on Implied Volatility Index (VIX) also supports the previous studies and show that VIX is the best forecast of future volatility.
The main purpose of the present work is to examine the information content and predictive power of implied volatility, based on Nifty index options. The study shows that call and put implied volatility of S&P CNX Nifty index option does contain information about future return volatility. The slope of historical volatility is observed to be “upward biased” and slope of implied volatility is “downward biased”. This is due to the potential problem of measurement errors\(^2\). The problem of EIV is controlled using instrumental variable estimation. The study documents that historical index return volatility does not add any significant information beyond that which is already contained in the option prices. In 2SLS estimation \(H\)-stat shows that call implied volatility is measured with errors and 2SLS estimate is more consistent than the OLS. The significance of the study is twofold: it is helpful in the volatility forecasting and option pricing.

The rest of the work is organized as follows: Section 2 presents data and sampling procedure, Section 3 reports the definition of variables, and Section 4 describes the methodology used for analysis of relationship between implied and realized volatility. Then, the empirical results are discussed in Section 5, and finally, the conclusions are offered in the last Section.

2. Data and sampling technique

The present study is based on the index options (one month) for S&P CNX Nifty. The NSE introduced trading in index options on June 4, 2001. The options contracts are of European type and cash settled. The study period is from June 4, 2001 to May 31, 2011, which covers the entire period of introduction of options on derivative segment of NSE. The Nifty options contracts expire on the last Thursday of the expiry month. If the last Thursday happens to be a trading holiday, the contracts expire on the previous trading day. Base prices are calculated using the BS option pricing model. The BS model for option pricing gives the price of a call and put option as follows:

\[
\begin{align*}
    c &= SN(d_1) - Xe^{-r(T-t)}N(d_2), \\
    p &= Xe^{-rT}N(-d_2) - SN(-d_1),
\end{align*}
\]

where \(d_1 = \{\ln(S/X) + (r + \sigma^2 / 2)(T-t)\} / \sigma \sqrt{T-t}, \ d_2 = d_1 - \sigma \sqrt{T-t}.\)

The variables are: \(S = \) Index price; \(X = \) Strike price; \(T-t = \) Time remaining until expiration, expressed as a percent of a year; \(r = \) Current continuously compounded risk-free interest rate (MIBOR); \(\sigma = \) Annual volatility of stock price (the standard deviation of the short-term returns over one year); \(ln = \) Natural logarithm; \(N(.) = \) Standard normal cumulative distribution function; and \(e = 2.72.\)

The sampling procedure has been used and is different as compared to that used in the previous studies in Indian context. For any options (call/put) to be included in the sample it should be: (i) Traded on business day close to but after expiry date and have expiration on the next expiry date; (ii) Close to at-the-money (ATM), i.e., \((S/X) \in (0.95, 1.05)\). 

\(^2\) Measurement errors occur when true values of implied and realized volatility are not observed.
5) where \( S_t \) is the index level and \( X_t \) is the strike price of the option; and (iii) is traded actively, i.e., have relatively high trading volume. Here Criterion (i) is used to avoid the overlapping of data. S&P CNX Nifty index options of the NSE follow the European style and expire on the last Thursday of the month. Every month three contracts are introduced: near month, two-month, and far month. For the present study only near month (one-month) contracts are taken into account for sampling purpose. Therefore, every year there will be 12 non-overlapping samples for call and put options. Criterion (ii) is used because the option pricing model for calculating implied volatility is more accurate for close to ATM options. Thus, implied volatility obtained from these options may result in less measurement errors. ATM options may have been thinly traded and their prices would not necessarily reflect market price, thus Criterion (iii) is required. Thus, sampling is done for the Nifty index options for the period June 4, 2001 to May 31, 2011 taking into account the above three criterion. The sampling process is repeated up to 120 monthly observations for each call and put options.

3. Definition of variables

3.1. Time-to-maturity \((T - t)\)

In this study, the life of an option is about one month ranging from 27 to 34 days (about 18 to 23 trading days). Here, \((T - t)/365\) indicates days-to-expiration which is measured by number of days from the \( t \) business day and the day immediately prior to expiry day divided by the number of calendar days per year which is 365. Here, the expiry day is not taken into account because all contracts expire on the day of expiration and are cash settled.

3.2. Implied volatility \((\sigma_{CV,t} \text{ and } \sigma_{PIV,t})\)

Implied volatility is the current volatility of a stock as reflected by its option price. Option pricing models cannot be inverted very easily, so implied volatility is calculated numerically. Implied volatility is estimated using the BS option pricing model employing the bisection method as follows:

\[
\text{Volatility Estimate} = \sigma_L + \frac{C - C_L}{C_H - C_L}(\sigma_H - \sigma_L),
\]

where \( \sigma_L \) and \( \sigma_H \) are the low and high volatility values respectively, \( C_L \) and \( C_H \) are the corresponding options values and \( C \) is the market price of the option.

