Empirical Testing of the Performance of Black and Scholes Option-Pricing Model with Alternative Volatility Measures

Abstract Submitted to the University of Delhi for the Award of the Degree of

Doctor of Philosophy

Research Supervisor
Prof. Y.P. Singh

Research Co-Supervisor
Prof. K.V. Bhanu Murthy

By
Sonal Sharma

Department of Commerce
Delhi School of Economics
University of Delhi
Delhi-110007

December, 2012
ABSTRACT

INTRODUCTION

Options have been considered to be the most dynamic segment of the securities market since the inception of the Chicago Board Options Exchange (CBOE) and introduction of the Black and Scholes (BS) option pricing model in April 1973. Since its invention, the BS formula has been widely used by traders to determine the price for an option. Since development of the model, numerous other complex option pricing models have been developed by various authors. Inspite of this overwhelming body of empirical evidence that refutes to a greater or lesser degree, the strict assumptions underlying the model and comes out with different refinements, extensions and alternatives, the BS model is still very much popular amongst the investors. Perhaps, it is exactly this overwhelming body of empirical evidence that leaves the BS model alive today. It is the only model that every participant in the financial markets understands and every researcher includes as a benchmark model with which to compare his/her research. Though the Black and Scholes (BS) model made a break-through in option pricing, but over time its underlying assumptions have been found to be inconsistent with the observed statistical properties of many asset price processes. Different authors have found different trends in over-or-under-pricing by the BS model in their studies. This contradiction attracts our attention and since very few studies have tested the BS model on Indian data, we through the present study, try to fill this gap. In other words we test the BS model’s pricing and hedging performance by applying it on Nifty index options.

MOTIVATION

Just preceding the recent global financial crisis, there was an explosion of derivative products and markets. The whole phenomenon of growth of structured products and the consequent excessive use of these products for hedging, leveraging and other financial strategies brought into question the entire gamut of financial derivatives. Moreover, the derivatives involved in the crisis were those derivatives, which were not regulated and traded on over-the-counter exchanges, where no market price data is available and the parties involved, sit together and decide upon the price of the derivative. When market prices are not available, the parties involved have to privately decide upon the price of the derivative by applying some model, or probably price them intuitively. This ultimately led to underpricing of the contracts and thereby to the market crisis. The role of a model thus, becomes very important for the market players in assessing the risk involved in the contract and accordingly deciding a price for it. There is a need to evaluate the pricing models so as
to examine their efficacy. The extant literature reviews the Black and Scholes model in context of other markets, but there are only few studies in the Indian context. Particularly, to our knowledge, there is no study which exhaustively tests for the sensitivity of the BS model in relation to varying volatility measures. There are a vast variety of volatility models available for the investor through which he can predict the future volatility. “Which volatility model one should opt for pricing an option through the BS model?” is a big and sensitive question and accordingly the investor has to select one out of them. Present study is an attempt to answer this question to some extent, since it takes twelve different volatility measures and test how these predicted volatilities change the performance of the BS model.

**OBJECTIVES AND HYPOTHESIS**

The present study is aimed at fulfilling the following objectives:

- To identify and measure various volatility model’s performance in predicting the future volatility of Nifty.
- To test the pricing efficiency of the BS model with various alternative volatility measures.
- To test the hedging efficiency of the BS model with various alternative volatility measures.

In addition to the above mentioned three main objectives, we also aim to achieve the following secondary objectives:

- To review the antecedents of option pricing.
- To test the Nifty return series for some seasonalities.
- To test whether higher order Generalized Autoregressive Conditional Heteroscedasticity model (GARCH (p,q)) model is better than the GARCH (1,1) model in predicting the Nifty volatility.
- To test whether combination models perform better than the individual models, especially in the Indian context.
- To test whether BS model performance, irrespective of the volatility measure, is independent of moneyness\(^1\).

There are basically three primary hypotheses and some secondary hypothesis, involved in the present study which can be described as follows:

---

\(^1\) Moneyness is defined as the ratio of the underlying asset’s spot price to the underlying asset’s spot price to the exercise price under the option contract.
1. For testing the performance of the various volatility models, the null and the alternate hypothesis are:
   Null Hypothesis: All the twelve volatility models perform equally.
   Alternate Hypothesis: All the twelve volatility models do not perform equally.

