An Analysis of Entropy of Volatility of Stock Return

Abstract

We attempt to find out the relative impact of volatility measure computed from four parameters, namely, closing, high and low quotes of the day as well as combination of these three, on the stock index return. In addition various time steps are also selected for such computation. For the purpose we select a middle-capitalization stock index in India called CNX Midcap 200, which represents around 72% of middle order firms listed in the National Stock Exchange of India. We use entropy of volatility to measure information content on various time steps as well as using various pricing parameters as given above. We also use GARCH (1,1) to find out impact of volatility. Based on the analysis of entropy, the results indicate that there has been relatively high impact of the volatility computed on high-low-closing prices and the lowest impact is found for volatility computed on high prices of the securities. The above result also confirms that the entropy of volatility is a valuable indicator for evaluating the performance of the volatility.

Debasis Bagchi Ph.d
Secretary and Executive Head
Institute of Cost & Works Accountants of India
12. Sudder Street
Kolkata 700 016
e-mail : dbagchi2002@yahoo.co.in
Alternate e-mail : bagchi.debasis@gmail.com

An Analysis of Entropy of Volatility of Stock Return

Introduction

The prediction of movements of stock prices has attracted the attention of both the investors and the academics. The academics are concerned with how prices are determined and whether future prices can be predicted. Both fundamental as well as model specification attempts are reported. The most developed theory on financial market that still withstood the test of time on the understanding of the movements of stock prices is based on market efficiency or Efficient Market Hypothesis. An important implication of EMH is, that the future market price of the security will be dependant on future information, which by nature is unpredictable and therefore, the future market price will also be unpredictable. This implication was subsequently challenged on the basis of overwhelming evidences that pointed out that apart from information that are reflected in the security prices, there are several other variables that can be relied upon to predict the future prices of the securities. That the past stock returns can predict future stock returns in short term, intermediate term and long term are well documented by a number of researchers, notably by Jegadeesh and Lehmann(1990), who found strong return reversal in short term time horizon (one month and six months), by Jegadeesh and Titman (1993) in intermediate time horizon (three to twelve months), Debondt and Thaler (1985) long term price reversals (three to five years). Lo and McKinley (1999) investigated short-run stock prices and found non zero serial correlation in the process, with the existence of a large number successive moves in the same direction. Reasons to substantiate such return phenomena are still not settled. The researchers have also documented several indicators that are found to have predictive ability. However, most important variable which is difficult to understand is the behaviour of volatility of the security prices.

It has been well documented that stock price behavior is influenced by volatility of stocks and theoretically one can compute volatility in various ways. One way is computing on the basis of past prices, which is called historical volatility, while the other is computed on the basis of Black-Scholes model, known as implied volatility. Although the implied volatility is important to compute option prices, historical volatility is nonetheless important to understand movement of prices of the underlying assets.

Motivation

There has been an argumentive difference between the efficiency of historical volatility and that of implied volatility. The latter is a better measure for understanding option prices, it is argued. However it can not be said with certainty that historical volatility is a poor measure for prediction purposes. Moreover, historical volatility can be calculated in different ways and on the basis of different time-steps. The historical volatility measure that we usually compute, is based on daily closing prices, of the past 10, 20, 30, 100 etc., daily returns. What makes the calculation defective is that all the returns get equal weightage, which is not true and this gives rise to plateauing effect of volatility. In order to eliminate this, we can compute volatility on exponentially moving average of the square of price return. It however eliminates plateauing effect but gives rise to exponential decay effect. We can also use day's high and low values of the assets prices to estimate volatility. It is argued that closing price is a fixed time price, while high and low prices could occur at random and at any point of time. This might give a better estimate of volatility. It is also possible to model volatility on high and low prices and include closing prices or even exponentially weighted average method to calculate volatility. An understanding of the effectiveness of the volatility calculated under the above different methods to predict future prices can bring more accuracy in the understanding of the behaviour of volatility, particularly pricing of the underlying assets.

It is also important to understand the information content of the volatility for different time steps. Here we use the information measure as the (Shannon's) entropy of the volatility, which is designed to measure information content of the volatility. We then compare information content of the different time step volatilities. If the information content is higher for a particular time step when compared to other time step, it signifies that there is a gain in information content and that may lead to higher predictive ability of the volatility for the discovery of future pricing of the underlying assets, while low information would lead to lower predictive ability. In sum we attempt to find out the following:

Case 1. Whether different volatility measures yields significant differences in information measures, i.e., entropy, and

Case 2. Whether analyzing information measures could lead to identification of enhanced predictive ability of different volatility measures calculated on high, low, closing prices and high-low-closing prices.