3.3. Realized volatility \((\sigma_{RV,t})\)

Shu and Zhang (2003) suggest that the predictive power of implied volatility can be improved and the measurement errors can also be minimized by constructing a more suitable measure of realized volatility. Realized volatility is calculated as the standard deviation of the daily index log-return during the remaining life of the option. Let \( n \) be the number of trading days before the expiration of an option, \( S_i \) be the index level, and \( R_i \) be the log-return on the \( i^{th} \) day during the remaining life of the option. Then, realized volatility is defined as follows:
where $i = 1, 2, 3, \ldots n,$

$$R_i = \ln \left( \frac{S_i}{S_{i-1}} \right),$$

$$\sigma_{RV,t} = \sqrt{\frac{252}{n-1} \sum_{i=1}^{n} (R_{i,t} - \bar{R}_t)},$$

where $\bar{R}_t = \frac{\sum_{i=1}^{n} R_i}{n}$ denotes the mean of daily log-return of the index at time $t$.

### 3.4. Historical volatility ($\sigma_{HV,t-1}$)

In previous studies, historical volatility at time $t$ is often defined as realized volatility at time $t-1$. If the above-mentioned technique of measurement is followed, then the information contained in the gap between the two consecutive contracts would be ignored (see Hansen 2001). Thus, for the present study different definition of historical volatility (as followed by Hansen 2001; Christensen, Hansen 2002; Li, Yang 2009) is used for a given contract with $T$ days to maturity at time $t$. The corresponding historical volatility is calculated by using the daily log-return of the period going back $T$ days from the time $t$ as follows:

$$\sigma_{HV,t-1} = \sqrt{\frac{252}{T-1} \sum_{i=1}^{T} (R_{i,t-1} - \bar{R}_{t-1})},$$

where $\bar{R}_{t-1} = \frac{\sum_{i=1}^{T} R_i}{T}$ denotes the mean of daily log-return of the index at time $t-1$.

### 3.5. Average implied volatility ($\sigma_{AVRIV,i,t}$)

To use all the months in the present dataset the new implied volatility measures are constructed (see Hansen 2001), as an average of both the call and put implied volatility. The average implied volatility $\sigma_{AVRIV,i,t} (i = 1, 2, 3)$ is defined as follows:

$$\sigma_{AVRIV1,t} = \frac{1}{2} \sigma_{c,t} + \frac{1}{2} \sigma_{p,t},$$

$$\sigma_{AVRIV2,t} = \frac{1}{2} \sigma_{c,t} + \frac{1}{2} \sigma_{p,t},$$

$$\sigma_{AVRIV3,t} = \exp\left\{\frac{1}{2} \ln \sigma_{c,t} + \frac{1}{2} \ln \sigma_{p,t}\right\},$$

where the implied volatility measures $\sigma_{AVRIV1,i}$ is constructed such that $\sigma_{AVRIV1,i}$ is the average of the implied variances, while $\sigma_{AVRIV2,i}$ is an average of the implied volatilities. The third measure is obtained by averaging the natural logarithm of the implied volatilities.

### 4. Methodology

From the review of past studies on the relationship between implied, realized and historical volatility, it is observed that option prices give the ex-ante volatility. In this work, an attempt has been made to show that the Indian option market is an efficient market.
and an option contains all the good and bad information. Based on few previous studies (e.g. Christensen, Prabhala 1998; Hansen 2001; Christensen, Hansen 2002; Panda et al. 2008; Li, Yang 2009) and on the assumption of rational expectation and efficiency of the market, the following models have been used for the purpose:

### 4.1. Conventional and encompassing OLS

Conventional OLS gives elementary results on the relationship between implied and realized volatility. It is based on the traditional measures of implied volatility, and therefore called conventional OLS:

\[
\sigma_{RV,t} = \alpha_o + \alpha_c \sigma_{CIV,t} + \alpha_p \sigma_{PIV,t} + \alpha_{HV} \sigma_{HV,t-1} + u_t. \tag{8}
\]

Similarly \( \sigma_{RV,t} = \alpha_o + \alpha_i \sigma_{AVRIV_i,t} + u_t \), where \( i = 1, 2, 3 \).

The present study tests four hypotheses:

**H1**: The intercept should be zero, \( \alpha_o = 0 \);

**H2**: The slope coefficients of call and put implied volatility should be equal to one, \( \alpha_c = 1, \alpha_p = 1 \);

**H3**: The coefficient of historical volatility should be zero, \( \alpha_{HV} = 0 \); and

**H4**: The residuals in all OLS estimation are white noise.

In encompassing OLS estimation one-period lagged values of dependent variable (\( \sigma_{RV,t} \)) are used, hence it is called encompassing regression. In encompassing regression the slope of lagged realized volatility is supposed to be zero:

\[
\sigma_{RV,t} = \delta_o + \delta_c \sigma_{CIV,t} + \delta_p \sigma_{PIV,t} + \delta_{RV} \sigma_{RV,t-1} + \epsilon_t. \tag{10}
\]

Encompassing regression model specified for average implied volatility is:

\[
\sigma_{RV,t} = \delta_o + \delta_i \sigma_{AVRIV_i,t} + \delta_{RV} \sigma_{RV,t-1} + \epsilon_t, \tag{11}
\]

where \( i = 1, 2, 3 \).

