Intraday volatility forecasting from implied volatility

Article *in* International Journal of Managerial Finance · February 2011

DOI: 10.1108/17439131111109017 · Source: RePEc

Suk Joon Byun
Korea Advanced Institute of Science and Technology

Sol Kim
Hankuk University of Foreign Studies

Intraday Volatility Forecasting from Implied Volatility

Suk Joon Byun, Dong Woo Rhee*, and Sol Kim

Abstract

The implied volatility from the stochastic volatility model has shown to improve the forecasting performance of future volatility which has generally conducted by the implied volatility from Black and Scholes (1973) model when daily data is investigated. This study conducts the intraday volatility forecast analysis and compares the performance of the implied volatility from the stochastic model with that from BS implied volatility. Contrary to the results from previous studies using daily data, this study using minute-by-minute data shows that the performance of BS implied volatility is better than that of the stochastic volatility model. In addition, this study analyzes whether the forecasting performance is changed if a specific market environment is selected. Four specific markets are investigated; high volatile market, low volatile market, high return market, low return market. The forecasting performances are increased under high volatile market and low return market. On the other hand, the forecasting performances are decreased under low volatile market and high return market.

Keywords: Options, Stochastic Volatility, Intraday, Forecasting, Granger Causality, VAR

JEL classification: G13, G14

Suk Joon Byun is an assistant professor of the Business School at the Korea Advanced Institute of Science and Technology

Dong Woo Rhee is a doctorial candidate of the Business School at the Korea Advanced Institute of Science and Technology.

Sol Kim is an assistant professor of School of Business at Hankuk University of Foreign Studies.

* Corresponding author, Dong Woo Rhee, Korea Advanced Institute of Science & Technology (KAIST) Business School , 87 Hoegiro Dongdaemoon-gu Seoul 130-722 KOREA.

Phone: 82-2-958-3968, Fax: 82-2-958-3618, E-mail: dwrhee@business.kaist.ac.kr

1. Introduction

Though intraday volatility forecast has hardly been studied, it has interested not only practitioners but also academic people as option markets grow. So far, the research on the volatility forecast has been focused on the performance of a month ahead future realized volatility which is estimated by daily closing data. According to the previous studies, Black and Scholes (1973) (henceforth BS) implied volatility from option prices is considered the best estimate of future volatility. Latane and Rendleman(1976), which is the early study on the volatility forecast, showed the positive correlation between the implied volatility and the ex-post realized volatility when the cross-section data on the stock option is analyzed. Day and Lewis (1992), Lamoureux and Lastrapes (1993), Jorion (1995), and Fleming (1998) conducted time series analysis and commonly showed that BS implied volatility is efficient but biased as an estimator of the future realized volatility.

Some later studies question that BS implied volatility is the best estimate of the future volatility and try to find out a better estimator. Poteshman (2000) and Byun, Rhee, and Kim (2009) use Heston (1993)'s model, which is one of the stochastic volatility (henceforth SV) models, to derive the implied volatility from option prices. They commonly believe that the bias from the BS implied volatility can be reduced if the implied volatility from the SV model is introduced because of the following reason. Owing to the volatility smile phenomenon, BS model is considered not to explain the option market correctly. Bakshi, Cao, and Chen (1997, 2000) show that the SV model improves on the problems of BS model the most. Thus, they introduce the SV model and compare the forecasting performance of Heston(1993)'s implied volatility with that of BS implied volatility on the future volatility. As they expect, the forecasting performance of Heston(1993)'s implied volatility is shown to be superior to that of BS implied volatility when the performance of a month ahead future realized volatility is analyzed.

In the empirical analysis on the forecasting performance of future volatility, we fill the following gaps that are not resolved in previous researches. First, this is the first study which shows the forecasting performance of intraday future volatility. In this study, the forecasting performance is measured as the strength of the lead-lag relationship between an implied volatility and a realized volatility. In this analysis, it can be also proven whether an implied volatility influences a realized volatility more or vise versa. This study check whether the performance of Heston (1993)'s implied volatility is still superior to that of BS implied volatility when the performance of a hour ahead future realized volatility which is estimated by minute-by-minute data is analyzed. **In addition, it is checked whether the trading strategy by taking advantage of forecasting performance of an implied volatility on a future realized volatility can be effective for making money.** Second, this study checks the forecasting performance of future volatility on KOSPI200 index option market. Most stock option markets are inadequate for analyzing the intraday volatility forecasting because of insufficient trade volume. KOSPI200 index option market, which is the biggest stock index option market in the world in terms of the trade volume, might be a single candidate for this analysis. According to Table 1, no market can beat KOSPI200 index option market in terms of the number of contracts. Finally, this study analyzes whether the forecasting performance can be changed when a specific market environment is selected. **As Kim, Kim, and Nam (2009), Fuertes, Izzeldin, and Kalotychou (2008) show, a market condition has an influence on the forecasting performance of an option.** In this study, four specific market environments are introduced; High volatile market, Low volatile market, High return market, and Low return market. **If the forecasting performance is increased under a certain market environment, it is also checked that the profit from the trading strategy is increased under the same market environment.**

The main finding from this study is that the implied volatility from the SV model does not show any improvement over BS implied volatility for the intraday volatility forecasting even if both implied volatilities are proven to be informative on an hour ahead future realized volatility. This result is opposite to the results from the previous studies which show the superiority of Heston(1993)'s implied volatility. This study also shows that the forecasting performances are increased under high volatile market and low return market. **This result is consistent with the fact that the trading performance by utilizing the forecasting performance of an implied volatility is more effective for making a profit under the same market environment.**

The remainder of this study is organized in the following way. Section 2 exhibits the models which are used in this study. The process of deriving the Heston (1993)'s model is presented, followed by the method of estimating realized volatility. In addition, economic analysis models used in this study are introduced. Section 3 explains the KOSPI200 option market and describes the data used in this study. Section 4 shows the empirical results. First, the volatility estimation results and descriptive statistics of the estimated volatilities are presented. In addition, in-sample pricing errors and out-of-sample pricing errors are described. Next, the lead-lag relationships between an implied volatility and a realized volatility are shown. Additionally, the lead-lag relationships under high volatile and low volatile markets and high return and low return markets are shown and the difference on each market environment is described. Finally, the trading strategy by utilizing the forecasting power of implied volatilities on future realized volatility and its simulation result are shown. Section 5 concludes the results from this study.