2. For testing the pricing efficiency of the BS model, the null and the alternate hypothesis are:
   Null Hypothesis: Pricing efficiency of the BS model is independent of the volatility measure.
   Alternate Hypothesis: Pricing efficiency of the BS model is dependent upon the volatility measure.

3. For testing the hedging efficiency of the BS model, the null and the alternate hypothesis are:
   Null Hypothesis: Hedging efficiency of the BS model is independent of the volatility measure.
   Alternate Hypothesis: Hedging efficiency of the BS model is dependent upon the volatility measure.

The Null and Alternate hypothesis corresponding to some of the secondary objectives can be listed out as follows:

1. For testing the seasonalities in the Nifty log return series, particularly the Monday effects, the Day-of-the-month effects and the Month-of-the-year effects.
   - For testing Day-of-the-week (Monday) effects:
     Null Hypothesis: the day-of-the-week has deterministic influence on the Nifty log-return series for the periods of study.
     Alternate Hypothesis: the day-of-the-week has no deterministic influences on the Nifty log-return series for the periods of study.
   - For testing Day-of-the-month effects:
     Null Hypothesis: the day-of-the-month has deterministic influence on the Nifty log-return series for the periods of study.
     Alternate Hypothesis: the day-of-the-month has no deterministic influences on the Nifty log-return series for the periods of study.
   - For testing Month-of-the-year effects:
     Null Hypothesis: the Month-of-the-year has deterministic influence on the Nifty log-return series for the periods of study.
     Alternate Hypothesis: the Month-of-the-year has no deterministic influences on the Nifty log-return series for the periods of study.
2. For testing the performances of higher order GARCH(p,q) and GARCH (1,1):
   Null Hypothesis: GARCH (p,q) and GARCH (1,1) perform equally.
   Alternate Hypothesis: GARCH (p,q) and GARCH (1,1) do not perform equally.

3. For comparing the performance of combination models with individual models:
   Null Hypothesis: Combination models and individual models perform equally.
   Alternate Hypothesis: Combination models and individual models do not perform equally.

4. For testing whether the BS model’s performance, irrespective of the volatility input, is linked to the moneyness of the contract:
   Null Hypothesis: The BS model performance, irrespective of the volatility input, is independent of the moneyness of the option contract.
   Alternate Hypothesis: The BS model’s performance, irrespective of the volatility input, is dependent of the moneyness of the option contract.

DATA AND METHODOLOGY

Data

Two categories of dataset are required for the study. The first dataset is on Nifty closing prices which are required for calculating the forecasted volatilities through the twelve volatility models applied, and the second dataset is on closing prices of Nifty options which are required for applying the BS model. As such, the total period spans from 1 July, 2002 through 30 June, 2011, which is further divided into two parts: the first part spans from July 1, 2002 to 30 June 2008, i.e. a period of six years, which is used for calculating various parameters of the different volatility models and is called “in-sample” period; and the second part spans from 1 July 2008 to 30 June 2011, i.e. a period of three years, and is called the “out-of-sample” period for which the volatilities from the various models would be forecasted. The daily NIFTY closing prices are converted into daily log returns. Since the forecasted volatilities are to be inputted into the BS model, therefore, the data on Nifty options on which the BS model will be applied is taken from July 1, 2008 to June 30, 2011, which has to be the same period as the out-of-sample period taken for forecasting volatilities.