### 4.2. Instrumental variable estimation

The present study deals with instrumental variable technique because of measurement errors. The problem of measurement errors was put forward by Harvey and Whaley (1991). There are several sources of measurement errors such as: (i) Unrealistic assumption of BS option pricing model (see Chance 2003); (ii) Dividend consideration; (iii) Non-synchronosity of index level (see Jorion 1995); (iv) Bid and ask spread of option prices; (v) Non-simultaneous price problem; (vi) Infrequent trading (see Harvey, Whaley 1991); and (vii) The jumps in the price level. Generally, EIV problem arises when regressors are correlated to the residuals, and as a result the slopes are biased and underestimated. To diagnose the presence of EIV the following auxiliary regression was suggested by Hausman (1978). Two different regressions are estimated. First, the suspected variable is regressed over the instruments and exogenous variables, and second, the original regression specification is estimated with additional regressors by including the residuals from the first regression.
First specification:

Here, an illustration for call implied volatility is presented:

\[ \sigma_{CIV,t} = \delta_o + \delta_1 \sigma_{CIV,t-1} + \delta_2 \sigma_{HV,t-1} + \epsilon_t, \]  

Eq. (12) gives the values of residual \( \epsilon_t \) which is used in the second regression as an explanatory variable.

Second specification:

\[ \sigma_{RV,t} = \alpha_o + \alpha_c \sigma_{CIV,t} + \alpha_H \sigma_{HV,t-1} + \alpha_e \epsilon_t + u_t. \]  

In the same way this auxiliary regression is estimated for put and average implied volatility. Measurement errors exist, if the slope of the residuals \( \alpha_e \) is different from zero (in Eq. (13)) and is statistically significant.

### 4.3. 2SLS estimation

To apply instrumental variable technique in 2SLS there should be at least as many instruments as explanatory variables. In this study, lagged call/put implied and historical volatility are considered as instruments. The reason for this is that the volatilities measured before period \( t \) is correlated with this period’s true implied volatility and plausibly uncorrelated with the measurement errors in month \( t \). One more reason for taking more instruments is to increase the explanatory power, i.e., adjusted \( R^2 \). Hausman proposed the \( H \)-statistic to test the null of ‘OLS estimates are consistent’ Vs ‘2SLS estimates are consistent’. \( H \)-Stat = \( (\beta^{2SLS} - \beta^{OLS})^2/\text{Var}(\beta^{2SLS}) - \text{Var}(\beta^{OLS}) \) follows chi-square distribution with one degrees of freedom. Thus, in this study the instrumental variable technique is used through which more consistent estimate can be obtained. In 2SLS procedure implied volatility (call/put) is regressed on instruments and exogenous variables. The fitted values of implied volatility are replaced in the second stage of OLS estimation with original implied volatility.

First stage regression estimates:

\[ \sigma_{CIV,t} = \beta_o + \beta_1 \sigma_{CIV,t-1} + \beta_2 \sigma_{CIV,t-2} + \beta_3 \sigma_{HV,t-1} + u_t, \]  

\[ \sigma_{PIV,t} = \beta_o + \beta_1 \sigma_{PIV,t-1} + \beta_2 \sigma_{PIV,t-2} + \beta_3 \sigma_{HV,t-1} + u_t. \]  

Second stage regression estimates:

\[ \sigma_{RV,t} = \alpha_o + \alpha_c \sigma_{CIV,t} + \alpha_p \bar{\sigma}_{PIV,t} + \alpha_H \sigma_{HV,t-1} + u_t. \]  

Similarly for average implied volatility 2SLS is modeled as follows:

First stage regression:

\[ \sigma_{AVRIV1,t} = \beta_o + \beta_1 \sigma_{AVRIV1,t-1} + \beta_2 \sigma_{AVRIV1,t-2} + \beta_3 \sigma_{HV,t-1} + u_t, \]  

\[ \sigma_{AVRIV2,t} = \beta_o + \beta_1 \sigma_{AVRIV2,t-1} + \beta_2 \sigma_{AVRIV2,t-2} + \beta_3 \sigma_{HV,t-1} + u_t, \]  

\[ \sigma_{AVRIV3,t} = \beta_o + \beta_1 \sigma_{AVRIV3,t-1} + \beta_2 \sigma_{AVRIV3,t-2} + \beta_3 \sigma_{HV,t-1} + u_t. \]  

Second stage regression:

\[ \sigma_{RV,t} = \alpha_o + \alpha_1 \bar{\sigma}_{AVRIV1,t} + \alpha_2 \bar{\sigma}_{AVRIV2,t} + \alpha_3 \bar{\sigma}_{AVRIV3,t} + \alpha_H \sigma_{HV,t-1} + u_t. \]
The main objective of this work is to estimate implied volatility from BS model, not pricing of options. ARCH/GARCH model is a better tool for pricing of option (see Lehar et al. 2002), while the main goal of 2SLS least squares estimation is to obtain a consistent estimate in the presence of measurement errors. The ARCH/GARCH model is mainly concerned with heterosedasticity. The main application of these models is to provide a volatility estimate, which can be used in portfolio selection, pricing of options (see Engle 2001; Lehar et al. 2002), risk management, and assets pricing.