2. Models

2.1 Stochastic Volatility Model

There are two types of SV models; continuous time SV model and discrete time SV model. Hull and White(1987), Johnson and Shanno(1987), Scott(1987), Wiggins(1987), Melino and Turnbull (1990), Stein and Stein(1991), and Heston (1993) assume the continuous time process, therefore, belong to the continuous time SV model. On the other hand, Duan(1995) and Heston and Nandi(2000) are considered the discrete time SV model. This study uses Heston (1993)'s model as a SV model because Bakshi, Cao, and Chen (1997, 2000) and Kim and Kim (2005) show the superiority of Heston(1993)'s model not only in terms of providing the closed form solution but also in terms of considering the correlation between the volatility and the return of the underlying asset.[1] Stochastic process of stock price and volatility which is assumed in Heston (1993)'s model is as below.

$$dS = \mu S dt + \sqrt{v_t} S dW_S \tag{1}$$

$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}\,dW_v \tag{2}$$

where, S is a stock value, μ is the return on the stock, W is a Wiener process, W_S and W_v have a correlation of ρ, v_t is an instantaneous variance at time t, κ is the speed parameter reverting to the long term average, θ, and σ is the volatility of volatility. Using the Fourier transform under the assumption of the stochastic process described above, the option pricing model follows below.

$$C = SP_1 - Ke^{-r\tau}P_2 \tag{3}$$

$$P_j = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-i\phi\ln[K]} f_j(x, v, \tau; \phi)}{i\phi}\right] d\phi \quad (j = 1, 2) \tag{4}$$

[1] The existing empirical studies showed the negative correlation between the volatility and the return of the underlying asset, risk neutral distribution with negative skewness and the low strike price which has large volatility, called volatility sneer. This is consistent with the leverage effect documented by Black (1976) and Christie (1982). The negative correlation phenomenon can be explained that falling stock prices will bring about relatively higher debt equity ratio, which in turn will have a leverage effect on the enterprise, which makes the volatility of the earnings per share greater, which eventually has the effect of amplifying the stock price volatility.

where, C is a call option, K is the strike price of the call option, r is risk-free interest rate, τ is time to maturity, Re[·]is the real number part of a complex number, i is the imaginary number, $\sqrt{-1}$, $f_j(x,\nu,\tau;\phi)=\exp(A(\tau;\phi)+B(\tau;\phi)\nu+i\phi x)$, x=ln(S), and A(·) and B(·) are functions of θ, κ, ρ, and σ.

Because the structural parameters are not observable, the parameters have to be estimated. Following Bakshi, Cao, and Chen (1997, 2000) and Bates (1991, 2000), we estimate the parameters every hour by minimizing the sum of squared percentage errors of the difference between the model price and the actual price. Compared to the estimation method using historical data, this method, which uses option prices, can take advantage of forward looking information contained in option prices. For Heston (1993)'s model, the parameters are estimated by minimizing the sum of squared percentage errors of the difference between the model price and the actual price in the following equation.

$$\min_{\sigma,\theta,\kappa,\rho,\nu_t} \sum_{i=1}^{N} \left[\frac{O_i^*(t,\tau;K)-O_i(t,\tau;K)}{O_i(t,\tau;K)} \right]^2 \quad (t=1,\cdots,T) \tag{5}$$

$O_i^*(t,\tau;K)$ is the model price of option i at time t, and $O_i(t,\tau;K)$ is the market price of option i, at time t. N is the number of options at time t, and T is the number of days in the sample.

2.2 Realized Volatility

Because this study has to estimate one hour realized volatility, high frequency data is inevitably needed. The shortest interval available in KOSPI200 index is a minute, therefore, minute-by-minute data is used for estimating the realized volatility. In addition, the use of the highest frequency data is supported by studies by Anderson, Bollerslew, Diebold, and Labys (2003) and Pong, Shackleton, Taylor, and Shu (2003). According to their studies, the realized volatility estimated by higher frequency data is better because it contains more data. Anderson, Bollerslev, Deiebold, and Ebens (2001) and Anderson, Bollerslev, and Diebold(2002) also mention that the measurement error from volatility estimates is decreased as the sampling frequency of the underlying is higher. Based on both practical and theoretical reasons above, the realized volatility is estimated as below.

$$Vol = \sqrt{\frac{1}{\Delta}\frac{1}{L-1}\sum_{i=1}^{L-1}\left[\ln\left(S_{i+1}/S_i\right)\right]^2} \tag{6}$$

where, Δ is the time interval between i and i+1 measured in years, L is the number of stock price data, S_i is the stock price at time i.

2.3 Econometric Analysis Models

To show the lead-lag relationship between an implied volatility and a realized volatility, granger causality test is carried out. The bivariate regression model is as below.

$$v_{i,t} = \alpha_0 + \sum_{n=1}^{6}\alpha_n v_{i,t-n} + \sum_{n=1}^{6}\beta_n v_{j,t-n} + \varepsilon_t \tag{7}$$
$$i, j = (1,2), (1,3), (2,1), (3,1)$$

where, $v_{1,t}$, $v_{2,t}$, and $v_{3,t}$ are realized volatility, BS implied volatility, and Heston (1993)'s implied volatility, respectively, corresponding to time t.

The granger causality test reports the F-statistics for the joint hypothesis, $\beta_1 = \cdots = \beta_6 = 0$.

The null hypothesis is that one volatility (v_j) does not lead the other (v_i).

A dummy variable is added to the granger causality test to check the lead-lag relationship between an implied volatility and a realized volatility when a specific environment is chosen. The bivariate regression model with a dummy variable is as below.

$$v_{i,t} = \alpha_0 + \sum_{n=1}^{6}\alpha_n v_{i,t-n} + \sum_{n=1}^{6}\beta_n v_{j,t-n} + \sum_{n=1}^{6}D\gamma_n v_{j,t-n} + \varepsilon_t \tag{8}$$
$$i, j = (1,2), (1,3), (2,1), (3,1)$$

where, $v_{1,t}$, $v_{2,t}$, and $v_{3,t}$ are realized volatility, BS implied volatility, and Heston (1993)'s implied volatility, respectively, corresponding to time t. D is the dummy variable which is set to 1 under the certain condition, and 0 elsewhere.