Data and Methodology

The data are collected on the cnxmidcap-200, which is a stock market index in India. The medium capitalised segment of the stock market is being increasingly perceived as an attractive investment segment with high growth potential. The primary objective of the CNX MidCap 200 Index is to capture the movement and be a benchmark of the midcap segment of the market. CNX Midcap 200 represents about 72% of the total market capitalization of the Mid-Cap Universe and about 70% of the total traded value of the Mid-Cap Universe (Mid-Cap Universe is defined as stocks having average six months market capitalization between Rs.75 crores and Rs.750 crores). (1crore = 10 million) (Source : www.nse-india.com)

The data are collected on daily basis commencing from April 1, 2002 to July 15, 2005. The stock market in India has undergone a stable growth during the time and after this period a high growth has occurred in the market. We decidedly exclude the high growth period in order that our results yield acceptable results free from biased and unwieldy movement in return and volatility. The stock return is log return of the day's closing price over the yesterday's closing price. Similar stock returns are calculated on the basis of high and low prices. The volatility is calculated on time step 5-day, 10-day, 15-day and 20-day basis.

The entropy measure of volatility is calculated as follows. The basic premise of the information theory is, that occurrence of an event may have several possibilities to which probabilities could be assigned. When information about the event is received these probabilities undergo changes. These changes in probabilities enable measurement of amount of information contained in the message. Mathematically[1], the information function could be expressed as :

$$h(p) = -\log p \ ; \ 0 \leq p \leq 1 \qquad (1)$$

The function monotonically decreases from ∝ to 0, that is, infinite information, when probability is zero and no (zero) information, when probability of occurrence of an event is 1 (one).

Extending the concept to n events, given by E_1, E_2, ………… E_n, with probability p_1, p_2, …… p_n, we find that, since one of the events is bound to occur, all the probabilities, p_1, p_2, …… p_n, will sum to one. Now if the event E occurs, the amount of information in the message stating the occurrence of E is h (p_i) and the expected information content of the message, i.e., the entropy of the distribution with probabilities $p_1, p_2, \ldots p_n$, is given by :

$$H(p_1, p_2, \ldots p_n) = \sum_{i=1}^{n} p_i \, h(p_i) = -\sum_{i=1}^{n} p_i \log p_i \ ; \qquad p_i \geq 0, \ \sum_{i=1}^{n} p_i = 1 \qquad (2)$$

and we define: $p_i \log p_i = 0$, when $p_i = 0$

The above expression is the entropy of a discrete distribution with probabilities $p_1, \ldots . p_n$.

Let us now consider a case that states that instead of certainty of happening of E as stated above, the message states that the probability of happening of the event has changed from *p to q*. In such case, the information content of the non-definite message will be :

$$h(p) - h(q) = -\log p + \log q = \log q / p \qquad (3)$$

---

[1] Shannon, C.E. "A mathematically theory of communication", Bell System Technical Journal XXVII (1948).

In other words, the expression (3) is the measure of uncertainty of volatility, since it is the difference of information content of the two states, i.e., the previous day's information content and the current day's information content.

Extending the above concept to n mutually exclusive events $E_1 \ldots\ldots E_n$ having probabilities of occurring $p_1, \ldots\ldots p_n$ which change to $q_1, \ldots\ldots q_n$. The expected information content of the non-definite message will be :

$$I = q_1 \log q_1 / p_1 + \ldots\ldots + q_n \log q_n / p_n \quad = \quad \sum_{i=1}^{n} q_i \log q_i / p_i \qquad (4)$$

The probabilities p (previous day's) and q (current day's) are calculated individually for various time steps, i.e., 5-day, 10-day, 15-day and 20-day. One of our objectives is to find out whether the uncertainty is important in evaluating volatility and in which time step it is important. The range and mean of entropy would provide such information. If the range is large, it signifies that uncertainty is more and when higher uncertainty is captured in the volatility its predictive ability is likely to be enhanced for discovery of future prices of the assets. Accordingly entropy of volatility is used to isolate the concerned time-step volatility so as to understand whether higher entropy leads to higher predictive ability, since, it is expected that increase in entropy leads to increase in information content of the volatility.