5. Results and analysis

5.1. Descriptive analysis

Table 1 summarizes the descriptive statistics of multivariate time series data. The descriptive statistics for raw series are presented in Panel A and for log-transformed in Panels B. Table 1 shows that the average realized volatilities \( \sigma_{RV,t} \) and \( \sigma_{HV,t-1} \) are smaller than the average put implied volatility \( \sigma_{PIV,t} \). Similar findings were reported by Hansen (2001), Panda et al. (2008), and Li and Yang (2009). This may be due to implementation of portfolio insurance as suggested by Harvey and Whaley (1991). On the other hand, the call implied volatility \( \sigma_{CIV,t} \) is found to be less than all realized volatilities.

Table 1. Descriptive statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>( \sigma_{CIV,t} )</th>
<th>( \sigma_{PIV,t} )</th>
<th>( \sigma_{RV,t} )</th>
<th>( \sigma_{HV,t-1} )</th>
<th>( \sigma_{AVRIV_{1,t}} )</th>
<th>( \sigma_{AVRIV_{2,t}} )</th>
<th>( \sigma_{AVRIV_{3,t}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.212</td>
<td>0.275</td>
<td>0.232</td>
<td>0.230</td>
<td>0.247</td>
<td>0.244</td>
<td>0.240</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.664</td>
<td>0.781</td>
<td>0.715</td>
<td>0.771</td>
<td>0.725</td>
<td>0.722</td>
<td>0.720</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.046</td>
<td>0.096</td>
<td>0.093</td>
<td>0.083</td>
<td>0.095</td>
<td>0.095</td>
<td>0.093</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.093</td>
<td>0.102</td>
<td>0.125</td>
<td>0.128</td>
<td>0.094</td>
<td>0.094</td>
<td>0.095</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>191</td>
<td>154</td>
<td>139</td>
<td>175</td>
<td>211</td>
<td>211</td>
<td>206</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>ln ( \sigma_{CIV,t} )</th>
<th>ln ( \sigma_{PIV,t} )</th>
<th>ln ( \sigma_{RV,t} )</th>
<th>ln ( \sigma_{HV,t-1} )</th>
<th>ln ( \sigma_{AVRIV_{1,t}} )</th>
<th>ln ( \sigma_{AVRIV_{2,t}} )</th>
<th>ln ( \sigma_{AVRIV_{3,t}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-1.634</td>
<td>-1.346</td>
<td>-1.570</td>
<td>-1.582</td>
<td>-1.456</td>
<td>-1.472</td>
<td>-1.490</td>
</tr>
<tr>
<td>Maximum</td>
<td>-0.408</td>
<td>-0.246</td>
<td>-0.334</td>
<td>-0.259</td>
<td>-0.321</td>
<td>-0.324</td>
<td>-0.327</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.415</td>
<td>0.339</td>
<td>0.450</td>
<td>0.461</td>
<td>0.338</td>
<td>0.344</td>
<td>0.356</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>14</td>
<td>3</td>
<td>10</td>
<td>8.659</td>
<td>6.</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Probability</td>
<td>0.001</td>
<td>0.243</td>
<td>0.007</td>
<td>0.013</td>
<td>0.049</td>
<td>0.070</td>
<td>0.159</td>
</tr>
<tr>
<td>Total Number</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
</tbody>
</table>

Notes: Table 1 shows DS for implied, realized and historical volatility for non-overlapping monthly ATM samples. The sampling procedure covers the period from June 4, 2001 to May 31, 2011.
This indicates that investors prefer the call index options least for their portfolio insurance pertaining to the Indian derivatives market. No significant difference is observed between the three combined implied volatility measures. They are found to be higher than realized volatility.

The second row of Panel A shows that maximum put implied volatility is 78%, while maximum call implied volatility is 66%. A comparison of SD’s of all volatility series reveals that both the realized volatilities are more volatile as compared to all other \textit{ex-ante} volatility series. As per the assumption of BS option pricing model, the annualized standard deviation should be constant and violating this may lead to serious measurement problem. The JB-statistic in Panel A indicates that the time series variable does not conform to normality. But, for the log-transformed series (as reported in Panel B), they approximately conform to normality. The last row of Panel B shows the \textit{p}-value of JB-statistic at 1%, 5% and 10% level of significance. The time series variables $\ln \sigma_{PIV,t}$, $\ln \sigma_{AVRIV_{1,2,3},t}$, and $\ln \sigma_{HV,t-1}$ are found to be normally distributed, except call and realized volatility. Therefore, in the econometric investigation of implied volatility, the focus will be on the log-transformed values.