The granger causality test with a dummy variable reports the F-statistics for the joint hypothesis, $\beta_1 + \gamma_1 = \cdots = \beta_6 + \gamma_6 = 0$. The null hypothesis is that one volatility (v_j) does not lead the other (v_i) when the factor on the degree of the volatility is taken into account.

Four dummy variables are introduced; high volatility, low volatility, high return, low return. The dummy variables on high volatility and low volatility are those which are set to 1 for the times belonging to the most volatile 10% days and for the times belonging to the least volatile 10% days, respectively. On the other hand, the dummy variables on high return and low return are those which are set to 1 for the times belonging to the highest 10% days based on a daily return and for the times belonging to the lowest 10% days based on a daily return, respectively.

3. Data

KOSPI200 option market in Korea is analyzed for testing forecasting performance of the implied volatilities for intraday future volatility. Because of the intraday analysis, the trade volume is the most important criteria for selecting the market investigated. Table 1 shows that KOSPI200 index options market is the biggest in the world in terms of the trade volume. Therefore, KOSPI200 index options market is selected for this intraday volatility forecast analysis. The maturity date of KOSPI200 index options is the second Thursday of the option contract month, and option contract months are consecutive three months and one more month from March, June, September, and December. There are at least five exercise prices per each option contract month, which can be increased as option prices move. KOSPI200 index option contract is fully automated and is European option which can be exercised only at maturity. **In order to get a sufficient amount of data to estimate structural parameters every hour, the nearest expiration contract and the second nearest expiration contract options are used.** The sample period extends from July 1, 2004 through June 29, 2007. The minute-by-minute transaction prices for the KOSPI 200 index options are obtained from the Korea Stock Exchange. Both calls and puts that are near-the-money (henceforth NTM) and out-of-the money (henceforth OTM) are used for calculating both BS implied volatility and Heston (1993)'s implied volatility. In-the-money (henceforth ITM) options are excluded because their trading volume is very small, as a result, the reliability of the transacted price is not fully satisfied. Only the last reported transaction price prior to each hour of each option contract is employed in the empirical test. The following rules are applied to filter data needed for the empirical test. As options with less than six days or more than sixty days to expiration may induce biases due to low prices and bid-ask spreads, they are excluded from the sample. To mitigate the impact of price discreteness on option valuation, prices lower than 0.02 are not included. Prices not satisfying the arbitrage restriction are excluded. **After filtering, 50,279 calls and 72,838 puts are used in the empirical test. Because there are 4,473 hours in the sample period, the average number of options used to estimate parameters for each hour is**

approximately 27. **Table 2 shows the average option price and the number of options based on moneyness and the type of the option (call or put).** The 3-month CD rates are used as risk-free interest rates.[2] This study estimates an hour realized volatility through one minute frequency KOSPI200 index data which is obtained from Korea Stock Exchange.

4. Empirical results

4.1 Volatility estimation results

Table 3 reports the mean and the standard error of the parameter estimates for each model. The implicit parameters are not constrained to be constant over time. While re-estimating the parameters hourly is admittedly potentially inconsistent with the assumption of constant or slow-changing parameters used in deriving the option pricing model, such estimation is useful for indicating market sentiment. For Heston(1993)'s model, the implied correlation coefficient is negative as we expected. This is consistent with the leverage effect documented by Black (1976) and Christie (1982), whereby lower overall firm values increase the volatility of equity returns, and the volatility feedback effects of Porterba and Summers (1986) whereby higher volatility assessments lead to heavier discounting of future expected dividends and thereby lower equity price. We evaluate the in-sample performance of each model by comparing market prices with model prices computed by using the parameter estimates from the current time. Table 4 reports the in-sample valuation errors for the BS and Heston(1993)'s models computed over the whole sample of options as well as across six moneyness and two option type categories. First, with respect to all measure, Heston(1993)'s model shows better performance than the BS model. This is rather an obvious result when the use of larger number of parameters in Heston(1993)'s model is considered. Second, all models show moneyness-based valuation errors. The models exhibits the worst fit for the near-the-money options. The fit, as measured by MAE, steadily improve as we move from near-the-money to out-of-the-money options. Overall, the SV model performs better for in-sample pricing. However, in-sample pricing performance can be biased due to the potential problem of overfitting to the data. **In the out-of-sample pricing, it can be checked whether the extra parameters do not improve the structural fitting, as a result, do cause overfitting. Table 5 presents the one hour ahead out-of-sample pricing results. Both MAE and MSE results support that Heston (1993)'model performs better than BS model for out-of-sample pricing. This result implies that more parameters Heston(1993)'s model has play positive roles in the structural fitting.**

[2] Korea does not have a liquid Treasury bill market, so the 3-month CD rates are used in spite of the mismatch of maturity of options and interest rates.

One interesting phenomenon observed in table 5 is that the fitting error is the biggest around near-the-money options as MAE measure, whereas that is the smallest as MSE measure. It can be explained that as moneyness moves to either in-the-money or out-of-the-money from at-the-money, the distribution of the fitting error looks more fat-tailed.

Table 6 shows the descriptive statistics for 4,473 hourly estimated realized volatilities and two implied volatilities on KOSPI200 option prices from 10:00 July 1 2004 through 15:00 June 30 2007. According to the results in Table 6, the estimated values on a realized volatility are averagely lower than those on implied volatilities. On the other hand, the standard deviation of the estimated realized volatilities is much larger than those of the estimated implied volatilities. ADF test statistics show that no volatility time series has a unit root.