Since our time step is small, we use conventional method to calculate volatility on the basis of high, low and closing prices. In addition we compute volatility taking high, low and closing prices as inputs and the following expression is used for the purpose:

$$\text{Volatility}_{HLC} = \sqrt{1/(n-1)d_t \left[\sum_{i=1}^{n} (0.5(\ln H_{(ti)} - \ln L_{(ti)})^2 - 0.39\,(\ln C_{(ti)} - \ln C_{(ti-1)})^2\right]}$$

Where $H_{(ti)}$, $L_{(ti)}$ and $C_{(ti)}$ are highest, lowest and closing stock prices on day t_i respectively, while d_t is the time step. (see Wilmott : Quantitative Finance p. 60)

Based on preliminary findings, a two-stage GARCH(1,1) model is used to find out the relative impact of volatility on the price discovery mechanism. In GARCH, the conditional variance is dependant on the past squared residual of the process. On that basis, we develop the following GARCH(1,1) model :

$$R_t = gVol_{t-1} + e_t$$

$$s_t^2 = a + be_{t-1}^2 + cs_{t-1}^2$$

where, R_t is the return on stock calculated on daily basis, s_t^2 is the one-period ahead forecast variance based on past performance, e_{t-1}^2 is the measure of lag of squared residuals of the past volatility, while the last period forecast variance is given by s_{t-1}^2.

The main purpose of GARCH test is to understand effect of volatility of stock computed on the basis of three parameters, viz, high, low and closing prices on the stock return and the combination of these three, by examining the behaviour of conditional mean and conditional variance of the volatility.

Results

The GARCH test is conducted using Eviews software. The GARCH results of volatility based on high, low, closing and high-low-closing prices are given in Table – 1, while the descriptive statistics on the value of entropy for the different time steps are given in Table – 2. We also test the series on possible existence of unit root. The values of ADF shown in Table-1 are all significant at 1 per cent level. We find that the ADF statistic rejects the hypothesis of existence of unit root in the series. An examination of Table 2 shows the mixed results. While the highest entropy value is found to be in the low price volatility for the time steps, viz., 20-day, 10-day and 5-day, while the minimum value of the same is found to be for volatility calculated on high-low-closing prices, except in the time step 5-day. The range which shows the difference of entropy values between maximum and minimum is also an indication of the uncertainty measured in the volatility. The maximum value in this regard is found to be for the volatility computed on low prices, while the minimum has been found to be for closing (10-day and 5-day) as well as high-low-closing (20-day and 15-day) price volatilities. There has been a clear indication of maximum and minimum mean values. The maximum mean has been found to be in respect of volatility calculated on closing prices while the minimum is found to be for the volatility calculated on high-low-closing prices. In the intra-group entropy values, it has been found that a steady increase occurs in the entropy values of volatility from 20-day to 5-day, in all the cases, except for the entropy values of volatility, based on closing prices where the opposite is the case. Accordingly, the information content of the volatility is found to increase with lower time step, while for the closing prices, the information content decreases with shorter time step. Such behaviour of entropy can not be explained given the available evidences. It is, therefore, apparent that as the difference of entropy gives the uncertainty measure of the volatility, volatility becomes more stable for the 20-day time step rather than for 5-day time step. The result from the analysis of the entropy is expected to be reflected on the impact of volatility on the security prices. It is expected, therefore, that the Υ will be an important indictor for understanding the character of volatility on the security prices.

On examination of Table - 1, Υ is found to be statistically significant in most of the cases and is found to be highest for the volatility based on high-low-closing prices. The 20-day time step emerges as having the highest impact in all cases of volatilities except for high price parameter volatility. It is, therefore, apparent that 20-day time step volatility is a more stable measure of the volatility and, accordingly, makes more impact on the future price of the securities in relation to other time steps.

The examination of Table - 1 reveals that the impact of conditional variance (b) of volatility measure is found to be statistically significant in all the cases. Here we find that there have been very little differences amongst the time steps and it is therefore apparent that volatility measured on 4 different methods has similar impact so far the conditional variance is concerned. Since coefficients are positive the impact of volatility on stock prices is directly proportionate. It therefore underlines that the impact of volatility measure computed on the basis of the high, low and closing prices and high-low-closing prices are in the positive direction on the price discovery mechanism. In addition, it is also observed that for all the volatilities the sum of co-efficient b and c is less than 1. It indicates that the process is a mean reverting variance process. In other words the volatility shocks that are generated in different time steps are transmitted to longer time steps in the process.