5.2. Time series plot of implied and realized volatility

Figures 1 and 2 show the graphical representation of raw time series and Figures 3 and 4 show that of the log-transformed values. The figures clearly show that implied volatilities do not conform exactly to realized volatilities. There is some gap between the two volatilities. This can be considered as the poor indication of information content of implied volatility as predictor of future realized volatility. This may be due to the causes identified by Harvey and Whaley (1991). This issue has been attributed empirically in the following sections.

As pointed out by Hansen (2001) volatility implied in the European option is an estimate of realized volatility until the expiration of life of the option. As per the assump-
Fig. 2. Time series plot for raw average implied, realized and historical volatility

Fig. 3. Time series plot for log-transformed values of implied, realized and historical volatility

Fig. 4. Time series plot for log-transformed values of average implied, realized and historical volatility
tion of BS option pricing model, volatility remains constant and the same volatility is used to price the European call and put option. Therefore, implied volatility estimated from both call and put options should be identical. However, it is clearly seen from Figures 1 to 4 that call and put implied volatilities are not identical for many months. This indicates measurement of implied volatility with noise.

5.3. Discussion

The empirical work is presented in three phases: First, by conventional OLS; second, by encompassing OLS; and third, by 2SLS using IV’s technique.

5.3.1. Conventional OLS estimation

The first row of Table 2 shows the slope of call implied volatility for log-transformed series as 0.68, while for the raw series it is 0.90 and is statistically significant. Consequently, call implied volatility does contain information about the realized volatility. This primary result provides strong base of the information content of implied volatility of the Nifty index option and supports the past literature (Christensen, Prabhala 1998; Hansen 2001; Li, Yang 2009). These results are different from the null hypotheses. The slope of call implied volatility is less than unity and the intercept is different from zero. This indicates that the call implied volatility is a biased estimate of future realized return volatility.

Table 2. Conventional OLS

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Other stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln \sigma_{RV,t}$</td>
<td>Intercept 0.47*** ln $\sigma_{CIV,t}$ 0.68*** ln $\sigma_{PIV,t}$ 0.82*** ln $\sigma_{HV,t-1}$ 0.54***</td>
<td>$R^2$ 0.38 F-stat 74.58*** LM-test (12) 0.61</td>
</tr>
<tr>
<td>$\ln \sigma_{RV,t}$</td>
<td>Intercept 0.46*** ln $\sigma_{CIV,t}$ 0.82*** ln $\sigma_{PIV,t}$ 0.54*** ln $\sigma_{HV,t-1}$ 0.29</td>
<td>$R^2$ 0.38 F-stat 73.66*** LM-test (12) 0.915</td>
</tr>
<tr>
<td>$\ln \sigma_{RV,t}$</td>
<td>Intercept 0.72*** ln $\sigma_{CIV,t}$ 0.54*** ln $\sigma_{PIV,t}$ 0.20*** ln $\sigma_{HV,t-1}$ 0.29</td>
<td>$R^2$ 0.29 F-stat 50.49*** LM-test (12) 1.211</td>
</tr>
<tr>
<td>$\ln \sigma_{RV,t}$</td>
<td>Intercept 0.42*** ln $\sigma_{CIV,t}$ 0.52*** ln $\sigma_{PIV,t}$ 0.20*** ln $\sigma_{HV,t-1}$ 0.29</td>
<td>$R^2$ 0.40 F-stat 40.04*** LM-test (12) 0.458</td>
</tr>
<tr>
<td>$\ln \sigma_{RV,t}$</td>
<td>Intercept 0.44*** ln $\sigma_{CIV,t}$ 0.65*** ln $\sigma_{PIV,t}$ 0.16 ln $\sigma_{HV,t-1}$ 0.38</td>
<td>$R^2$ 0.38 F-stat 38.14*** LM-test (12) 0.770</td>
</tr>
<tr>
<td>$\ln \sigma_{RV,t}$</td>
<td>Intercept 0.33*** ln $\sigma_{CIV,t}$ 0.36*** ln $\sigma_{PIV,t}$ 0.41*** ln $\sigma_{HV,t-1}$ 0.42</td>
<td>$R^2$ 0.42 F-stat 29.70*** LM-test (12) 0.602</td>
</tr>
</tbody>
</table>

Notes: Table 2 shows OLS results of realized, implied and historical volatility based non-overlapping monthly ATM samples. The LM-statistic tests the hypothesis “there is no autocorrelation”. Asymptotic t-stat is shown in square bracket. ***1%, **5%, *10% significant.

3 Results of the analysis done on raw series are not reported here and can be obtained from the author on request.
Now, the question arises regarding the predictive power of put implied volatility. The slope of put implied volatility is found to be 0.82 and is statistically significant; the analysis reveals that put implied volatility also contains information about future volatility. Comparing the two slopes, the slope of put implied is found to be greater than that of the call implied volatility. This implies that put implied volatility rather than call implied volatility is the best forecast of the future volatility. This also reveals that investors prefer/buy more put options to protect their portfolio. Sinclair (2008) rightly stated that “in order to profit from our forecast of implied volatility we need to hedge, so that our risk is actually the realized volatility. The trader trade in volatility to protect the portfolio through insurance, thus there is a risk premium in buying the insurance”.