4.2 Lead-Lag relation

Panel A of Table 7 is the granger causality test results between an implied volatility and a realized volatility for the total period, from 1 July 2004 to 30 June 2007. An implied volatility is derived from either BS model or Heston(1993)'s model. Though this results indicate that a realized volatility grander causes any implied volatility, it is found that the implied volatilities grander cause a realized volatility stronger. This result means that an implied volatility influences a realized volatility more than vise versa. Between the two implied volatilities, BS implied volatility lead a realized volatility more significantly than Heston (1993)'s implied volatility. **This result is opposite to those appeared in both in-sample pricing and out-of-sample pricing analyses. It can be inferred from the following reason. In pricing performance, five parameters in Heston(1993) model function simultaneously. As a result, the pricing performance is attributed not only from a volatility parameter but also from other four parameters. On the contrary, the grander causality result is only influenced by the volatility parameter. This is why the forecasting performance of Heston(1993)'s implied volatility can be inferior to that of BS implied volatility even if the pricing performance of the former is superior to that of the latter.** This result is also different from those of Poteshman (2000) and Byun, Rhee, and Kim (2009) which analyze the 30 days forecasting performance of both BS implied volatility and Heston (1993)'s implied volatility on a realized volatility. In their findings, Heston (1993)'s implied volatility is commonly superior to BS implied volatility for the forecasting performance of future volatility. The results shown in Table 7, on the contrary, indicates that the empirical results using intraday data does not support the superiority of Heston (1993)'s implied volatility for the forecasting performance of future volatility. Panel B of Table 7 shows the sub periods results of the granger causality test results. There are 6 sub periods, each of which is composed of six months. In these results, implied volatilities do not lead a realized volatility if any specific time period is selected. Thus, it can be mentioned that implied volatilities do not always give information on a realized volatility.

4.3 Lead-Lag relation under high volatile market and low volatile market

Table 8 shows the results of the granger causality test when the factor on the degree of the volatility is taken into account. To catch this effect, two dummy variables are introduced. One dummy variable which recognizes a high volatile market is set to 1 for the times belonging to the most volatile 10% days. The other which recognizes a low volatile market is set to 1 for the times belonging to the least volatile 10% days. Panel A of Table 8 is the granger causality test results when the dummy variable on a high volatile environment is added to the original ganger causality test which is presented in Table 7. Compared with the results in Table 7, the values on F-statistics are largely increased. In particular, it is shown that both BS implied volatility and Heston (1993)'s implied volatility lead a realized volatility more strongly. This fact can be understood as follows. The positive relationship between the trade volume and the volatility is empirically proven in previous studies such as Hiemstra and Jones (1994), Bessembinder and Seguin (1993), and Karpoff (1987). The fact that the high price efficiency can be achieved from the high trade volume is intuitively accepted. Thus, the implied volatilities which are forward-looking volatilities in a market can lead a realized volatility better in a highly volatile environment. On the contrary, Panel B of Table 8 is the granger causality test results when the dummy variable on a low volatile environment is added to the original granger causality test. As it is expected from the results of Panel A, the values on F-statistics are largely decreased compared with the results presented in Table 7. Panel B indicates that implied volatilities do not lead a realized volatility statistically in a low volatile market.

4.4 Lead-Lag relation under high return market and low return market

Table 9 shows the results of the granger causality test when the factor on the degree of the return is taken into account. In order to induce these results, two dummy variables are used as well. One dummy variable which recognizes high return market is set to 1 for the times belonging to the highest 10% days based on a daily return. The other which recognizes low return market is set to 1 for the times belonging to the lowest 10% days based on a daily return. Panel A of Table 9 is the granger causality test results when the dummy variable on high return market is added to the original ganger causality test which is presented in Table 7. Even if the values on F-statistics are largely statistically significant, those are slightly decreased compared with the results in Table 7. On the contrary, the granger causality results which are shown at Panel B indicate that the values on F-statistics are increased when the dummy variable on low return market is applied. Upon the results shown in Table 8, it can be mentioned that implied volatilities lead a realized volatility more strongly in a bad performing period. This result is consistent with that presented in Table 8 in that the negatively skewed feature observed in stock markets implies that a bad performing period is generally volatile. Though the degree of the lead-lag relationship between an implied volatility and a realized volatility depends on the return of a stock market, it is less sensitive to a return rather than volatility.

4.5 Trading Simulation

The following trading strategy is set to utilize the forecasting performance of implied volatilities on one hour ahead future volatility. First, an estimated future realized volatility is derived from either lagged BS implied volatilities or lagged Heston(1993)'s implied volatilities. If the estimated future realized volatility is greater (smaller) than BS implied volatility which is derived from the option price at an observation hour, the option is considered underpriced (overpriced). Long (Short) strategy, which longs (shorts) one unit of call option and shorts (longs) delta unit of index, and unwinds the positions in an hour, is executed when the option is the most underpriced (overpriced) 10% hours among the sample selected. Among several options traded in the same hour, NTM option is selected for this trading strategy. The delta, which is needed to make delta-neutral strategies above, is calculated as below.

$$\Delta_t = \frac{\partial C_t}{\partial S_t} = N(d_1) \tag{9}$$

where,

$$d_1 = \frac{\ln(S_t / K_t) + (r_t + 0.5 \times \sigma_t^2) \times \tau}{\sigma_t \sqrt{\tau}},$$

C_t is the call option price at time t, S_t is KOSPI200 index at time t, K_t is the near-the-money strike price at time t, r is the risk free rate at time t, σ_t is BS implied volatility at time t, and N(·) is the cumulative normal density function. The profit or loss from the trading strategy at time t is obtained as below.

$$Long: \quad \pi_t = [(C_{t+1} - C_t) + \Delta_t (S_t - S_{t+1})] \tag{10}$$

$$Short: \quad \pi_t = [(C_t - C_{t+1}) + \Delta_t (S_{t+1} - S_t)] \tag{11}$$

Table 10 reports the trading simulation results. When full period is selected, Long strategy earns a positive profit regardless of the way of estimating future realized volatility. Short

strategy, on the contrary, loses money. These results can be inferred from the following reason. A person who has long (short) option and short (long) delta neutral underlying gets a positive profit when a convexity oriented gain from the movement of underlying is bigger (smaller) than the decrease of option value from time decaying effect. However, in this trading strategy whose duration is only an hour, time decaying effect is too small to impact on the trading performance. Therefore, Long strategy shows averagely positive performance, whereas Short strategy shows averagely negative performance. This trading strategy is also applied to four specific market environments; high volatile market, low volatile market, high return market, and low return market. High (Low) volatile market sample covers the hours belonging to the most (least) volatile 10% days, and high (low) return market sample covers the hours belonging to the highest (lowest) 10% days based on a daily return. The trading performance from high volatile market is better than that from low volatile market and the trading performance from low return market is better than that from high return market. This result is consistent with those observed at lead-lag relationship analysis. It can be inferred that lagged implied volatilities have stronger influence on a future realized volatility under high volatile or low return market, as a result, the trading strategy executed under that kind of market environment performs better. The way of deriving estimated future realized volatility does not impact significantly on the trading performance. However, because z statistics results presented at Table 10 are low, it is risky to follow this trading strategy. Even, these trading simulation results do not consider transaction costs.