Methodology

Present study applies twelve different volatility models for predicting the future volatility of the Nifty returns. The random walk model (RW) is simple and just involves including the immediate past volatility as the forecast for the next period in the series. The historical mean (LTM) model involves taking an average of all the past observed volatilities available up to the time period t for forecasting the volatility in time period t+1. The moving average (MA) models (there are three MAs taken, that is n=15, 30 and 60 days) involve taking averages
over the ‘n’ number of past observations to forecast the next period’s volatility. The next model is the exponentially weighted moving average model (EWMA), where lambda is the only parameter involved which is taken to be constant, as per the Riskmetrics database, and equal to 0.94. In the simple regression (SR) model, we run a regression of actual volatilities on lagged values of the Nifty returns. It involves running the first Autoregression on the in-sample data, which is meant for estimating the parameters, and the estimates thus obtained are used for forecasting the volatility in the out-of-sample period. The parameters of the regression equation are updated every six months. Under the ARCH/GARCH category of volatility models, we first try to define the mean equation by analyzing the autocorrelations (ACF) and the partial autocorrelations (PACF) to define its lag structure. Then various calendar effects are tested to know whether they have any influence on the mean equation or not. Once, the mean equation is finalized, we then apply the basic ARCH model with the variance equation including up to 9 lags. From the comparison of these nine different ARCH models, the AIC (Akaike’s Information Criterion), BIC (Bayesian Information Criterion) and the log-likelihood criterion selected ARCH (1) model as the most efficient. We then apply ARCH (1) and check for any misspecification in it through the ACFs, PACFs, ARCH test, Ljung-box test, etc; and found it to be misspecified. Then we applied GARCH (p,q) and found through various criteria that GARCH (4,2) is the best amongst all the 81 GARCH models tested, that can be defined with \( p \in [1,2,\ldots,9] \) and \( q \in [1,2,\ldots,9] \). After applying GARCH (4,2), we test it for any misspecification, as was done for ARCH (1) and found that it cleared all the misspecification tests. Lastly, we apply the GARCH (1,1) model and through the various misspecification tests found it to be efficient in removing the ARCH effects and fitting the data nicely. After this, the two GARCH models are used to forecast the Nifty volatility for the out-of-sample period. The parameters of the two GARCH models are updated every six months, just like we did for the simple regression model.

Then we apply the implied volatility (IV) model, which is defined to be a market-based forecast of future volatility. It involves putting all the inputs required into the BS model, except for volatility, and equating it to the actual option prices prevailing in the market in order to extract the only left out variable, i.e., the standard deviation of the underlying asset. This extracted volatility is called the implied volatility.

The last category of volatility models is the combination models. We apply two combination models: first combination is called as “COMB1” and is a simple average of all the ten previously calculated volatility forecasts; and second combination is called as “COMB2”, which is again a simple average of all the previous forecasts except the implied volatility model.
All these volatility models are then evaluated through three categories of evaluation measures. First is the symmetrical loss measure in which we calculate the mean error (ME), the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the root mean squared errors (RMSE). The second category includes the asymmetrical loss errors, in which we calculate two measures, the mean mixed error which penalizes the under-predictions more heavily (MME(U)) and the mean mixed error which penalizes over-prediction more heavily (MME(O)). Lastly, we calculate a directional measure called the “correct directional measure (CDC)” in order to measure whether the models are able to predict the direction of change in the volatility or not. We give ranking to all the twelve models according to the above-mentioned three categories of evaluation measures and then in the end apply the Index of Rank Dominance (IRD) and the Relative Index of Rank Dominance (RIRD) in order to analyze, which model tends to dominate according to the various evaluation criteria.

In order to test the pricing performance of the BS model, we input the volatility forecasts from the twelve models, and accordingly, get the twelve different model prices each day. These model prices are subtracted from the market prices to get the absolute and percentage errors, which are then averaged according to “all-options” basis and moneyness basis. To test the hedging efficiency of the BS model, the model prices are used to locate the over-or-under-priced options on day \( t \). On the same day \( t \) we create a hedge, where overpriced/under-priced call options are sold/bought at the market price and \( n d_{t} \) (that is, the hedge ratio) number of index contracts are bought (sold) in the market, and the investment required for creating the hedge is calculated. The hedged position is maintained till the next day \( t+1 \) at which time it is closed out or liquidated at the \( t+1 \) prices and the excess returns from the hedged position are calculated. The hedged position is then re-established on day \( t+1 \) through creation of a new hedge. All the hedge returns are then averaged according to the same two bases as was applied in pricing errors.

RESULTS

The results regarding the respective hypothesis can be summarized briefly in the following paragraphs. Notably, the best model according to MAE is the GARCH (1,1) model whereas the worst is the IV model. According to the MAPE metric, the best is MA (30) and the worst is IV model, whereas according to RMSE metric, the best and the worst are the EWMA and IV model respectively. In sharp contrast to this the CDC criteria ranks the IV model as the best and RW as the worst model.