The third row of Table 2 shows the slope of historical volatility 0.54 and statistically significant. The plausible reason for additional information content of historical volatility may be the fact that historical volatility does contain more information beyond that contained in call/put implied volatility. This may be the consequences of existence of a serious problem EIV which is discussed in the later sections. It is essential to compare the call(put) implied volatility with historical volatility, therefore some more regressions are estimated. In multiple regressions the slopes of call(put) and historical volatility are found to be 0.52(0.65) and 0.20(0.16) respectively and are statistically significant. On including historical volatility as an additional regressor, the slope of call implied volatility declines from 0.68 to 0.52. This implies that slope of call implied volatility is ‘downward biased’ and that of historical volatility is “upward biased”. This may be due to existence of noise in the measurement of call (put) implied volatility. As shown in the sixth row of Table 2, when call and put implied volatility and historical volatility are taken together as regressors, the estimated slopes of these three regressors are found to be 0.36, 0.41 and 0.06 respectively. In this regression one of the interesting facts is that historical volatility is found to be insignificant.

To make this study more informative some more forecasting measures of future realized volatility are estimated in Table 3. The reason for combining both implied volatility is to reduce the measurement errors. The slopes for log-transformed values are found to be 0.87, 0.86, and 0.83 respectively. The results are found to be more accurate and precise in comparison to the slope obtained from OLS estimation (as shown in Table 2). This implies the presence of potential measurement errors. The slope of $\ln \sigma_{\text{AVRIV1,1}}$ is found to be 0.87 and is statistically significant and more than the slope of separate call and put implied volatility. It is clearly seen that the explanatory power is 0.43 which is higher than all previous regressions. In multiple OLS, when historical volatility is included as an additional regressor one of the interesting facts noticed is that there is no significant change in the explanatory power due to addition of one more regressor. This implies that the previous regression models are not miss-specified.
Table 3. Conventional OLS for average implied volatility

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Other stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnσ_{RV,t}</td>
<td>Intercept</td>
<td>0.30***</td>
</tr>
<tr>
<td></td>
<td>lnσ_{AVRIV1,t}</td>
<td>0.87***</td>
</tr>
<tr>
<td></td>
<td>[-2.15]</td>
<td>[9.45]</td>
</tr>
<tr>
<td>lnσ_{RV,t}</td>
<td>-0.30**</td>
<td>0.86***</td>
</tr>
<tr>
<td></td>
<td>[-2.20]</td>
<td>[9.53]</td>
</tr>
<tr>
<td>lnσ_{RV,t}</td>
<td>-0.33**</td>
<td>0.83***</td>
</tr>
<tr>
<td></td>
<td>[2.46]</td>
<td>[9.49]</td>
</tr>
<tr>
<td>lnσ_{RV,t}</td>
<td>-0.30***</td>
<td>0.82***</td>
</tr>
<tr>
<td></td>
<td>[-2.15]</td>
<td>[5.21]</td>
</tr>
<tr>
<td>lnσ_{RV,t}</td>
<td>-0.30**</td>
<td>0.81***</td>
</tr>
<tr>
<td></td>
<td>[-2.19]</td>
<td>[5.31]</td>
</tr>
<tr>
<td>lnσ_{RV,t}</td>
<td>-0.33**</td>
<td>0.77***</td>
</tr>
<tr>
<td></td>
<td>[-2.43]</td>
<td>[5.28]</td>
</tr>
</tbody>
</table>

Notes: Table 3 shows OLS results of realized, implied and historical volatility based non-overlapping monthly ATM samples. The LM-statistic tests the hypothesis “there is no autocorrelation”. Asymptotic t-stat is shown in square bracket. ***1%, **5%, *10% significant.

5.3.2. Encompassing OLS estimation

Christensen and Prabhala (1998) attempted to examine the information content of realized volatility as a lagged independent variable. The results of estimation of one-period lagged realized volatility with call and put implied volatility are shown in Table 4. Christensen and Prabhala (1998) found the slope of lagged realized volatility to be 0.57. In this study the slope of lagged realized volatility is found to be 0.56 and is statistically significant. The results lead us to believe that the past value of realized volatility does subsume the information regarding the future volatility. But, while estimating the second and third, with call/put implied volatility separately the slope of realized volatility declines dramatically (from 0.56 to 0.22/0.20) and the explanatory power also increases from 0.31 to 0.35/0.39. This signifies that call/put implied volatility is superior to realized volatility in the prediction of future volatility. Finally, in the last row of Table 4, results show that the slope of lagged realized volatility (0.10) is statistically not significant. While the slopes of call and put implied volatility are 0.36 and 0.39 respectively and are statistically significant, the explanatory power is also higher (0.42) as compared to the first three regressions.