5. Conclusion

There have been no previous researches which tried to find out the best model for the forecasting performance of intraday future volatility. As a volatility market grows, it is getting more and more important to check the forecasting performance of intraday future volatility not only for the academic prospective but also for the practical aspect. Since KOSPI200 index options market is the most liquid equity option market in the world in terms of trading volume, this market is examined for the intraday volatility forecasting performance. Among various models, BS and Heston(1993) models are selected because those are previously compared and proven to be superior to others in terms of the forecasting performance of future volatility.

The lead-lag relation results show that both BS and Heston(1993)'s implied volatilities lead a realized volatility. Unlike previous studies, such as Poteshman(2000) and Byun, Rhee, and Kim (2009), which analyzed the forecasting performance of one month ahead future realized volatility, Heston(1993)'s implied volatility is not proven to be superior to BS implied volatility in terms of the forecasting performance. In this aspect, to make the assumptions flexible in

Heston (1993)'s model, such as non-zero market price of volatility risk and non-zero correlation between innovations to the level and volatility of the underlying asset, is not the meaningful effort to improve the empirical deficiency which BS model has when intraday volatility is forecasted.

The lead-lag relationship between an implied volatility and a realized volatility is changed when a specific market environment is selected. On a high volatile market, the lead-lag relationship is stronger. This result implies that implied volatilities are more meaningful for the future realized volatility. On the contrary, the lead-lag relationship is weaker under low volatile market. On the other hand, the return on KOSPI200 index is another important factor for the strength of the lead-lag relationship. Whereas the lead-lag relationship is weaker under the high return environment, that is stronger under the low return environment. These results appear to make sense in that the negatively skewed feature observed in stock markets implies that a bad performing period is generally volatile.

Trading simulation results support the lead-lag relation results between an implied volatility and a realized volatility. When the trading strategy which utilizes the forecasting performance of an implied volatility on a future realized volatility is applied, positive performance can be achieved. In addition, the trading strategy seems to be more effective when it is applied under high volatile market and low return market. However, as z statistics of the trading performance imply, it does not look safe to follow the trading strategy.

References

Andersen TG, Bollerslev T, Diebold FX (2002) Parametric and Nonparametric Volatility Measurement. Working paper, prepared for Handbook of Financial Econometrics

Andersen TG, Bollerslev T, Diebold FX, Ebens H (2001) The Distribution of Stock Return Volatility. J Financial Econometrics 61:43-76

Andersen TG, Bollerslev T, Diebold FX, Labys P (2003) Modeling and Forecasting Realized Volatility. Working Paper, National Bureau of Economic Research

Bakshi G, Cao C, Chen Z (1997) Empirical performance of alternative option pricing models. J Finance 52:2003-2049

Bakshi G, Cao C, Chen Z (2000) Pricing and hedging long-term options. J Econometrics 94:277-318

Bates D (1991) The Crash of ’87: Was It Expected? The Evidence from Options Market. J Finance 46:1009-1044

Bates D (2000) Post-’87 Crash Fears in the S&P 500 Futures Option Market. J Econometrics 94:181-238

Bessembinder H, Segin PJ (1993) Price volatility, trading volume, and market depth: evidence from the futures markets. Journal of Financial and Quantitative Analysis 28:21-39

Black F (1976) Studies of Stock Price Volatility Changes. Proceedings of the 1976 Meeting of the American Statistical Association, Business and Economic Statistic Section:177-181

Black F, Scholes M (1973) The pricing of options and corporate liabilities. J Political Economy 81:637-659

Byun SJ, Rhee DW, Kim S (2009) Forecasting Future Volatility from Option Prices Under the Stochastic Volatility Model. Working paper, KAIST

Christie A (1982) The Stochastic Behavior of Common for Alternative Stochastic Processes. J Financial Economics 10:407-432

Day T, Lewis C (1992) The information content of option prices and a test of market efficiency. J Econometrics 52:267-287

Duan JC (1995) The GARCH Option Pricing Model. Math Finance 5:13-32

Figlewski S (1997) Forecasting volatility. Financial Market Institutions & Instruments 6:1-88

Fleming J (1998) The quality of market volatility forecasts implied by S&P 100 index option prices. J Empir Finance 5:317-345

Fuertes AM, Izzeldin M, Kalotychou E (2008) On Forecasting Daily Stock Volatility: the Role of Intraday Information and Market Conditions. Working paper, City University London

Heston SL (1993) A Closed-Form Solutions for Options with Stochastic Volatility with Applications to Bond and Currency Options. Rev Financial Studies 6:327-343

Heston SL, Nandi S (2000) A Closed-Form GARCH Option Valuation Model. Rev of Financial Studies 13:585-625

Hiemstra C, Jones JD (1994) Testing for linear and nonlinear Granger causality in the stock price-volume relation. Journal of Finance 49:1639-1664

Hull J, White A (1987) The Pricing of Options with Stochastic Volatilities. J Finance 42:281-300

Johnson H, Shanno D (1987) Option Pricing when the Variance is Changing. J Financial and Quantitative Analysis 22:143-151

Jorion P (1995) Predicting volatility in the foreign exchange market. J Finance 50:507-528

Karpoff JM (1987) The relation between price changes and trading volume: a survey. Journal of Financial and Quantitative Analysis 22:109-126