Conclusion

We attempt to find out the relative impact of volatility measure computed from 4 parameters, namely, high, low, closing and high-low-closing prices of the day on the stock index return. In order to measure the effect on uncertainty as reflected in the volatility, we use Shannon's entropy measure. The difference of entropy shows the stability of the volatility measure. Lower the difference of entropy between maximum and minimum values, lower is the uncertainty associated with it. We attempt to find out whether time step is important in volatility measure and if so whether shorter or longer time step is significant By the above approach, the volatilities are categorized for their expected predictive ability. We then employ volatilities to find out validity of the above hypothesis.

For the purpose we select a mutual capitalization stock index called CNX, Midcap 200 which represents around 72 per cent of middle order forms listed in the National Stock Exchange in India. The return of the index is computed continuously from 20-day, 15-day, 10-day and 5-day respectively. The volatility is likewise computed on the same above basis, i.e. for the period 20-day, 15-day, 10-day and 5-day continuously. The results show that there are divergence in the intra-time step entropy values as also for different price parametric volatility entropy values based on high, low, closing and high-low-closing prices. Since the difference of entropy values, i.e. the range, is an indication of the uncertainty measure of the volatility, it is expected that the volatility measure will behave likewise for the prediction of future security prices.

To understand such behaviour, we use GARCH (1,1) model for evaluating the impact of volatility on the security prices based on conditional mean and conditional variance. It has been found that there has been relatively high impact of the volatility based on high-low-closing prices and the lowest impact is found for volatility based on high prices of the securities. The above result also confirms that the entropy of volatility is a valuable indicator for evaluating the performance of the volatility measure.

Bibliography

Bekaert, Geert and Campbell R. Harvey, "Emerging equity market volatility," *Journal of Financial Economics*, 43, 1997, pp.29-77.

Black, F., : 'Noise' , *Journal of Finance,* 41 (1986), pp.529-544

Bollerslev, T., "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 1986, pp.307-27.

Dicky, D. A. and W. A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of American Statistical Association*, 74, 1979, pp.427-431.

Fama, E., : 'Efficient Capital Markets : A review of Theory and Empirical Work', *Journal of Finance,* 25 (1970), pp 383-417

Lo. A. and C. Mackinlay, : 'A Non-Random Walk Down Wall Street'. Princeton, NJ, Princeton University Press (1999)

Table 1

Estimation results GARCH(1,1) model for CNX MidCap 200 Index returns, volatility computed on closing, high and low quotes obtained on daily basis for the period between 1st April, 2002 and 15th July, 2005, using the following model:

$$R_t = gVol_{t-1} + e_t$$
$$s_t^2 = a + be_{t-1}^2 + cs_{t-1}^2$$

where, R_t is the return on stock calculated on daily basis, s_t^2 is the one-period ahead forecast variance based on past performance, e_{t-1}^2 is an ARCH term which is the measure of lag of squared residuals of the past volatility, while the last period forecast variance is given by s_{t-1}^2 and
where c represents the conditional variance of the stock index return, R_t, at time t and e_t is the residual which is assumed to follow normal distribution with zero-mean, and a time varying standard deviation, s_t.

Volatility based on day's highest price

Time step	ϒ	*a*	*b*	*c*	ADF statistic
20-day	.006675 (3.129)*	.000000171 (1.5921)	.15 (2.8516)*	.60 (3.7177)*	-5.7473*
15-day	.9620 (3.3348)*	.00000684 (2.4248)	.161 (4.0717)*	.811 (21.68)*	-6.0513*
10-day	0.2124 (5.2837)*	.00000817 (2.6959)*	.1684 (4.0796)*	.797 (20.408)*	-5.9110*
5-day	.4321 (5.5024)*	.00000775 (2.7339)*	.1645 (3.9754)*	.8031 (21.70)*	-8.2670*

* Significant at 1% level.