The MME (O) statistics, which penalizes over-predictions more heavily, prefers the RW model over other models, while giving the last rank to the IV model. The MME (U)
statistics prefers COMB1 model, while the worst model is the RW model. If one tries to combine all the results and see the overall performance of all the volatility models according to IRD and RIRD measures, the best model is the EWMA model and for the worst model there is a tie between the IV and SR models. In summary, the ranking of the different models significantly depends upon the choice of the error metrics and the null hypothesis that different volatility models tested performs equally is rejected. Moreover, it seems better not to combine the implied volatility information into the averages for getting the combination models as it reduces the performance, and thereby the ranking, of the combined model. Accordingly the secondary null hypothesis that the combination models and individual models perform equally is rejected.

The implied volatility graph depicted the shape of a “smirk” rather than a full “smile” which indicates that the implied volatility for ITMs is much higher than those for NTMs and OTMs in India. The results for the absolute pricing errors from the BS model indicate in the first instance, that IV model is the best input for the BS model to price the call options, though it was the worst performer according to the MAE, MAPE and RMSE statistical measures, as was seen previously. If the purpose of the investor is to price an option, then the implied volatility is the best choice as it consistently provides minimum errors irrespective of the moneyness of a call. There is no single model comes out to be an overall consistent winner according to the percentage errors. Secondly, it can be seen that ITMs are better priced than OTMs by the BS model irrespective of the volatility input. And lastly, barring a few exceptions, all volatility inputs through the various models in the BS formula leads to decreased percentage errors as moneyness increases. The null hypothesis that the BS model’s pricing efficiency doesn’t depend upon the volatility input, is rejected, though this rejection is based on the assumption made while performing the test, that is, the options markets are efficient. Moreover, the secondary null hypothesis is also rejected by the results.

The hedging results of the BS model indicates that the model’s performance for the short-term call options depend upon the category of calls hedged and the volatility input used to locate the over or under priced options. Overall, the model has a tendency to give better hedging performance for the OTMs and the NTMs as compared to the ITMs. The ability to register profits changes drastically if the correct volatility input is used to locate the over or under priced OTMs and NTMs. So, the null hypothesis that the BS model’s hedging performance is independent of the volatility intake, is rejected. Even the secondary null hypothesis that the BS model’s performance, irrespective of the volatility input, is independent of the moneyness of the option is also rejected. The best volatility model seems to be the simple regression model, which can enhance the BS model’s performance for all categories of options, some to major extent, whereas others to only some extent.
CONCLUSIONS

The celebrated Black and Scholes (BS) model has been a very effective tool for both the valuation and risk management of derivative assets. Mathematical option pricing models require estimation of volatility. There are a vast variety of volatility models available for the investor through which he can predict the future volatility. “Which volatility model he should opt for pricing an option through the BS model?” is a big and sensitive question and accordingly the investor has to select one out of them. This problem is faced by every investor in the derivatives market who aims for trading an option. Present study is an attempt to answer this question to some extent, since it takes twelve different volatility measures and test how these predicted volatilities change the performance of the BS model. In addition, it also tests the twelve volatility model’s performances independently in order to know which volatility model matches with the actual volatility and give minimum errors. In other words, our attempt was to incorporate various alternative volatility measures along with the set of statistical tests to verify the robustness of these alternative measures with a view to the following: (1) to identify the best volatility model for predicting future volatility and (2) to test the BS model for its predictive value through these alternative volatility measures.

Our results indicate that the ranking of the different volatility models significantly depends upon the choice of the error metrics. The market participants should take into consideration this sensitivity in ranking and should use that error statistics which serves their purpose the best. Considering the BS model’s pricing performance the results indicate that IV model is the best according to the absolute pricing errors, in pricing the call options. Regarding the BS model’s hedging performance, it can be concluded that the model’s performance for the short-term call options depends upon the category of calls hedged and the volatility input used to locate the over or under priced options. The best volatility model seems to be the simple regression model, which can enhance the BS model’s performance for all categories of options, some to major extents whereas others to only some extent. In summary, all the primary as well as the secondary hypotheses are rejected by the results. Which volatility model should be inputted into the BS model depends upon the purpose of the investor.