Kim IJ, Kim S (2005) Empirical Comparison of Alternative Stochastic Volatility Option Pricing Models: Evidence from Korean KOSPI 200 Index Options Market. Pac Basin Finance J 12:117-142

Kim S, Kim IJ, Nam SO (2009) The lead-lag relationship between stock index options and the stock index market: Model, moneyness, and news. International Journal of Managerial Finance 5:311-332

Lamoureux D, Lastrapes W (1993) Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. Rev Financial Studies 6:293-326

Latane HA, Rendleman RJ (1976) Standard deviations of stock price ratios implied in option prices. J Finance 31:369-381

Melino A, Turnbull SM (1990) Pricing Foreign Currency Options with Stochastic Volatility. J Econometrics 45:239-265

Pong S, Shackleton M, Taylor S, Xu X (2003) Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models. Working paper

Poterba J, Summers L (1986) The persistence of volatility and stock market fluctuations. American

Economic Review 76:1142-1151

Poteshman AM (2000) Forecasting Future Volatility from Option Prices. Working paper, University of Illinois at Urbana-Champaign

Scott LO (1987) Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application. J Financial and Quantitative Analysis 22:419-438

Stein EM, Stein JC (1991) Stock Price Distribution with Stochastic Volatility: An Analytic Approach. Rev Financial Studies 4:727-752

Wiggins JB (1987) Option Values under Stochastic Volatility: Theory and Empirical Estimates. J Financial Economics 19:351-377

Table 1: The World's Top 10 Derivative Contracts

This table shows the ten most active derivative contracts, measured in millions of contracts from the year 2001 to the year 2007. The rank is determined based on the trading volume of the year 2007.

Rank	Contract	2001	2002	2003	2004	2005	2006	2007
1	KOSPI 200 Options, Korea Exchange	823	1,890	2,838	2,522	2,535	2,414	2,642
2	Eurodollar Futures, CME	184	202	209	298	410	502	621
3	E-mini S&P500 Futures, CME	39	116	161	167	207	258	415
4	10y T-Note Futures, CME	58	96	147	196	215	256	349
5	Euro-Bund Futures, Eurex	178	191	244	240	299	320	338
6	DJ Euro Stoxx 50 Futures, Eurex	38	86	116	122	140	214	327
7	Eurodollar Options on Futures, CME	88	106	101	131	188	269	313
8	DJ Euro Stoxx 50 Options, Eurex	19	40	62	71	91	150	251
9	1d Inter-Bank Deposit Futures, BM&F	46	49	58	100	121	161	221
10	3m Euribor Futures, Liffe	91	106	138	158	167	202	221

Source: Futures Industry Association (http://www.futuresindustry.org)

Table 2: KOSPI 200 Options Data

This table reports the average option price and the number of options based on moneyness and the type of the option (call or put). The sample period is from 9:00 1 July 2004 to 15:00 30 June 2007. The minute-by-minute information from the last traded prices prior to every hour of each option contract is used to get the summery statistics. S and K denote the spot price and the exercise price of KOSPI200, respectively.

Call Options			Put Options		
Moneyness	Price	Number	Moneyness	Price	Number
S/K<0.94	0.3514	18421	1.00<S/K<1.03	2.6527	14404
0.94<S/K<0.96	1.0682	16360	1.03<S/K<1.06	1.4604	13706
0.96<S/K<1.00	2.5478	15498	S/K>1.06	0.4282	44728
Total	1.2616	50279	Total	1.0623	72838

Table 3: Parameters

The table reports the mean and the standard error of the parameters which are hourly estimated for the Black-Scholes (1973) option pricing model (BS) and Heston's (1993) option pricing model (SV). For the BS and SV models, each parameter is estimated by minimizing the sum of squared errors between model and market option prices every hour.

	σ				
BS	0.2020				
	(0.0005)				
	κ	θ	σ_v	ρ	v_t
SV	15.5809	0.5708	1.1575	-0.5634	0.0478
	(1.7198)	(0.0450)	(0.0127)	(0.0043)	(0.0004)

Table 4: In-Sample Pricing Errors

This table reports in-sample pricing errors for the KOSPI 200 option with respect to moneyness. S/K is defined as moneyness where S and K denote the spot price and the strike price, respectively. Each model is estimated every hour during the sample period and in-sample pricing errors are computed by the difference between the model price using estimated parameters and the actual prices which are used for estimating the parameters. MAE and MSE denote mean absolute errors and mean squared errors, respectively and they are calculated as below.

$$MAE = \sum_{i=1}^{N} |\varepsilon_i| / N$$

$$MSE = \sum_{i=1}^{N} (\varepsilon_i)^2 / N$$

Where, ε_i is the difference between the model price and the actual price for each sample and N is the number of the sample. BS is the Black-Scholes (1973) option pricing model and SV is Heston's (1993) option pricing model.

	Moneyness	BS	SV
MAE	S/K<0.94	0.2060	0.0353
	0.94<S/K<0.96	0.3081	0.0526
	0.96<S/K<1.00	0.3144	0.1680
	1.00<S/K<1.03	0.3462	0.2015
	1.03<S/K<1.06	0.3735	0.0733
	S/K>1.06	0.2183	0.0554
	Total	0.2727	0.0853
MSE	S/K<0.94	0.9251	0.1006
	0.94<S/K<0.96	0.3360	0.0391
	0.96<S/K<1.00	0.0671	0.0197
	1.00<S/K<1.03	0.0301	0.0096
	1.03<S/K<1.06	0.1163	0.0122
	S/K>1.06	0.5551	0.1800
	Total	0.4096	0.0906

Table 5: One hour out-of-Sample Pricing Errors

This table reports one hour out-of-sample pricing errors for the KOSPI 200 option with respect to moneyness. S/K is defined as moneyness where S and K denote the spot price and the strike price, respectively. Each model is estimated every hour during the sample period and one hour out-of-sample pricing errors are computed by the difference between the model price using the parameters estimated by the options which are traded one hour before and the actual prices. MAE and MSE denote mean absolute errors and mean squared errors, respectively and they are calculated as below.

$$MAE = \sum_{i=1}^{N} \left| \varepsilon_i \right| / N$$

$$MSE = \sum_{i=1}^{N} \left(\varepsilon_i \right)^2 / N$$

Where, ε_i is the difference between the model price and the actual price for each sample and N is the number of the sample. BS is the Black-Scholes (1973) option pricing model and SV is Heston's (1993) option pricing model.