Note: The result on Encompassing OLS for average implied volatility remains identical.
Table 4. Encompassing OLS

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Other stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>lnσ_{CIV, t}</td>
</tr>
<tr>
<td>lnσ_{RV, t}</td>
<td>-0.69***</td>
<td>0.56***</td>
</tr>
<tr>
<td></td>
<td>[-5.51 ]</td>
<td>[7.34]</td>
</tr>
<tr>
<td>lnσ_{RV, t}</td>
<td>-0.40***</td>
<td>0.50***</td>
</tr>
<tr>
<td></td>
<td>[-2.99 ]</td>
<td>[4.40]</td>
</tr>
<tr>
<td>lnσ_{RV, t}</td>
<td>-0.42***</td>
<td>0.52***</td>
</tr>
<tr>
<td></td>
<td>[-3.18 ]</td>
<td>[4.04]</td>
</tr>
<tr>
<td>lnσ_{RV, t}</td>
<td>-0.32**</td>
<td>0.358</td>
</tr>
<tr>
<td></td>
<td>[-2.34 ]</td>
<td>[2.79]***</td>
</tr>
</tbody>
</table>

Notes: Table 4 shows encompassing OLS of implied and lagged realized volatility based non-overlapping monthly ATM samples. The LM-statistic tests the hypothesis “there is no autocorrelation”. Asymptotic t-stat is shown in square bracket. ***1%, **5%, *10% significant.

5.3.3. EIV problem and 2SLS

To diagnose the problem of EIV, Hausman (1978) specification test is used and Eqs (12) and (13) are estimated as auxiliary regressions. In the case of auxiliary regression, the slope of estimated residuals $\alpha_e$ is different from zero, but statistically insignificant. However, as the value of slope of residual is non-zero, the presence of EIV problem cannot be rejected. The estimated slopes of residuals for regressions of call and put implied volatility are obtained as 0.29 and 0.35 respectively which are different from zero, and that implies the potential presence of EIV problem. Similarly, the slopes based on the other three combined implied volatility are obtained as 0.34, 0.14 and 0.08 respectively.

The results of first stage regression for traditional measures of implied volatility are obtained using Eqs (14)–(15). In this first stage OLS, instruments are used to obtain the fitted value of call/put implied volatility. Panda et al. (2008) obtained the explanatory power of first stage OLS as only 0.57, while in this study it is found to be 0.71. This happens because of more precise model specification and no serious problem of multicollinearity. The results of 2SLS estimation are shown in Table 5. The first two rows of Table 5 show that the slope of call and put implied volatility is 0.81 (H-stat = 4.83, significant) and 0.90 respectively. The results obtained from the present study are more precise than those of the previous works in the Indian context. The reason might be the use of longer time series data and more improved measure of volatility. In multiple regressions, when historical volatility is included as an additional regressor the slopes of call/put implied volatility are statistically significant, while that of historical volatility is not significant. This implies that implied volatility dominates historical volatility.

On using average implied volatility, one notices that the slope of implied volatility increases dramatically. This may be due to the potential problem of EIV. The results of first stage OLS for combined implied volatilities are obtained using Eqs (17)–(19). Now, these fitted values are replaced in the second stage of OLS. Table 6 shows the first three
### Table 5. 2SLS estimation

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Other stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>ln $\sigma_{CIV,t}$</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.25</td>
<td>0.81***</td>
</tr>
<tr>
<td></td>
<td>[-1.52]</td>
<td>[8.06]</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.36**</td>
<td>0.90***</td>
</tr>
<tr>
<td></td>
<td>[-2.30]</td>
<td>[7.85]</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.28</td>
<td>0.74***</td>
</tr>
<tr>
<td></td>
<td>[-1.50]</td>
<td>[3.06]</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.41**</td>
<td>0.71**</td>
</tr>
<tr>
<td></td>
<td>[-2.41]</td>
<td>[2.45]</td>
</tr>
</tbody>
</table>

**Notes:** Table 5 shows 2SLS results based on Eq. (16). The LM-statistic tests the hypothesis “there is no autocorrelation”. H-stat is reported to compare OLS and 2SLS estimates with one degrees of freedom follows chi-square distribution $H$-statistic $= (\hat{\beta}^{2SLS} - \hat{\beta}^{OLS})^2/Var(\hat{\beta}^{2SLS}) - Var(\hat{\beta}^{OLS})$. Asymptotic t-stat is shown in square bracket. ***1%, **5%, *10% significant.