Volatility based on day’s lowest price

Time step	Υ	*a*	*b*	*c*	ADF statistic
20-day	.59673 (2.5655)*	.00000705 (2.5253)	.163 (4.1559)*	.809 (22.22)*	-5.8740*
15-day	.4715 (2.3213)	.00000699 (2.4930)	.164 (4.2161)*	.808 (22.42)*	-5.9514*
10-day	.4643 (4.735)*	.00000771 (2.6573)*	.168 (4.0398)*	.801 (21.36)*	-5.4826*
5-day	.3503 (5.3517)*	.00000796 (2.7164)*	.1741 (4.0695)*	.7941 (21.101)*	-8.1083*

Volatility based on day’s closing price

Time step	Υ	*a*	*b*	*c*	ADF statistic
20-day	.2124 (5.2637)*	.00000817 (2.6959)*	.168 (4.0796)*	.797 (20.41)*	-5.9110*
15-day	.0434 (5.0299)*	.00000734 (2.5616)	.166 (4.5070)*	.804 (21.2618)*	-4.5322*
10-day	.0558 (5.0946)*	.00000778 (2.6704)*	.170 (4.0585)*	.798 (21.14)*	-5.3243*
5-day	.0769 (4.4726)*	.00000767 (2.6780)*	.1698 (4.067)*	.800 (21.68)*	-6.9905)*

Volatility based on day’s highest, lowest and closing prices

Time step	Υ	*a*	*b*	*c*	ADF statistic
20-day	.7486 (4.3281)*	.00000738 (2.5764)*	.1628 (4.0575)*	.8072 (21.41)*	-4.4080*
15-day	.6166 (4.1295)*	.00000741 (2.5722)	.1628 (4.0646)*	.8072 (21.53)*	-5.1906*
10-day	.4356 (3.5178)*	.00000726 (2.5397)	.1618 (4.1171)*	.8088 (21.97)*	-6.708*
5-day	.2984 (3.3567)*	.00000725 (2.5318)	.161 (4.0812)*	.810 (21.89)*	-9.2391*

* Significant at 1% level.

Table 2

Entropy values of volatility for different time step

Estimated entropy calculated on the basis of following expression :

$$H(p_1, p_2, \ldots p_n) = \sum_{i=1}^{n} p_i\, h(p_i) = -\sum_{i=1}^{n} p_i \log p_i\,; \qquad p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1$$

The probability p is calculated individually for various time-steps, i.e., 5-day, 10-day, 15-day and 20-day respectively from the volatility calculated on the basis of day's high, low, closing prices as well as calculated on the basis of following expression for high-low-closing price :

$$\text{Volatility}_{HLC} = \sqrt{1/(n-1)d_t \left[\sum_{i=1}^{n}\left(0.5(\ln H_{(ti)} - \ln L_{(ti)})^2 - 0.39(\ln C_{(ti)} - \ln C_{(ti-1)})^2\right]\right.}$$

Based on volatility calculated on day's Highest price

Time step	Maximum	Minimum	Range	Mean	Standard deviation	Skew-ness	Kurtosis	Sum
20-day	.00457	.00013	.00443	.000996	.000787	2.324	6.426	.78917
15-day	.00623	.00021	.00602	.00122	.000999	2.490	7.600	.96640
10-day	.01069	.00022	.01046	.00157	.0013	3.085	14.471	1.24337
5-day	.02254	.00024	.02230	.002282	.00238	3.485	19.360	2.23217

Based on volatility calculated on day's Lowest price

Time step	Maximum	Minimum	Range	Mean	Standard deviation	Skew-ness	Kurtosis	Sum
20-day	.00734	.00021	.00713	.00141	.00132	2.318	5.805	1.11574
15-day	.00927	.00025	.00902	.00164	.00153	2.559	7.634	1.29508
10-day	.01374	.00018	.01356	.00201	.00193	2.884	11.074	1.58962
5-day	.03108	.00020	.03088	.00339	.00354	3.911	21.520	2.68875

Based on volatility calculated on day’s Closing price

Time step	Maximum	Minimum	Range	Mean	Standard deviation	Skew-ness	Kurtosis	Sum
20-day	.0144	.00828	.00613	.00924	.00109	2.663	7.900	7.31527
15-day	.01311	.00637	.00674	.00788	.00957	2.566	8.771	6.23891
10-day	.01118	.00561	.00557	.00634	.00777	3.438	15.655	5.02049
5-day	.01106	.00373	.00733	.0044	.00715	4.578	31.019	3.43378

Based on volatility calculated on day’s High, Low and Closing price

Time step	Maximum	Minimum	Range	Mean	Standard deviation	Skew-ness	Kurtosis	Sum
20-day	.00373	.00012	.00360	.000878	.000717	1.887	3.578	.69525
15-day	.00526	.00015	.00511	.00107	.000911	2.160	5.351	.8415
10-day	.00818	.00013	.00806	.00139	.00122	2.539	8.684	1.10326
5-day	.01823	.00027	.01796	.00224	.00218	3.403	16.340	1.2067