	Moneyness	BS	SV
MAE	S/K<0.94	0.2062	0.0431
	0.94<S/K<0.96	0.3098	0.0755
	0.96<S/K<1.00	0.3212	0.1910
	1.00<S/K<1.03	0.3519	0.2217
	1.03<S/K<1.06	0.3740	0.0970
	S/K>1.06	0.2182	0.0626
	Total	0.2745	0.1000
MSE	S/K<0.94	0.9546	0.1111
	0.94<S/K<0.96	0.3469	0.0464
	0.96<S/K<1.00	0.0691	0.0229
	1.00<S/K<1.03	0.0311	0.0120
	1.03<S/K<1.06	0.1171	0.0172
	S/K>1.06	0.5546	0.1921
	Total	0.4158	0.0988

Table 6: Descriptive Statistics

Descriptive Statistics for 4,473 hourly estimated realized volatilities and two implied volatilities from option prices using BS and Heston (1993) Models on KOSPI200 option for the period from 10:00 July 1 2004 through 15:00 June 30 2007 are presented. Realized volatilities are annualized through multiplying an hour basis volatility by $\sqrt{6 \times 250}$. There are 4 principals to filter data for empirical analysis. First, the latest transacted option data prior to every hour observed are selected. Second, if there are more than one transaction data for each hour, only one data is used. Third, the option whose price is below 0.02 is excluded. Fourth, the option which doesn't meet the arbitrage restriction is not included. Realized volatilities are estimated by 60 minute-by-minute data beginning from the every targeted hour. Mean, Media, Max, Min, Standard deviation, Skewness, Kurtosis, and Augmented Dickey-Fuller (ADF) test statistics for 4,473 hourly estimated realized volatility, BS implied volatility (BS), Heston (1993)'s implied volatility (SV) are presented.

Statistics	RV	BS	SV
Observations	4473	4473	4473
Mean	0.148916	0.201962	0.212653
Median	0.121031	0.196401	0.20837
Max	1.632191	0.331017	0.6949
Min	0.061580	0.130101	0.000112
Std. Dev.	0.094423	0.035200	0.051026
Skewness	4.205564	0.677648	0.546646
Kurtosis	33.42331	3.230714	5.407161
ADF Test Statistic	-41.4899	-4.31224	-7.68093

Table 7: Granger Causality Test Results

This table reports the granger causality test results between an implied volatility and a realized volatility. An implied volatility is derived from option prices using either BS or Heston (1993) Model. The bivariate regression model is as below.

$$v_{i,t} = \alpha_0 + \sum_{n=1}^{6} \alpha_n v_{i,t-n} + \sum_{n=1}^{6} \beta_n v_{j,t-n} + \varepsilon_t$$
$$i, j = 1,2,3 \; (i \neq j)$$

$v_{1,t}$, $v_{2,t}$, and $v_{3,t}$ are realized volatility (RV), BS implied volatility (BS), and Heston (1993) implied volatility (SV), respectively, corresponding to time t.
The reported F-statistics are the Wald statistics for the joint hypothesis, $\beta_1 = \cdots = \beta_6 = 0$. The null hypothesis is that one volatility (v_j) does not lead the other (v_i).
*** , **, and * are statistically significant at 1%, 5%, and 10% levels.

Panel A: Total Period

	BS leads RV	SV leads RV	RV leads BS	RV leads SV
F-stat	21.9893***	14.8804***	7.8982***	7.2747***

Panel B: Sub Periods

	BS leads RV	SV leads RV	RV leads BS	RV leads SV
2004.7.1 ~ 2004.12.31				
F-stat	2.6449**	1.4574	1.0481	1.8109*
2005.1.1 ~ 2005.6.30				
F-stat	1.6386	0.9603	2.3815**	4.3107***
2005.7.1 ~ 2005.12.31				
F-stat	2.1695**	2.8434***	3.6525***	3.8665***
2006.1.1 ~ 2006.6.30				
F-stat	5.9440***	6.3337***	0.8640	1.2799
2006.7.1 ~ 2006.12.31				
F-stat	11.5946***	9.0718***	0.2798	0.4986
2007.1.1 ~ 2007.6.30				
F-stat	1.1129	0.1481	9.6821***	1.7803

Table 8: Granger Causality Tests Results with Dummy Variables on Volatility

This table reports the granger causality test results between an implied volatility and a realized volatility with dummy variables whose values are determined by the degree of the volatility on KOSPI200 in the date those are applied to. An implied volatility is derived from option prices using either BS or Heston (1993) Model. The bivariate regression model with dummy variables is as below.

$$v_{i,t} = \alpha_0 + \sum_{n=1}^{6} \alpha_n v_{i,t-n} + \sum_{n=1}^{6} \beta_n v_{j,t-n} + \sum_{n=1}^{6} D_{HV} \gamma_n v_{j,t-n} + \varepsilon_t$$

$$v_{i,t} = \alpha_0 + \sum_{n=1}^{6} \alpha_n v_{i,t-n} + \sum_{n=1}^{6} \beta_n v_{j,t-n} + \sum_{n=1}^{6} D_{LV} \gamma_n v_{j,t-n} + \varepsilon_t$$

$$i, j = (1,2), (1,3), (2,1), (3,1)$$

$v_{1,t}$, $v_{2,t}$, and $v_{3,t}$ are realized volatility (RV), BS implied volatility (BS), and Heston(1993)'s implied volatility (SV), respectively, corresponding to time t. D_{HV} and D_{LV} are dummy variables which are set to 1 for the times belonging to the most volatile 10% days and for the times belonging to the least volatile 10% days, respectively.

The reported F-statistics are the Wald statistics for the joint hypothesis, $\beta_1 + \gamma_1 = \cdots = \beta_6 + \gamma_6 = 0$. The null hypothesis is that one volatility (v_j) does not lead the other (v_i) when the factor on the degree of the volatility is taken into account.