### Table 6. 2SLS estimation for average implied volatility

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Other stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>ln $\tilde{\sigma}_{AVRIV1,t}$</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.27*</td>
<td>0.89***</td>
</tr>
<tr>
<td></td>
<td>[-1.74]</td>
<td>[8.37]</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.28*</td>
<td>0.88***</td>
</tr>
<tr>
<td></td>
<td>[-1.77]</td>
<td>[8.44]</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.28*</td>
<td>0.86***</td>
</tr>
<tr>
<td></td>
<td>[-1.83]</td>
<td>[8.46]</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.30*</td>
<td>0.81***</td>
</tr>
<tr>
<td></td>
<td>[-1.68]</td>
<td>[2.99]</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.30*</td>
<td>0.81***</td>
</tr>
<tr>
<td></td>
<td>[-1.69]</td>
<td>[3.09]</td>
</tr>
<tr>
<td>ln $\sigma_{RV,t}$</td>
<td>-0.30*</td>
<td>0.81***</td>
</tr>
<tr>
<td></td>
<td>[-1.72]</td>
<td>[3.17]</td>
</tr>
</tbody>
</table>

**Notes:** Table 6 shows 2SLS results based on Eq. (20). The LM-statistic tests the hypothesis “there is no autocorrelation”. H-stat is reported to compare OLS and 2SLS estimates with one degrees of freedom follows chi-square distribution $H$-statistic $= (\hat{\beta}^{2SLS} - \hat{\beta}^{OLS})^2/Var(\hat{\beta}^{2SLS}) - Var(\hat{\beta}^{OLS})$. Asymptotic t-stat is shown in square bracket. ***1%, **5%, *10% significant.
slopes of implied volatility as 0.89, 0.88 and 0.86 respectively and they are statistically significant with explanatory power of 0.43. In next three multiple regressions, when historical volatility is included as an additional regressor, historical volatility appears to be non-significant for all the three regressions. This indicates that historical volatility does not contain any additional information about future volatility other than what is already contained in the option prices. Another interesting result is that when historical volatility is included as an additional regressor, adj. $R^2$ does not change significantly. This shows that univariate regression model is a precise model for forecasting realized volatility.

The Appendix clearly shows that the results of the present work support the results of the previous studies. The results of this study are strongly consistent with those of Hansen (2001) for the Danish options market, which is a thinly-traded option market. Thus, the present study is an extension in the context of Indian options market, as nowadays Nifty options are considered to have the highest trading volume. Volatility trading is mainly useful in finding profitable trades and in risk management. This study educates the volatility traders regarding how volatility can be estimated and how volatility determines the option price.

Conclusions

In this empirical work the implied volatility from call and put options are studied as predictor of future realized return volatility. This study uses the non-overlapping sampling procedure to sample the one-month ATM CNX Nifty index options and its implied volatility measures are calculated. The sampling procedure gives 120 monthly non-overlapping data points which covered a time span of 10 years. The results corroborate the past study on information content and predictive power of implied volatility. The novel aspects of the study are: it deals with longer estimation period; uses different definition of historical and realized volatility; minimizes the measurement error by averaging the call and put implied volatility; and uses more recent data.

The results show that implied volatility (call/put/average) subsumes the information about future realized return volatility. The estimation results show that the slope of historical volatility is ‘upward biased’ and that of implied volatility is ‘downward biased’ and this is due to the existence of the potential problem of measurement errors. During the analysis period, historical volatility is found to be a biased estimate of future volatility and does not contain any additional information other than what is already contained in the option prices. More intuitively, it is concluded that the slopes for average implied volatilities are larger than the slope of call and put implied volatility. This implies that average implied volatilities explain better the future realized volatility. The EIV problem is controlled by employing instrumental variable estimation and it is found that the historical index return volatility did not add any incremental information.

In 2SLS estimation, the $H$-statistic shows that call implied volatility is measured with errors and 2SLS slope is more consistent than the OLS estimate. Based on the findings of this study it is suggested that the investors can use implied volatility as a predictor of future index return volatility in risk management.
References


Shaikh, I.; Padhi, P. 2014b. Inter-temporal relationship between India VIX and Nifty equity index, *Decision*. http://dx.doi.org/10.1007/s40622-014-0046-0


## APPENDIX

<table>
<thead>
<tr>
<th>Previous studies</th>
<th>Call IV</th>
<th>Put IV</th>
<th>Combined IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christensen and Prabhala (1998)</td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hansen (2001)</td>
<td>0.59</td>
<td>0.55</td>
<td>0.52</td>
</tr>
<tr>
<td>Christensen and Hansen (2002)</td>
<td>0.83</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Li and Yang (2009)</td>
<td>0.45</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>Panda et al. (2008)</td>
<td>0.42</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td><strong>Our results</strong></td>
<td>0.68</td>
<td>0.82</td>
<td>0.87</td>
</tr>
</tbody>
</table>

*Source: Author’s calculation.*

Puja PADHI is an Assistant Professor of Economics in the Department of Humanities & social sciences at the Indian Institute of Technology, Bombay. Before joining IITB, she was working as an Assistant Professor at Pondicherry Central University, India. Her current areas of research interests include financial economics, time series econometrics and economic theory.

Imlak SHAIKH is working as an Assistant Professor in the area of Accounting and Finance in Management Department at Birla Institute of Technology & Science, BITS Pilani. Before joining BITS Pilani Dr Shaikh has been worked as a lecturer in KBS Commerce & NATARAJ professional Sciences College Vapi, Gujarat, India. He has published several papers in national and international journals such as: *Transition Studies Review, Economic Change and Restructuring, Decision, Global Business Perspectives, Emerging Market Finance, Margin: Applied Economic Research, Global Business Review, Journal of Indian Business Research.*