*** , **, and * are statistically significant at 1%, 5%, and 10% levels.

Panel A: High Volatility

	BS leads RV	SV leads RV	RV leads BS	RV leads SV
F-stat	35.1451***	55.8664***	5.8002***	4.4478***

Panel B: Low Volatility

	BS leads RV	SV leads RV	RV leads BS	RV leads SV
F-stat	1.5159	0.0865	0.1820	0.6182

Table 9: Granger Causality Tests Results with Dummy Variables on Return

This table reports the granger causality test results between two volatilities selected from a realized volatility and two implied volatilities from option prices using BS and Heston (1993) Models with dummy variables whose values are determined by the degree of the return on KOSPI200 in the date those are applied to. The bivariate regression model with dummy variables is as below.

$$v_{i,t} = \alpha_0 + \sum_{n=1}^{6} \alpha_n v_{i,t-n} + \sum_{n=1}^{6} \beta_n v_{j,t-n} + \sum_{n=1}^{6} D_{HR} \gamma_n v_{j,t-n} + \varepsilon_t$$

$$v_{i,t} = \alpha_0 + \sum_{n=1}^{6} \alpha_n v_{i,t-n} + \sum_{n=1}^{6} \beta_n v_{j,t-n} + \sum_{n=1}^{6} D_{LR} \gamma_n v_{j,t-n} + \varepsilon_t$$

$$i, j = (1,2), (1,3), (2,1), (3,1)$$

$v_{1,t}$, $v_{2,t}$, and $v_{3,t}$ are realized volatility (RV), BS implied volatility (BS), and Heston(1993)'s implied volatility (SV), respectively, corresponding to time t. D_{HR} and D_{LR} are dummy variables which are set to 1 for the times belonging to the highest 10% days based on a daily return and for the times belonging to the lowest 10% days based on a daily return, respectively.
The reported F-statistics are the Wald statistics for the joint hypothesis, $\beta_1 + \gamma_1 = \cdots = \beta_6 + \gamma_6 = 0$. The null hypothesis is that one volatility (v_j) does not lead the other (v_i) when the factor on the degree of the return is taken into account.
*** , **, and * are statistically significant at 1%, 5%, and 10% levels.

Panel A: High Return

	BS leads RV	SV leads RV	RV leads BS	RV leads SV
F-stat	16.2800***	7.8994***	3.3811***	0.3500

Panel B: Low Return

	BS leads RV	SV leads RV	RV leads BS	RV leads SV
F-stat	23.6066***	16.0076***	31.3276***	10.3163***

Table 10: Trading Simulation Results

This table reports the trading simulation results by utilizing the relative value between an implied volatility and estimated one hour ahead future volatility. Long (Short) strategy, which longs (shorts) one unit of call option and shorts (longs) delta unit of index, and unwinds the positions in an hour, is executed when an implied volatility is the most underpriced (overpriced) 10% hours among the sample selected. The implied volatility in the strategies is BS implied volatility at the time when a strategy is executed. One hour ahead future volatility is estimated by two ways; One is estimated by six lagged BS implied volatilities, and the other is estimated by six lagged Heston(1993)'s implied volatilities. The coefficients on lagged variables are calculated by ordinary least squared (OLS) method. The delta, which is needed to make delta-neutral strategies above, is calculated as below.

$$\Delta_t = \frac{\partial C_t}{\partial S_t} = N(d_1)$$

where,

$$d_1 = \frac{\ln(S_t / K) + (r + 0.5 \times \sigma^2) \times \tau}{\sigma\sqrt{\tau}},$$

C_t is the call option price at time t, S_t is KOSPI200 index at time t, K_t is the near-the-money strike price at time t, r is the risk free rate at time t, σ_t is BS implied volatility at time t, and $N(\cdot)$ is the cumulative normal density function. Among several option prices corresponding to strike prices traded, only NTM option is selected for the trading simulation. High (Low) volatile period sample covers the times belonging to the most (least) volatile 10% days, and high (low) return period sample covers the times belonging to the highest (lowest) 10% days based on a daily return. The profit or loss from the trading strategy at time t is obtained as below.

$$Long: \quad \pi_t = [(C_{t+1} - C_t) + \Delta_t(S_t - S_{t+1})]$$
$$Short: \quad \pi_t = [(C_{t+1} - C_t) + \Delta_t(S_t - S_{t+1})]$$

Average, max, min, standard deviation, and z statistics of the trading performance are reported. The performances of Short strategy are in parentheses.

Panel A: Estimated realized volatility is derived by lagged BS implied volatilities

Sample Period	Sample Number	Average	Max	Min	S.D.	Z statistics
Full	447(447)	0.01(-0.01)	2.37(0.55)	-0.37(-2.64)	0.17(0.22)	0.06(-0.04)
High Volatile	45(45)	0.05(-0.01)	1.27(0.28)	0.27(-0.28)	0.24(0.12)	0.21(-0.09)
Low Volatile	45(45)	-0.01(-0.04)	0.20(0.15)	-0.34(-1.82)	0.10(0.29)	-0.10(-0.13)
High Return	45(45)	0.08(0.00)	2.37(0.21)	-0.37(-0.39)	0.40(0.11)	0.20(0.04)
Low Return	45(45)	0.10(-0.05)	1.72(0.28)	-0.21(-2.07)	0.32(0.34)	0.32(-0.13)

Panel B: Estimated realized volatility is derived by lagged Heston(1993)'s implied volatilities

Sample Period	Sample Number	Average	Max	Min	S.D.	Z statistics
Full	447(447)	0.02(-0.00)	2.46(0.34)	-0.57(-2.08)	0.26(0.14)	0.09(-0.00)
High Volatile	45(45)	0.05(-0.00)	0.61(0.14)	-0.27(-0.16)	0.17(0.07)	0.31(-0.03)
Low Volatile	45(45)	0.04(-0.04)	2.46(0.19)	-0.43(-1.82)	0.39(0.33)	0.10(-0.13)
High Return	45(45)	0.05(-0.00)	2.37(0.20)	-0.57(-0.26)	0.41(0.08)	0.13(-0.03)
Low Return	45(45)	0.11(-0.00)	2.64(0.28)	-0.61(-0.19)	0.52(0.08)	0.22(-0.04)